Libro de resúmenes

VII Congreso de la
Sociedad Internacional de

HUMANIDADES DIGITALES HISPÁNICAS

Valencia, 22-24 de octubre de 2025

Compromiso cívico y social de las Humanidades Digitales

Daniel Fernández Rodríguez
Víctor Huertas Martín
Rafael Massanet Rodríguez (eds.)

HDH Humanidades Digitales Hispánicas
Universitat de València 525 anys
UVcultura

VII Congreso de la Sociedad Internacional de Humanidades Digitales Hispánicas

Compromiso cívico y social de las Humanidades Digitales

(Libro de resúmenes)

Daniel Fernández Rodríguez, Víctor Huertas Martín,
Rafael Massanet Rodríguez (eds.)

Valencia, 22-24 de octubre de 2025

Organizan:

Colaboran:

Edición: Daniel Fernández Rodríguez, Víctor Huertas Martín, Rafael Massanet Rodríguez

ISBN: 978-84-9133-830-7

DOI: https://doi.org/10.7203/PUV-OA-9788491338307

Esta obra está bajo una Licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional.

Junta Directiva de la HDH

Fátima Díez Platas (presidenta)
César González Pérez (vicepresidente)
Carlota Cristina Fernández Travieso (secretaria)
María Luisa Díez Platas (tesorera)
Alejandro García Reidy (vocal 1)
Patricia García Sánchez-Migallón (vocal 2)
Patricia Martín Rodilla (vocal 3)
David Merino Recalde (vocal 4)

Comité Local Organizador (Universitat de València)

Dirección

Teresa Ferrer Valls

Comité Asesor

Ester Alba Pagán
Joan Oleza Simó
Antonio Briz Gómez
Marta Haro Cortés
Jesús Tronch Pérez

Comité Ejecutivo

Josefa Badía Herrera
María Estellés Arguedas
Jorge Sebastián Lozano
Adrián Cabedo Nebot
Jaime Yefi Verdugo
Ariadna Fuertes Seder
Luz Souto Larios
José Martínez-Rubio
Daniel Fernández Rodríguez
Clara Monzó Ribes
Rafael Massanet Rodríguez
Purificació Mascarell
Verónica Romero Gómez
Jaume Segura Garcia
Dorota Kotwica
Cristina Villalba Ibáñez
Cristina Portalés Ricart
María del Mar Gaitán Salvatella
Vanesa García López de Andújar

Víctor Huertas Martín
Rosa Álvarez Sellers

Comité Científico

Fausta Antonucci (Università Roma Tre)
Ignacio Arellano Ayuso (Universidad de Navarra)
Nieves Baranda Leturio (Universidad Nacional de Educación a Distancia)
Sònia Boadas Cabarrocas (Universitat Autònoma de Barcelona)
Miriam Bouzouita (Universität Berlin)
José Calvo Tello (Göttingen State and University Library)
Pascual Cantos Gómez (Universidad de Murcia)
Jodi Cranston (Boston University)
Iria da Cunha (Universidad Nacional de Educación a Distancia)
Bárbara de Cock (Université Catholique de Louvain)
Gimena del Río Riande (CONICET)
Santiago Fernández Mosquera (Universidade de Santiago de Compostela)
Alejandro Gardía Reidy (Universidad de Salamanca)
Leonardo Impett (University of Cambridge)
Simon Kroll (Investigador independiente)
Sagrario López Poza (Universidade da Coruña)
José Manuel Lucía Megías (Universidad Complutense de Madrid)
Ricardo Maldonado Soto (Universidad Nacional Autónoma de México)
María Dolores Martos Pérez (Universidad Nacional de Educación a Distancia)
Rocío Ortuño Casanova (Universidad Nacional de Educación a Distancia)
Nieves Pena Sueiro (Universidade da Coruña)
Santiago Pérez Isasi (Universidade de Lisboa)
Marco Presotto (Università di Trento)
Nuria Rodríguez Ortega (Universidad de Málaga)
Antonio Rojas Castro (Universität Tübingen)
Dolores Romero López (Universidad Complutense de Madrid)
Jorge Ignacio Serrano Cobos (Universitat Politècnica de València)
Paul Spence (King's College London)
Simone Trecca (Università Roma Tre)
Alejandra Ulla Lorenzo (Universidad de Santiago de Compostela)
Germán Vega García-Luengos (Universidad de Valladolid)
Maurizio Vitella (Università degli studi di Palermo)

Índice

Comunicaciones

El Objeto Digital Aurático (ODA): una propuesta para la preservación crítica del arte digital en la cultura contemporánea

Elena Battaner Moro; Juan Alonso López-Iniesta 17

Entre píxeles y cal: gemelos digitales y narrativas visuales para la preservación del mural moderno

Leticia Crespillo Marí 21

Tecnologías avanzadas para la conservación del patrimonio cultural: ChemiNova y HBIM como nuevos Sistemas de Gestión de Contenido para bienes culturales

Esperanza Villuendas Ferri 30

Realidad Aumentada como herramienta participativa para la activación del patrimonio: el caso de TinajAR

Bibiana Martínez Torrecilla; Lucía Vera Mahiques; Cristina Portalés Ricart; Carlos Martínez Barragán 34

El juez de residencia y su desempeño. Un análisis desde la representación cartográfica con SIG

Víctor González-Melgar 41

Uso de los proyectos Wikimedia en instituciones y proyectos de patrimonio cultural

Gustavo Candela; Ángel Obregón Sierra 46

Interpretar, documentar y gestionar digitalmente la movilidad como patrimonio cultural: una propuesta conceptual y metodológica

Sonia Medina Gordo; César González Pérez; Esther Travé Allepuz 52

Inteligencia artificial y visión por computador para la memoria colectiva: preservación digital del patrimonio en riesgo

Laya Targa Hernández 59

Digitalización de fondos documentales: el buscador de pasajeros de la Compañía Trasatlántica Española

Silvia Sanz Prieto 64

Nueva museografía y museología del Museo Fray Pedro Gocial de Quito desde las Humanidades Digitales

Beatriz Garrido-Ramos; Ana María Zabía; Mercedes Sevillano Pareja . 68

Scripta manent, sed quid de digitalibus? Desafíos y resultados de un proyecto de Historia Medieval tras trece años de investigación entre archivos y pantallas

Cristina Jular Pérez Alfaro; Raúl Villagrasa Elías 73

Pablo Picasso y su red: un análisis del ecosistema cultural del siglo XX

Fatiha Idmhand 79

Búsqueda por contenido en bases de datos federadas: una nueva funcionalidad para ASODAT

Ariadna Fuertes Seder; Jaime Yefi Verdugo 81

Nuevas posibilidades de exploración filológica para el teatro áureo: búsqueda por contenido en ASODAT

Alejandro García-Reidy; Teresa Ferrer Valls; Josefa Badía Herrera; Jesús Tronch Pérez; Alejandra Ulla Lorenzo; Lola Josa González; Héctor Urzáiz Tortajada 85

Hacia el reconocimiento automático de tipos impresos antiguos: el proyecto RETIPIA

Paula Casariego Castiñeira; Alejandra Ulla Lorenzo 90

Procesamiento de datos prosódicos y sociolingüísticos del corpus PRESEEA-Valencia: generación de informes automatizados con R, Shiny y ChatGPT

Adrián Cabedo Nebot 94

De la construcción a la exploración: Humcor, un corpus oral multimodal para el estudio del humor en español

Doina Repede 98

La manifestación de cortesía valorizadora y mitigadora en el contexto del malentendido digital: un análisis pragmático de interacciones en Reddit

Yuezhe Du 102

Discurso, coherencia y conflicto patrimonial: análisis computacional de estructuras retóricas en torno a los ataques patrimoniales durante el *Black Lives Matter*

Patricia Martín-Rodilla; Martial Pastor 105

Leer a Aristóteles con IA: un modelo ampliado de lectura profunda de textos filosóficos

José Higuera Rubio; Salvador Ros 113

Nigromancia Digital: Hacia una epistemología pos-hermenéutica para las Humanidades Digitales

Ángel M.ª Lumbreras 118

Escenas conectadas: Humanidades Digitales e Inteligencia Artificial para el estudio y difusión del teatro musical en español (1650–1800)

Juan Pablo Fernández-Cortés 125

IA, prensa histórica y literatura hispanofilipina: las HD y el estudio de narrativas postcoloniales en español

Diana Villanueva-Romero; Emilio Vivó Capdevila 130

Los modelos 3D como herramienta de análisis y preservación patrimonial: el caso del Obrador Botifoll (Martorell)

Isidre Ot Padilla; Esther Travé Allepuz 135

ChemiNova: Nuevas Herramientas para la Inspección y Documentación del Patrimonio Cultural en Riesgo

Carmen Cano; Belén Palma; Esperanza Villuendas; Laya Targa; Álvaro Solbes; Sergio Casas; Ester Alba; Cristina Portalés..... 142

Recreando un final del Camino que no pudo ser. El modelo virtual de la Rúa Galicia ideada por Antonio Palacios para Santiago de Compostela (1932)

Jesús Ángel Sánchez García..... 150

Herramientas digitales para la conservación, análisis y activación del patrimonio del diseño en el Arxiu Valencià del Disseny

Arabella León Muñoz..... 158

Cartografías entrelazadas: modelar la transnacionalidad en el ecosistema expositivo español (2000-2023) mediante análisis de redes

Nuria Rodríguez-Ortega; Bárbara Romero-Ferrón 163

Desafíos y oportunidades en la preservación digital de las colecciones textiles de seda en museos europeos

Mar Gaitán..... 173

Persuasión y emoción: la prosodia como herramienta argumentativa en el discurso político oral

Carlos Castelló Vercher; Noelia Ruano Piqueras 177

Corpus ADUMAT: claves y retos para la creación de un corpus de habla de adultos mayores

Cristina Villalba Ibáñez 182

Memoria colectiva y reconstrucción virtual orientados a la terapia del deterioro cognitivo: estudio piloto del caso de la ciudad de Burgos en 1970

Roberto Martín González; Montserrat Santamaría Vázquez; Olalla Sáiz Vázquez..... 186

Confluencias significantes: explorando las relaciones de producción y las conexiones iconográficas entre las ediciones ilustradas de las *Metamorfosis* de Ovidio y los *Emblemata* de Alciato de los siglos XVI y XVII

Fátima Díez Platas; María Luisa Díez Platas; Paloma Centenera Centenera..... 192

Hécuba: uso de las HD para el estudio comparado de un mito en la enseñanza preuniversitaria

Teresa Santa María; María Luisa Díez Platas 196

Compartimos experiencias con AMIS (Advanced Metadata Intelligent System)

Fatiha Idmhand; Amelia Sanz 202

DraMÆ (base de datos, estrategias dramatúrgicas y motivos): una nueva herramienta digital para el estudio del teatro de la memoria democrática en España

Jessica De Matteis 205

Hacia un paradigma dataísta de la memoria. Potencialidades de MNLAB en el almacenamiento, tratamiento y análisis de datos sobre la literatura de la memoria en España

José Martínez Rubio; Luz C. Souto 210

GEXEL Digital: nuevos enfoques sobre la literatura del exilio republicano español de 1939

Lucía Cotarelo Esteban 215

Humanismo digital y lengua latina: el latín como espacio de reflexión crítica y metalingüística

Germán González Muñoz 220

Inquietar inteligencias, despertar conciencias: humanos artesanos. La IA en los procesos de formación de docentes colaborando en la construcción de competencias humanizadoras.

Patricia Bibiana Ramírez Lombardi 225

Tecnología y preservación del patrimonio cultural en contextos de emergencia: herramientas y desafíos contemporáneos, el caso de la Dana en Valencia

Alejandra Nieto Villena; Álvaro Solbes García; Carmen Cano Sola 230

El Archivo Histórico Riva-Agüero (PUCP) y el acceso digital a colecciones patrimoniales privadas

Magally Alegre Henderson 235

La Red Goya: Humanidades Digitales para mapear la recepción de Goya en el siglo XIX

Paula Fayos Pérez; Anita Verö 238

Soluciones y nuevos modelos para la edición crítica digital de la obra de Calderón

Candela Iglesias Balsa 244

Primeras calas en *Valor, ingenio y fineza*: edición digital y análisis de redes

Carles Márquez Molins 248

Una estrategia sostenible para la edición crítica digital del teatro del Siglo de Oro: el caso de PROLOPE

Anna Abate; Emanuele Leboffe; David Merino Recalde 255

Proyecto SCORE: investigación musicológica, edición digital y modelo ontológico a partir de la correspondencia de Kurt Schindler (1882-1935)

Sara Escuer Salcedo 261

Lo que aprendimos sobre el procesamiento de entrevistas de historia oral en español chileno para TF-IDF y análisis de sentimientos

Nicole Larrondo; Eduardo Cerda 271

Herramientas digitales transmedia y enseñanza ELE: cómo insertar las leyendas en el aula

Carmen Cazorla Vivas 278

Hacia un ecosistema FAIR de recursos lingüísticos en gallego: corpus oral Nos_TranscriSpeech-GL

Adina Ioana Vladu; Daniel Fernández López; Elisa Fernández Rei .. 2281

Lingüística de corpus e inteligencia artificial: desambiguación semántica de *pobo* (‘pueblo’) en gallego

Vítor Míguez Rego 288

Las redes sociales como herramientas para la enseñanza y difusión de la lengua gallega entre la juventud: iniciativas protagonizadas por docentes de Secundaria y Primaria

Carme Silva Domínguez 293

ESPRINT: un corpus digital para el análisis interaccional del conflicto verbal en contextos íntimos

Marta Albelda Marco; Maria Estellés Arguedas 296

Hacia un corpus de películas y series de televisión españolas del español coloquial: criterios de selección y metadatos

Ana Llopis Cardona 300

Análisis pragmalingüístico del discurso digital con IA: evidencialidad y vigilancia epistémica en debates sobre teorías de la conspiración

Dorota Kotwica 304

Modelo metodológico para el análisis semántico de corpus históricos: integración de OCR, ontologías y SPARQL

Rafael Prezado 307

Transcripción de textos históricos en español con diferentes fuentes y formatos: análisis y comparación de resultados con tres enfoques basados en IA

Ana García-Serrano; Jaione Macicior; Marina Míguez; Yanco Torterolo 313

Análisis de documentos novohispanos por medio del Reconocimiento y Desambiguación de Entidades Nombradas (NERD)

Diego Jiménez-Badillo 320

La memoria individual como herramienta de documentación etnográfica para la reconstrucción digital: el caso del penal de Valdenoceda

María del Pilar Bermúdez Martínez; Roberto Martín González; Mario Alaguero Rodríguez 333

Ni son todas las que están, ni están todas las que son: Estructuras, márgenes y mujeres artistas en el sistema expositivo andaluz

Bárbara Romero-Ferrón; Nuria Rodríguez-Ortega 339

De los átomos a las ondas: estilometría rítmica para el estudio de la autoría en el teatro del Siglo de Oro

Álvaro Cuéllar 346

Retrieval Augmented Generation como herramienta para el estudio de la censura teatral de los Siglos de Oro

Joel Reverter Lainez 351

Diseño e Integración de dispositivo IoCT en aplicación de Realidad Virtual

Guillem Montalbán Faet; Jaume Segura García; Mario Alberto Montagut Climent 356

Reconstrucción virtual inmersiva de teatros del *Siglo de Oro* desaparecidos: La Casa de Comedias de l'Olivera (València)

Eugenio Giuseppe Nerelli; Álvaro Egea Benavente; Jaume Segura-Garcia; J. Antonio De Rus; Mario Montagud 363

Geolingüística digital: el Atlas Lingüístico Galego en línea

Xulio Sousa; Pedro García-Louzao 372

El papel del reconocimiento de entidades para la investigación histórica.

Renata Vieira; Fernanda Olival; Helena Freire Cameron; Ana Sofia Ribeiro 377

Experiencias con el proyecto COST *Action Distant Reading for European Literary History* (CA16204) y el desarrollo del corpus ELTeC

Borja Navarro Colorado 384

Salvando el desajuste entre formación y necesidades del ámbito laboral en las Humanidades Digitales

María Goicoechea de Jorge 388

Data driven learning revisitado: del corpus a la IA generativa en el aula de español

Virginia González García 395

Mejora de la escritura en español mediante Inteligencia Artificial y reglas explícitas en contextos académicos

Natalia Centeno Alejandre; Pedro Manuel Moreno-Marcos; Marina Serrano Marín 399

Visualización 3D para la documentación, análisis y conservación de cuevas rituales ibéricas

Agustín Diez Castillo; Sonia Machause López; Daniel Negueruela Villanueva 403

Cartografiar la cerámica arquitectónica valenciana: tecnología y difusión en Nollamap

Vanesa García López de Andújar; Ángela López Sabater; Xavier Laumain 409

HILAME frente al espejo. Una mirada al futuro desde el presente

Agurtzane Paz Moro; Ana Galdós Monfort; Arsenio Dacosta 416

Canon clásico y academia hispanohablante: patrones de canonización en textos académicos del siglo XXI

Luisa Ripoll-Alberola; Andreas Niekler; Manuel Burghardt 422

Cartografiar el infinito: una lectura interactiva de *El infinito en un junco* (2019) de Irene Vallejo a través de ArcGIS StoryMaps

Soledad Castaño Santos 432

Construcción del relato turístico durante el desarrollismo: un estudio de textos oficiales sobre Zaragoza ORTS con Voyant Tools

Andrea Todolí Moreno; Ángela Celis Sánchez 437

Ciencia de datos para las Humanidades: estudio del lenguaje inclusivo con CRISP-DM

Ángela Celis; Gema Fernández-Avilés 442

La gestión y el procesamiento de datos del corpus Ameresco a través de la herramienta computacional Oralstats Aroca

Andrea Carcelén Guerrero; Adrián Cabedo Nebot 447

AIMeRiC: análisis informatizado del metro y el ritmo del catalán

Pere Garau Borràs 451

La *Gramàtica essencial de la llengua catalana*: una gramática online

Maria Josep Cuenca 456

El *Llibre dels feits* del rey Jaime I de Aragón: de la oralidad artística a la reproductibilidad técnica

Anna Saumell Almirall; Clàudia Raventós Albà; Pere Poy Baena 464

Hacia una transcripción asistida de la documentación inquisitorial: lineamientos para el entrenamiento de modelos (siglos XV-XIX)

Bárbara Santiago Medina 469

Modelizar para compartir. CÁLICE y la ecología de datos abiertos en la historiografía de la censura inquisitorial

Mathilde Albisson 474

TEI y RDF aplicados a la documentación de los procesos inquisitoriales

Andrés Felipe Echavarría Peláez 480

Inteligencia Artificial aplicada al análisis de corpus en prensa musical histórica: metodologías computacionales para el estudio de la crítica musical en español

María Palacios Nieto 485

Bases de datos relacionales para el estudio del libro musical en València (1413–1641)

Ferran Escrivà Llorca; Alex Wingate 490

Mapping Pliegos. Un catálogo colectivo y biblioteca digital de literatura de cordel

Juan Gomis Coloma; Pilar Martínez Olmo 494

Uso de técnicas avanzadas de imagen en la documentación de obras afectadas por la DANA: un caso de estudio en el proyecto ChemiNova

Carmen Cano Sola; Álvaro Solbes García; Ester Alba Pagán; Arabella León Muñoz; Esperanza Villuendas Ferri 498

Transformación semántica de museos virtuales locales

María Luisa Díez Platas; Susana Quirós Alpera; Luis Pedraza Gómara; Paloma Centenera 505

Sistematización metodológica para el análisis multidimensional del subsistema lingüístico-textual del dominio de las exposiciones artísticas

Nuria Rodríguez-Ortega; María Ortiz Tello; Ángel Lumbreras Fernández; María Luisa Díez-Platas 510

Novus ordo seclorum? Retos y posibilidades para la visualización de datos espacio temporales en las humanidades digitales

Jorge Sebastián Lozano 524

Analíticas y visualización de datos a partir de modelos heterogéneos de discurso

César González-Pérez; Beatriz Calderón-Cerrato; Martín Pereira-Fariña 528

Una métrica de similitud entre modelos argumentales para análisis de discurso

Ángel Méndez-Vázquez; César González-Pérez 535

Análisis cualitativo y de errores en un dataset sintético de question-answering para RAG sobre Trafalgar, de Galdós: ventajas y limitaciones

Sofía Micaela Roseti; Yanco Amor Torterolo Orta; Antonio Moreno Sandoval 541

The nearest neighbor problem en el análisis estilométrico del *Entremés del triunfo de los coches*, de Gaspar de Barrionuevo y Carrión

Cristina Ruiz Urbón 548

Constelaciones culturales en la Biblioteca Digital Mnemosine: La responsabilidad ética y social de las bibliotecas digitales con la docencia y la investigación universitarias

Amanda San Román Sastre; Dolores Romero López 552

Sobre datos, grafos y redes en abierto: Una lectura distante de María Lejárraga (1874-1974)

Dolores Romero López; Patricia García Sánchez-Migallón 556

Desafíos y posibles soluciones para el diseño y la compilación de corpus lingüísticos multimodales en línea: el caso de WA2

Elia Puertas Ribés 562

La extracción automática de datos con Python y Selenium aplicada a la lingüística de corpus

Elia Puertas Ribés 566

Catálogo y Biblioteca Digital de Relaciones de Sucesos: análisis y visualización de las colecciones que conforman la base de datos

Carlota Fernández Travieso; Nieves Pena Sueiro 569

Catalogación automatizada de incunables con la herramienta CACIN

Hala Neji; Javier Nogueras-Iso; Javier Lacasta; Manuel José Pedraza-Gracia 573

Jugabilidad sagrada: estructuras sacras y santas en el videojuego

Rubén Gregori Bou; Nadia McGowan 579

Imaginarios de la marginalidad urbana contemporánea en Chile: una aproximación interdisciplinaria al caso de la música trap y su difusión en videoclips a través de la plataforma Youtube

Rodrigo Ganter Solís; Angela Rivera Martínez 584

Inteligencias artificiales: ¿una reinvención del historiador como profesional? Menos clics, más inteligencia histórica

Chimo Soler Herreros; Antonio Carrasco Rodríguez 589

El diseño del Archivo Digital Sebastià Juan Arbó: el empleo del modelo no relacional para la construcción de la base de datos bibliográfica

Joan Antoni Forcadell; Joel Reverter Lainez 594

Creación de un modelo del HTR para letra cortesana a partir de procesos inquisitoriales del siglo XVI

Maryam Kebli Jebari 599

Adaptación de los proyectos de investigación en Humanidades Digitales al paradigma de la Ciencia Abierta

José Manuel Barrueco Cruz; Cristina García Testal 606

Tomando posiciones ante la descolonización del museo: análisis de la presencia africana en el ecosistema expositivo andaluz

Ángel María Lumbreras Fernández; María Ortiz Tello; Nuria Rodríguez Ortega 611

¿Quién custodia lo enseñado?: La preservación de objetos educativos digitales durante y después de la pandemia por el Covid-19 en escuelas de Colombia y España

Jennifer Restrepo Monsalve 618

La construcción discursiva del actor político en entornos digitales: análisis de la actancialidad en titulares mediáticos en redes sociales

Sofía Zakhir Puig 623

La ilusión de la conversación: marcadores del discurso en interacciones generadas por inteligencia artificial

Alessio Alessandro Santamato Peroni; Italo Cosentino 627

El Siglo de Oro en la prensa periódica (SILEM III): primera aproximación cuantitativa a los textos de la Biblioteca Digital Prensa y Canon

Laura Hernández-Lorenzo 631

Metodología de identificación de entidades mediante PLN en el corpus biográfico de DICAT

Ariadna Fuertes Seder; Verónica Romero Gómez; Luis de la Fuente Valentín 637

Anotación y detección de las emociones en cartas de mujeres del siglo XVI en español: los retos actuales

Albina Sarymsakova; Patricia Martín-Rodilla 642

Escritura, afectos y legitimación: redes sociales de autoras españolas (1890-1920)

Gema Duarte Abós 647

Pósteres

Progetto Mambrino. Tabula Imaginaria

Antonio Azaustre Lago; Stefano Bazzaco; Stefano Neri; Patricia Sánchez Migallón 653

Visualización de patrones de uso del vocabulario 'mobiliario y enseres' en AMERLEX mediante Voyant tools para acciones de transferencia

Mª Teresa Cáceres-Lorenzo; José-Ignacio Salas-Cáceres 655

El uso de LogosLink para la unificación de datos discursivos y etnográficos

Beatriz Calderón-Cerrato; Paula Sánchez-Carrera; César González-Pérez 658

Metodología de impresión 3D para recintos amurallados: aplicación y evaluación en Toledo y Algeciras

Daniel Castilla Rodríguez; Alfredo Miguel Maximiano Castillejo..... 662

Buscador de recursos TEISEL. De la ciencia abierta a las buenas prácticas en la investigación sobre segundas lenguas

Mar Cruz Piñol; Sophie McBride; Carmen López Ferrero..... 665

La dieta ideal en el más allá: análisis iconográfico y digitalización de tumbas egipcias

Inmaculada Delage González; Beatriz Garrido Ramos; Antonio Criado Martín; Juan Francisco Romero Collazos; Esteban García Viñas; Myriam Seco Álvarez; Daniel Becerra Fernández 668

Rasgos fonéticos y léxicos del español canario en las cartas de Rosalía Franchi (1820-1821)

Miguel Antonio Di Novella..... 673

Preservación y digitalización, un viaje de 150 años por el Archivo Histórico de la Sociedad del Gran Teatro del Liceo de Barcelona

Ana Escañuela Barradas 677

De Diccionarios históricos a Ontologías: el "Abub" y la Organología del s. XIX en LexiMus

Eugenia Gallego Cañellas; María Ordiñana Gil; Gorka Rubiales..... 680

Una ventana digital a la historia natural. Virtualización y divulgación cien]fica accesible desde el Museu de la Universitat de València de Historia Natural

Anna García Forner; José Antonio Villena Gómez; Natalia Conejero Ortega; Teresa Sáez Máñez; Sergio Montagud Alario 683

Cartografía bio-bibliográfica de un campo literario: el caso de la República Democrática Alemana

Jörn Kreutel; Axelle Lecroq; Thomas Möbius 685

ARStimuli: tecnologías digitales para el estudio de las obras de arte

Carme López Calderón 687

La biblioteca universitaria como apoyo a la docencia en innovación educativa y humanidades digitales

María del Don Martínez Aliaga, Elena Viqueira Lagares; Elisa Pastor Planells; Eva Montilla Navas 689

Preservación en formato futuro de la memoria en riesgo: difusión y conservación digital de la tradición abaniquera de Aldaia

Marta Segovia Cobo 695

El Objeto Digital Aurático (ODA): una propuesta para la preservación crítica del arte digital en la cultura contemporánea

Elena Battaner Moro
Juan Alonso López-Iniesta
Universidad Rey Juan Carlos

Palabras clave: Objeto digital aurático; Arte digital; Preservación; Humanidades digitales; Metadatos culturales
Keywords: Auratic Digital Object; Digital Art; Preservation; Digital Humanities; Cultural Metadata

Problema de investigación

La transformación radical del ecosistema cultural contemporáneo, marcado por la digitalización, la hiperconectividad y la aceleración de los procesos de creación y circulación de contenido, ha producido un nuevo tipo de objeto cultural: el objeto digital-arte. Este nuevo tipo de objeto desafía las herramientas tradicionales de análisis, archivo y preservación utilizadas por museos, bibliotecas, archivos y otros entornos patrimoniales. La pregunta central que guía esta investigación es: ¿es posible identificar y preservar una forma de "aura" en los objetos digitales contemporáneos? Y si es así, ¿cómo puede diseñarse un marco conceptual y técnico que permita su conservación a largo plazo?
Esta investigación propone el concepto de *Objeto Digital Aurático* (ODA) como una categoría teórica y una herramienta archivística destinada a identificar, describir y preservar las cualidades auráticas de los objetos digitales-arte. El objetivo es ofrecer un modelo integral que aborde tanto las dimensiones estéticas como las técnicas y contextuales de este tipo de objetos, dentro del marco de las humanidades digitales.

Metodología

Nuestra investigación se articula a través de cinco secciones que combinan una metodología teórico-analítica, histórico-crítica y propositiva:

- Marco conceptual: Relectura del concepto de "aura" propuesto por Walter Benjamin en diálogo con Heidegger, Brea, Boltanski y otros teóricos contemporáneos. A partir de esta base se propone una ampliación crítica del término "aurático" que permita aplicarlo a entornos digitales y desmaterializados.

- Análisis histórico-tecnológico: Revisión de las distintas eras de la imagen (imagen-materia, imagen-film, e-image, imagen-flujo) y de la evolución del objeto artístico desde su producción manual hasta su desmaterialización digital. Se problematizan las nociones de autenticidad, originalidad y valor simbólico en el contexto del arte digital.
- Propuesta del modelo ODA: Definición del concepto de Objeto Digital Aurático, articulado en torno a cinco criterios fundamentales: autenticidad, unicidad, presencia en red, reproducibilidad controlada y contexto cultural. Estos criterios son operacionalizados en una tabla (ver figura adjunta).
- Diseño archivístico y técnico: Desarrollo de un esquema de metadatos propio (ODA), basado en modelos como Europeana EDM y RiC-CM. El modelo describe atributos técnicos, estéticos, contextuales y curatoriales que permiten capturar la "condición aurática" del objeto.
- Propuesta institucional: Creación de la Fundación ODA como entidad gestora del modelo. La Fundación se concibe como un órgano interdisciplinar de preservación, evaluación y gobernanza del arte digital aurático, con la capacidad de operar en colaboración con instituciones museográficas, académicas y patrimoniales.

Resultados obtenidos

Los resultados principales de la investigación se pueden agrupar en tres bloques:

a) Conceptualización y redefinición del aura

Se plantea que el aura no debe entenderse exclusivamente como una propiedad ligada a la unicidad material del objeto, sino como un estado de percepción y valoración emergente de la interacción entre obra, creador, institución y público. Esta concepción se adapta a entornos digitales mediante el desarrollo de un continuum aurático, que permite categorizar los objetos digitales según su intensidad aurática.

b) Herramienta técnica de archivo: metadatos ODA

El esquema de metadatos ODA proporciona un marco sistemático para describir objetos digitales-arte atendiendo a:

- su contexto de creación (intencionalidad artística, autoría, medio de producción)
- su condición técnica (formato, duración, plataforma)
- su circulación (número de copias, control de versiones, contexto de exhibición)
- su recepción (curaduría, crítica, uso público)

Este modelo facilita la interoperabilidad con sistemas institucionales existentes y permite su aplicación a casos reales (archivos digitales de media art, piezas de net art, obras NFT desmaterializadas, etc.).

c) Aplicación institucional: la Fundación ODA

Se establece una infraestructura institucional que permite implementar y

extender el uso del modelo. La Fundación ODA actúa como centro de documentación, evaluación y mediación crítica. Su actividad se enfoca en:

- la certificación de objetos digitales-arte como auráticos
- la formación y asesoramiento a instituciones culturales
- la integración con colecciones digitales europeas e internacionales

Este ecosistema institucional y normativo apunta a crear condiciones de estabilidad para la valoración y preservación del arte digital, más allá de las lógicas especulativas del mercado (como en el caso de los NFTs).

Conclusiones y discusión

La propuesta del ODA representa una contribución original a los estudios en cultura contemporánea (artes y humanidades digitales), en tanto que:

- Propone una relectura activa de los marcos estéticos clásicos para aplicarlos a los desafíos de la cultura digital.
- Ofrece una herramienta técnica interoperable y adaptable a múltiples tipos de objetos digitales.
- Establece una conexión entre teoría, archivo y gestión institucional del patrimonio cultural contemporáneo.

A través del modelo ODA, se propone una estrategia de resistencia crítica frente a la desauratización acelerada del arte digital, afirmando su valor simbólico, su potencial narrativo y su lugar legítimo en las políticas culturales del presente.

Bibliografía

ALCALÁ MELLADO, J. R., & LÓPEZ INIESTA, J. A. (2021), «TetraArt: Un sistema de conservación y existencia del arte digital vinculado al código de programación», en *Conservación de Arte Contemporáneo*, Madrid, Museo Nacional Centro de Arte Reina Sofía, pp. 11–22.

BATTANER MORO, E., & LÓPEZ INIESTA, J. A. (2025), *El objeto digital aurático (ODA). Gestión, archivo y preservación de la cultura contemporánea digital*, Oxford / Bern / Berlin / Bruxelles / New York / Wien, Peter Lang.

BENJAMIN, W. (1973), «La obra de arte en la era de su reproductibilidad técnica», en *Discursos interrumpidos I* (J. Aguirre, trad.), Taurus, pp. 15–49.

BOLTANSKI, L., & ESQUERRE, A. (2022), *Enriquecer: Una crítica de la mercancía*, Editorial Katz.

BREA, J. L. (1998), «La era de la imagen técnica», en *Los últimos días: conversaciones en el siglo de la imagen técnica*, Cátedra.

BREA, J. L. (2010), *Las tres eras de la imagen*, Akal.

FUNDACIÓN ESPAÑOLA PARA LA CIENCIA Y LA TECNOLOGÍA (FECYT) (2023), *Libro Blanco de la interrelación entre arte, ciencia, tecnología y sociedad en España*.

GARCÍA-MORALES, L. (2019), *Teoría de la Conservación Evolutiva*, Cátedra.

HEIDEGGER, M. (1977), *El origen de la obra de arte* (S. Ramos, trad.), Alianza Editorial.

PREMIS EDITORIAL COMMITTEE (2005), *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*, OCLC Online Computer Library Center.

Entre píxeles y cal: gemelos digitales y narrativas visuales para la preservación del mural moderno

Leticia Crespillo Marí
Universidad de Málaga

Palabras clave: Patrimonio mural; Dordio Gomes; fotogrametría digital; curaduría expandida; análisis computacional
Keywords: Mural Heritage; Dordio Gomes; digital photogrammetry; expanded curation; computational analysis

Introducción: patrimonio mural y desafíos contemporáneos

En un momento en el que las formas de mediación, conservación y narración del patrimonio cultural desde un punto de vista museográfico evolucionan a un ritmo vertiginoso, el proyecto *Lime, Pigments, and Pixels. Bringing Modern Mural Heritage Back to Public Life: A Critical Approach to Computational and Digital Tools for Art Historical Curatorship*[1], aborda una cuestión central: ¿cómo puede la digitalización del patrimonio mural ir más allá de la documentación técnica y convertirse en una forma de reactivación crítica, narrativa, y visualmente situada del pasado?

La propia digitalización del patrimonio mural plantea actualmente un desafío metodológico, técnico y cultural de gran relevancia. Frente a la vulnerabilidad creciente de los frescos históricos en contextos urbanos sujetos a transformaciones estructurales. En tal sentido, el proyecto que aquí se presenta aborda la generación de gemelos digitales de catorce murales al fresco del artista portugués Dordio Gomes (1890–1976), ubicados en la ciudad de Oporto.

Dordio Gomes fue una figura clave del arte portugués del siglo XX, con un corpus mural relevante que abarca edificios institucionales, espacios religiosos y arquitecturas del Estado Novo (1933-1974). Pese a su importancia, muchos de estos frescos han quedado fuera de los circuitos patrimoniales, en parte por su difícil acceso físico, por su ubicación periférica o por la falta de estrategias de preservación digital y activación discursiva por parte de instituciones públicas y entidades privadas.

Se trata, por tanto, de una iniciativa transdisciplinar orientada a registrar, modelar, curar y narrar estos murales desde una doble perspectiva: como legado visual en riesgo y como fuente activa de interrogación cultural. El proyecto

[1] Financiado por fondos nacionales a través de la FCT – Fundação para a Ciência e a Tecnologia, I.P., en el ámbito del proyecto 2023.11110.PEX, DOI: https://doi.org/10.54499/2023.11110.PEX

propone un giro metodológico: no se trata solo de conservar imágenes del pasado, sino de activar su potencial narrativo y relacional en clave contemporánea. El objetivo ha sido no solo documentar con precisión estas obras, sino también abrir nuevas vías para su estudio, conservación preventiva, mediación curatorial y relectura crítica en el marco de las humanidades digitales desde una perspectiva no solo historiográfica.

Problema de investigación

Se parte de la necesidad urgente de registrar y poner en valor un conjunto de murales de alto valor histórico y artístico y, al mismo tiempo, de cuestionar los modos en que estos procesos de digitalización transforman no solo el objeto artístico, sino también sus condiciones de recepción, interpretación y circulación. Las condiciones de conservación de estos murales, entre los que se encuentran conjuntos del Café Rialto, la Cámara Municipal, el Tribunal da Relação o la Iglesia de los Redentoristas, son extremadamente frágiles debido a la falta de protección institucional, al deterioro de los inmuebles que los albergan y a la progresiva pérdida de memoria visual en el imaginario colectivo. En muchos casos, se trata de espacios de acceso restringido, abandonados o incluso demolidos, lo que convierte a estas obras en piezas de difícil estudio e inminente desaparición material.

Este contexto nos exige repensar el papel de la digitalización no solo como herramienta de archivo, sino como instrumento activo de intervención curatorial, académico-científica y divulgativa, produciendo un conocimiento múltiple y transdisciplinar sobre las propias características técnicas, materiales y simbólicas del mural moderno en constante relación con su pasado. además, ¿pueden entenderse los resultados de la propia digitalización fotogramétrica como una “alternativa poética”[2] que complementa y enriquece la transferencia de conocimiento del usuario?

Metodología

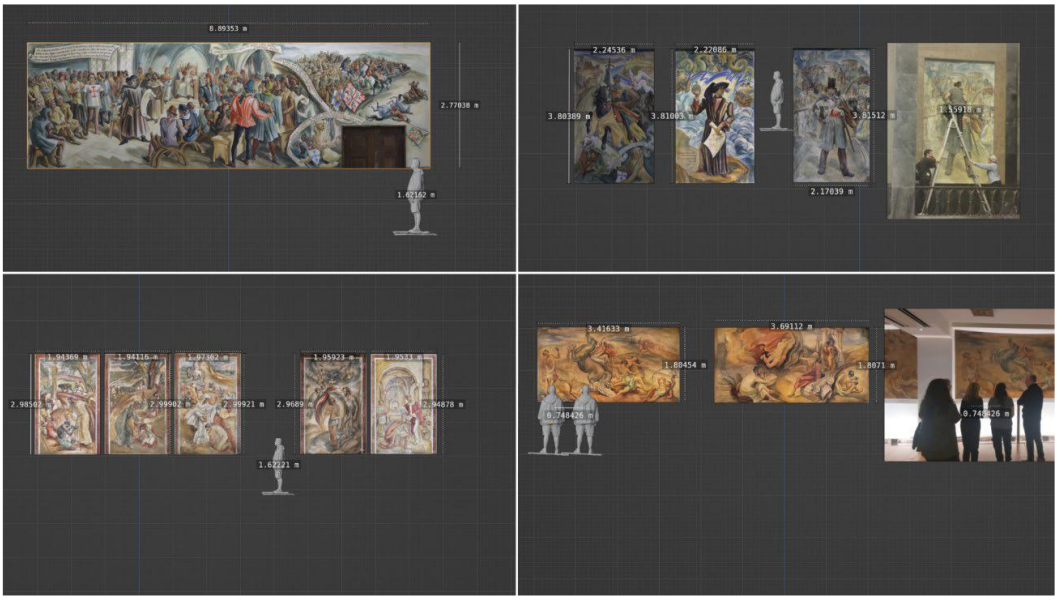
El diseño metodológico combina técnicas de documentación tridimensional de alta precisión con análisis computacional multiescalar y criterios museográficos de proyección narrativa. La investigación se enmarca en el proyecto *Lime, Pigments, and Pixels – Bringing Modern Mural Heritage Back to Public Life* (indicado anteriormente), financiado por la FCT portuguesa, y se apoya en los *Principios de Sevilla (2017)* como protocolo normativo para garantizar la fiabilidad geométrica de los modelos generados.

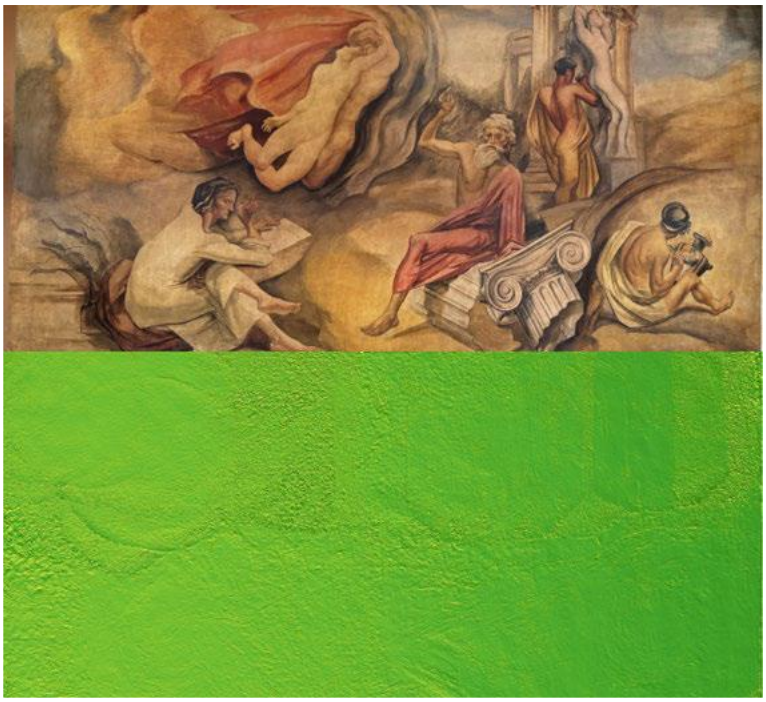
La fase inicial consistió en la toma de datos mediante fotografía réflex (Canon EOS1300D) en formato RAW y JPEG, con valores calibrados para capturar tanto texturas como colores con la mayor fidelidad posible. La captura

[2] Nubes de puntos generadas con Agisoft Metashape o Capturing Reality, análisis de coordenadas y parámetros RGB derivadas de la misma; ¿podemos ir más allá de la contemplación visual de estos elementos como soportes de información?

se adaptó a la escala y geometría de cada mural, utilizando pértigas, trípodes y disparadores remotos. En total, se generaron estereopares con solape del 60–70 %, ajustados a diferentes ángulos y alturas, lo que permitió cubrir con precisión murales de gran formato en entornos complejos. Los datos fueron procesados en *Capturing Reality,* donde se generaron nubes de puntos, mallas de alta resolución y texturas realistas que luego se optimizaron con *Instant Meshes* y *Blender* (bakeado de texturas).

Para validar la escala de los modelos, se contrastaron las medidas obtenidas con las dimensiones registradas por Laura Castro. Además, se usaron modelos 3D de escala humana para verificar la proporcionalidad espacial. Este procedimiento garantizó la precisión milimétrica de las reconstrucciones, esenciales tanto para análisis técnicos como para su uso museográfico.





A partir de un modelo de control (un panel de 1x1 m al fresco, del corpus de Baco)[3], se desarrolló un análisis microtopográfico basado en una muestra de 50.000 puntos. Empleando herramientas como CloudCompare, QGIS y librerías de Python (OpenCV, Numpy, Scikit-learn, XGBoost, etc.) se analizó la distribución de elevaciones, texturas y colores. Gracias a ello, se pudieron extraer los datos de coordenadas XYZ y los códigos HEX[4] relacionados con los puntos de la nuve. Se normalizó la altura (Z) y se aplicaron técnicas de clusterización K-Means y DBSCAN, para agrupar patrones cromáticos y topográficos. Posteriormente, se asignaron pigmentos históricos probables (87,23 % de éxito) mediante clasificación automatizada. Asimismo, se generaron mapas de vulnerabilidad pigmentaria, interpolaciones de elevación y visualizaciones 3D del riesgo de deterioro mural.

Todo esto, nos permitió construir un índice de riesgo tridimensional, relacionando vulnerabilidad cromática, relieve superficial y desgaste potencial. Una información que, sin duda, se traduce en visualizaciones ampliamente útiles para la toma de decisiones preventivas respecto de los murales del artista que aquí estudiamos.

Los propios modelos 3D generados con combinación de software especializado (indicado en párrafos precedentes) han sido preparados para su proyección a escala real en 16K en la exposición final del proyecto en abril de 2026. Esta tendrá lugar en el pabellón de la Faculdade de Belas Artes da Universidade do Porto, frente al mural de Prometeo (donde Dordio Gomes fue profesor). Además, las versiones en 1080p están utilizándose por parte de los

[3] Este se pudo digitalizar en febrero de 2025, ya que el panel se encontraba en el domicilio del nieto del artista.

[4] Código que permite a una web poner un color exacto, ya que indica a la pantalla cuánto color mostrar.

alumnos del Máster en Imagen y Sonido de la Universidad Católica de Portugal, para desarrollar animaciones experimentales sobre los murales digitalizados, inspiradas en obras como *Le Déjeuner sur l'Herbe (2011)* o *The Battle of San Romano (2017).* Esto nos demuestra que el gemelo digital, más allá de ser archivo, deviene herramienta crítica y creativa para el diseño de nuevos formatos de activación curatorial y patrimonial.

Resultados (hasta la fecha)

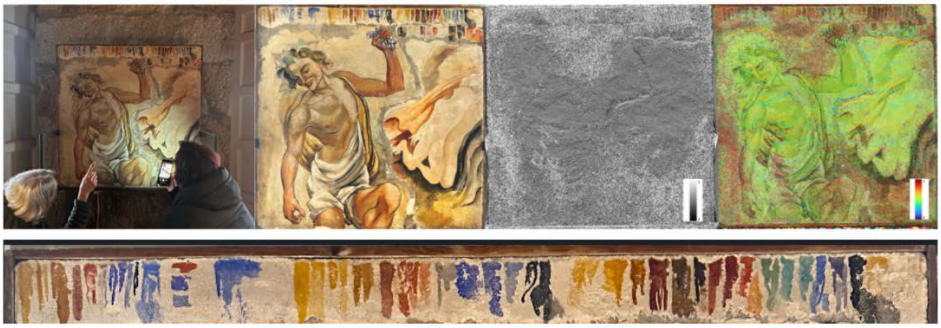
Podemos agrupar los resultados del proyecto en tres grandes bloques que incluyen documentación técnica, análisis conservativo y producción curatorial.

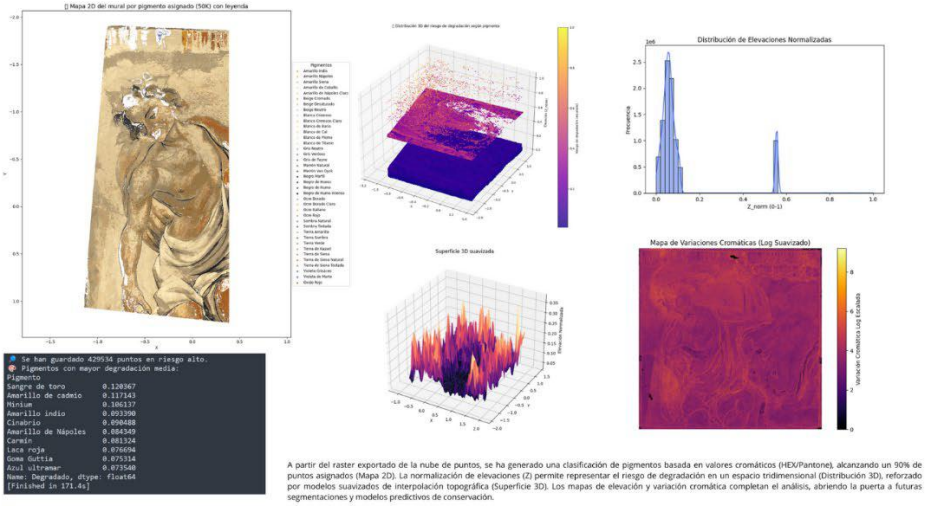
Bloque 1. Documentación técnica de alta resolución: se digitalizaron catorce murales con una precisión de 183 µm/píxel, incluyendo representaciones que ya no son accesibles físicamente (como los del Café Rialto, actual BCP Millennium). Obtuvimos modelos 3D texturizados y optimizados para navegación online e integración museográfica y se diseñó un flujo de trabajo reproducible y escalable, adaptable a otros conjuntos murales.

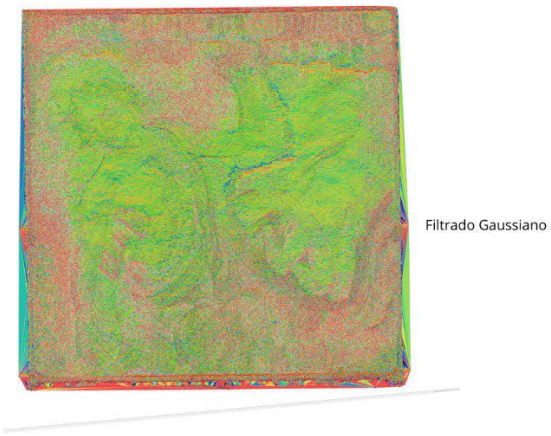
Bloque 2. Diagnóstico conservativo y análisis predictivo: se han podido detectar costuras, microelevaciones, barnices alterantes y zonas restauradas. Este análisis controlado del panel 1x1 está demostrado ser una herramienta fiable para correlacionar color, pigmento y estado de conservación, permitiendo extrapolar métodos de análisis a gran escala.

Se han generado mapas de degradación cromática y morfología superficial (a través del análisis de una pequeña muestra que luego puede ser ampliable con más medios, como el super ordenador Picasso de la Universidad de Málaga) y se ha construido un sistema de alertas visuales que permite identificar áreas críticas sin intervención directa (a trabajar con Patrícia Monteiro, restauradora profesional y miembro del equipo, encargada del análisis pigmentario de las obras de Gomes, desde una perspectiva tradicional). Esto permite trabajar la interpretación algorítmica computacional (intangible) vs. Intervención directa material (tangible). Podemos decir que esta combinación entre análisis físico y

químico y modelado computacional refuerza la multidimensionalidad del planteamiento que ofrece el proyecto.



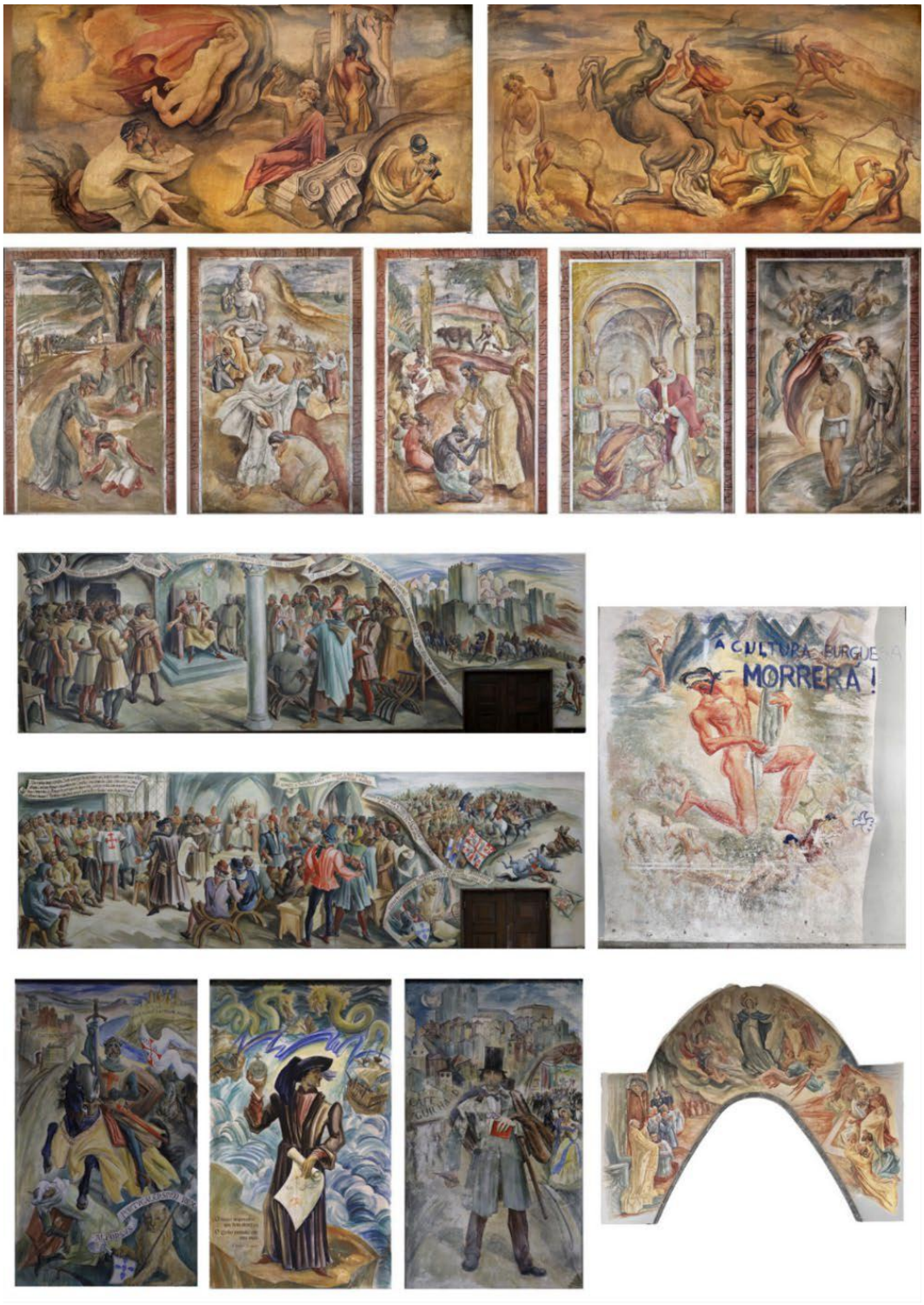




Bloque 3. Producción narrativa, curatorial y museográfica: los modelos (renders en HQ) se integrarán en una exposición inmersiva dividida en zonas temáticas (contextual iconografía, técnica, digitalización) que busca activar la memoria patrimonial. Tanto los renders como las diferentes visualizaciones obtenidas se utilizarán para diseñar disposiciones en sala, adaptadas ergonómicamente al espacio del que disponemos.

No se trata solo de un dispositivo académico-científico para “conocedores” de la materia, sino que se trabaja con estrategias educativas que tienen un fin

pedagógico e identitario. Lo digital aquí no se plantea solo como una reproducción visual neutra, sino como dispositivo narrativo, documental y proyectivo, capaz de preservar y reactivar el valor estético, técnico y cultural de los frescos de Dordio Gomes en consonancia con el muralismo contemporáneo y los modos curatoriales del siglo XXI.



Conclusión

Lime, Pigments, and Pixels. Bringing Modern Mural Heritage Back to Public Life: A Critical Approach to Computational and Digital Tools for Art Historical Curatorship trata de demostrar cómo las tecnologías de documentación tridimensional pueden trascender su función técnica para convertirse en vectores de conocimiento crítico, preservación, divulgación y activación cultural del patrimonio mural. Más allá del valor documental que pueda ofrecer, este proyecto apuesta por una curaduría digital/computacional situada, que reconoce los murales como expresión artística y como dispositivo político, social e identitario. La digitalización se consolida aquí como un acto de relectura crítica del patrimonio mural en riesgo, integrando datos técnicos, memoria visual y mediación cultural en un mismo proceso.

Bibliografía

CASTRO, L. (2021), «Obras e projectos para edifícios públicos», en *Dordio Gomes (1890–1976)* (ed. L. Castro), Porto, Árvore – Cooperativa de Actividades Artísticas, pp. 129–184.

CRESPILLO MARI, L. (2023), «Registrando intangibles lumínicos. Desafíos museográficos desde la renovación catalográfica inmersiva», *Arbor*, 199(810), a735–a735. https://doi.org/10.3989/arbor.2023.810015

FINANCE, L., & MONFORT, M. (2013), *In Situ. Revue des patrimoines*, n.º 22: «La peinture murale: héritage et renouveau», octubre. Disponible en: http://journals.openedition.org/insitu/10812

FISCHER, F., MANTONAN, D., & TRAMELLI, B. (2023), «Building relations and enhanced relationality as the backbone of methodologies in the digital and public humanities», *Magazén*, 4(1), pp. 7–12. https://edizionicafoscari.unive.it/media/pdf/article/magazen/2023/1/art-10.30687-mag-2724-3923-2023-07-000.pdf

JEFFERIES, L. (2023), «How was it for you? Qualitative observational methods to assess the experiential value of two immersive augmented reality stories, a case study», *Presence: Virtual and Augmented Reality*, 30(2), pp. 125–148. https://doi.org/10.1162/pres_a_00368

LOBO, P. R., & BRITO ALVES, M. (2017), «Espaço, fotografia e "factografia" na propaganda do SPN», *Comunicação Pública*, 12(23). https://doi.org/10.4000/cp.1287

TORRAS, B. F. (2016), «(No) Laughing Matter: Modernism and Xavier Nogués' cartoons», *RIHA Journal*, n.º 0134, julio. https://www.riha-journal.org/articles/2016/0131-0140-special-issue-southern-modernisms/0134-farre-torras

TORRAS, B. F. (2020), «Noucentisme, pintura mural i debats de l'art modern a Europa», en J. FALGÀS CASANOVAS & J. M. PUIGVERT I SOLÀ (eds.),

Noucentisme: el Noucentisme català en el context cultural europeu, Girona, Universitat de Girona / Fundació Rafael Masó, pp. 105–119.

TORRAS, B.; CRESPILLO MARI, L.; & SOARES, M. (2024), «Post wow, is less more? A critical approach to animated mapped projection for art historical knowledge sharing. The twentieth-century mural as a case study», *Magazén*, 5(1), pp. 149–172. https://doi.org/10.30687/mag/2724-3923/2024/01/006

Tecnologías avanzadas para la conservación del patrimonio cultural: ChemiNova y HBIM como nuevos Sistemas de Gestión de Contenido para bienes culturales

Esperanza Villuendas Ferri
Universitat de València

Palabras clave: Sistemas de Gestión de Contenido; Innovación; Conservación; Patrimonio cultural; Humanidades digitales
Keywords: Content Management Systems; Innovation; Conservation; Cultural Heritage; Digital Humanities

La protección del Patrimonio Cultural (PC) es una obligación de las generaciones presentes. Para lograr esta protección, es fundamental el conocimiento y registro exhaustivo de los bienes, lo que se realiza a través de inventarios y catálogos. Estos instrumentos son el punto de partida para cualquier acción de conservación (Alba Pagán, E., 2014).

Sin embargo, gracias a la irrupción de las TIC (Tecnologías de la Información y la Comunicación) en el ámbito del PC, los inventarios y catálogos deben de regirse por una nueva idiosincracia digital. Los Sistemas de Gestión de Contenido (SGC) permiten la organización, el almacenamiento y la preservación de grandes volúmenes de datos sobre bienes culturales. Al integrar funciones de inventario y catalogación, los SGC van más allá de un simple listado, ofreciendo un seguimiento del objeto patrimonial desde su adquisición, catalogación y referencias documentales y bibliográficas, hasta información sobre su estado de conservación, tratamientos de restauración y movimientos, tanto dentro como fuera de un centro de custodia. La implementación de SGC es crucial para la conservación preventiva del patrimonio, ya que facilitan la monitorización continua del estado de los bienes. Esta capacidad de seguimiento detallado permite identificar patrones de deterioro o posibles riesgos, lo que a su vez posibilita la toma de decisiones informadas para aplicar estrategias de conservación de manera proactiva. Además, la digitalización de estos registros en SGC centralizados, como el proyecto CER.es, mejora significativamente la accesibilidad y la gestión colaborativa del patrimonio cultural, promoviendo una comprensión integral de su valor y facilitando su preservación a largo plazo.

Con todo, las TIC siguen evolucionando y revolucionando nuestra comprensión del PC al proporcionar herramientas avanzadas para su adquisición digital, almacenamiento, conservación, reconstrucción y representación antes no planteadas en los SGC. La integración de las TIC facilita la creación de experiencias inmersivas, lo que permite a los usuarios interactuar

de formas novedosas. Los museos, antes vistos como instituciones conservadoras, están evolucionando hacia centros de intercambio digital y diálogo intergeneracional, priorizando la accesibilidad cultural para todos (Portalés, C., 2018). El auge de las tecnologías interactivas, como la realidad aumentada (RA) y la realidad virtual (RV), permite reubicar y recontextualizar los bienes del PC, mejorando la comprensión y apreciación del usuario, presentándose como nuevas formas no sólo de difundir y poner en valor, sino como herramientas para la gestión y documentación.

Como historiadora del arte y conservadora de museo, la reflexión sobre la importancia y necesidad de disponer de recursos digitales que agilicen y ayuden a preservar la conservación de las colecciones histórico-artísticas de un museo, me parece cada vez más prioritaria. El historiador del arte debe permanecer en los procesos tecnológicos que configuran la digitalización del Patrimonio Histórico-Artístico, no como personal colaborador, como estamos acostumbrados, sino como agente activo del proceso. Para ello, es imprescindible realizar una simbiosis curricular dentro de ambas áreas, la Tecnológica y la de Historia del Arte. ¿Cómo la tecnología podría ayudar al historiador del arte, al catalogador, al conservador de colecciones, edificios y monumentos histórico-artísticos (perfiles que concibo como un mismo profesional)? ¿Qué herramientas digitales actuales pueden hacer que la catalogación y la conservación de estos bienes culturales sea visto como un todo inseparable y que, además, pueda difundirse para los ciudadanos?

Buscando respuestas a estas preguntas, las Humanidades digitales (HD) se presentan como la disciplina con la cual indagar, profundizar y transformar en humanidades, y a la vez aplicar métodos computacionales y abordar el impacto de las prácticas digitales (Schwandt, 2020). En mi caso concreto, mediante una investigación que surge desde preguntas propias de la Historia del Arte, y aplica tecnologías avanzadas para la conservación del patrimonio cultural. El estudio realiza una revisión sistemática que abarca desde los primeros inventarios y catálogos de Bienes de Interés Cultural (BIC), la digitalización de colecciones y la gestión cultural por parte de museos e instituciones por medio de los SGC hasta el tratamiento de datos propios del patrimonio cultural y el desarrollo de un Data Management Plan (DMP) en proyectos europeos de investigación. Seguidamente, evalúa la aportación de nuevos SGC de bienes culturales, con los que gestionar, catalogar y realizar labores documentales de gestión preventiva. Comparando los estándares del proyecto europeo de investigación ChemiNova, y los de la metodología HBIM, se propone un marco práctico para la aplicación de estos sistemas en futuras iniciativas de gestión documental y conservación preventiva patrimonial (Wang, W., 2024); y también como un marco teórico de investigación histórico-artística.

El estudio, además, pone atención en la evolución histórica de la profesionalización en Historia del Arte, desde la época “analógica” hasta la era digital, culminando en la formación interdisciplinaria y tecnológica. Esta

perspectiva permite a historiadores del arte liderar las labores de aplicación de tecnologías digitales avanzadas.

La metodología adoptada en el presente estudio se articula en torno a un enfoque multidisciplinar que combina el análisis teórico con la observación empírica. En primer lugar, se lleva a cabo una exhaustiva revisión bibliográfica y documental, centrada en el estudio de inventarios y catálogos de bienes culturales, los procesos de digitalización en museos, los sistemas de gestión de contenido (SGC) y el tratamiento de datos patrimoniales en el contexto de proyectos europeos de investigación. Esta fase permite establecer un marco teórico sólido y actualizado sobre las prácticas contemporáneas en la gestión del patrimonio cultural. Posteriormente, se desarrolla un estudio de casos que incluya el análisis detallado de los sistemas HBIM (Heritage Building Information Modeling) y ChemiNova, evaluando su aplicabilidad en la conservación y gestión documental del patrimonio. Se examinan sus procesos operativos, beneficios potenciales y limitaciones técnicas, con el fin de valorar su eficacia en contextos reales de intervención patrimonial (Yang, X., 2020).

En este sentido, la Unión Europea y los principales organismos en materia de conservación del Patrimonio Cultural europeo, como el ICCROM, llevan varios años promoviendo el uso de la metodología HBIM por su capacidad para transformar la conservación del Patrimonio Cultural mediante la integración de tecnologías digitales. Además, este plan ha demostrado ser fructífero en investigaciones como las llevadas a cabo por la Dra. Mónica Giuliano de los Musei Vaticani sobre "Applicazione della tecnologia 3D nei Musei Vaticani: Vantaggi e criticità", el Dr. Livio de Luca del CNRS-MAP de París con "Nôtre-Dame de Paris: a cathedral of digital data and multidisciplinary knowledge in heritage science", y el Dr. Atsushi Noguchi de la Komatsu University de Japón con "Recent Advancements in the Digitalization and HBIM Application of Wooden Structures", las cuales son referencias ineludibles para este estudio.

Por su parte, el proyecto “ChemiNova - Novel technologies for on-site and remote collaborative enriched monitoring to detect structural and chemical damages in cultural heritage assets” cuenta con el financiamiento del programa Horizonte Europa. El proyecto está coordinado desde la Universitat de València con un enfoque interdisciplinar dual: desde el Departamento de Historia del Arte, encabezado por la Doctora Ester Alba Pagán; y desde el Instituto de Robótica y Tecnologías de la Información y las Telecomunicaciones (IRTIC) por la Doctora Cristina Portalés Ricart. El consorcio está compuesto por doce socios, y propone una solución integral para la conservación preventiva del PC mediante la creación de modelos e3D y el uso de inteligencia artificial, sensores portátiles, realidad aumentada y plataformas colaborativas. Su objetivo no es únicamente mejorar la detección y evaluación del deterioro, sino también fomentar la implicación activa de las comunidades locales en la protección del patrimonio (Cordis, 2023).

Complementariamente, se realizarán entrevistas semiestructuradas y cuestionarios dirigidos a expertos en tecnologías avanzadas aplicadas a la

conservación del PC y a las HD. Esta estrategia permitirá recoger perspectivas cualificadas sobre la evolución del rol del historiador del arte en un entorno cada vez más digitalizado y tecnificado.

También se llevará a cabo un ejercicio de comparación crítica entre los sistemas tradicionales de gestión documental y las soluciones tecnológicas avanzadas, con el objetivo de identificar sus respectivas contribuciones a la conservación preventiva y a la eficiencia en la gestión de la información patrimonial. Esta comparación servirá de base para la formulación de propuestas orientadas a la implementación efectiva de tecnologías digitales en el ámbito patrimonial, subrayando el papel emergente del historiador del arte como técnico especializado y humanista digital.

Finalmente, el estudio aspira a consolidar un marco teórico que contribuya al desarrollo de una línea de investigación histórico-artística centrada en la digitalización patrimonial, al tiempo que aspira, a su vez, a proponer un modelo práctico para la aplicación de estas tecnologías en futuras iniciativas de documentación y conservación del PC.

Bibliografía

ALBA PAGÁN, E. (2014), «Catálogo e inventario como instrumentos para la gestión del patrimonio cultural», en *Actas de la Jornada "Educació i formació en torn al patrimoni cultural de l'interior"*, recurso en línea. https://roderic.uv.es/rest/api/core/bitstreams/c9a1ae08-2a28-4f79-a607-537ba3a7d6e9/content

CORDIS (2023, 10 de noviembre), «Novel technologies for on-site and remote collaborative enriched monitoring to detect structural and chemical damages in cultural heritage assets», *CORDIS | European Commission*, recurso en línea. https://cordis.europa.eu/project/id/101132442

PORTALÉS, C., RODRIGUES, J. M. F., RODRIGUES GONÇALVES, A., ALBA, E., & SEBASTIÁN, J. (2018), «Digital Cultural Heritage», *Multimodal Technologies and Interaction*, 2(3), art. 58. https://doi.org/10.3390/mti2030058

SCHWANDT, S. (2020), *Digital Methods in the Humanities: Challenges, Ideas, Perspectives*, recurso en línea. https://library.oapen.org/handle/20.500.12657/47133

WANG, W., ZAHEER, Q., QIU, S., WANG, W., AI, C., WANG, J., WANG, S., & HU, W. (2024), *Digital Twin Technologies in Transportation Infrastructure Management* (1.ª ed. 2024), Singapore, Springer Nature Singapore. https://doi.org/10.1007/978-981-99-5804-7

YANG, X., GRUSSENMEYER, P., KOEHL, M., MACHER, H., MURTIYOSO, A., & LANDES, T. (2020), «Review of built heritage modelling: Integration of HBIM and other information techniques», *Journal of Cultural Heritage*, 46, pp. 350–360. https://doi.org/10.1016/j.culher.2020.05.008

Realidad Aumentada como herramienta participativa para la activación del patrimonio: el caso de TinajAR

Bibiana Martínez Torrecilla
Lucía Vera Mahiques
Cristina Portalés Ricart
Universitat de València
Carlos Martínez Barragán
Universitat Politécnica de València

Palabras clave: Realidad aumentada; patrimonio inmaterial; instalación inmersiva; memoria colectiva; humanidades digitales
Keywords: Augmented reality; intangible heritage; immersive installation; collective memory; digital humanities

El auge de las tecnologías inmersivas ha transformado los modos de producción, transmisión y conexión con la memoria cultural, abriendo nuevas posibilidades de participación con los bienes patrimoniales, tanto materiales como inmateriales. La incorporación progresiva de entornos digitales en museos y espacios culturales ha ampliado las formas de interacción, aunque persiste el desafío de diseñar dispositivos que, más allá de la espectacularidad tecnológica, logren activar memorias colectivas, establecer vínculos con el territorio y generar experiencias accesibles.

En este contexto, la Realidad Aumentada (RA) se presenta como una herramienta adecuada para dinamizar estos procesos, actuando como dispositivo de visualización, interacción y contextualización capaz de establecer diálogos entre pasado y presente. La RA es una tecnología que permite superponer información digital —imágenes, sonidos, animaciones, vídeos o datos— sobre el entorno físico en tiempo real, enriqueciendo así la percepción del mundo tangible mediante capas virtuales (Azuma, 1997). En el ámbito del patrimonio cultural, esta tecnología ha demostrado su capacidad para activar nuevas formas interacción y resignificación del pasado.

Más allá de la superposición visual, la RA puede generar experiencias que apelan al cuerpo y a la memoria, provocando una implicación afectiva del visitante. En este sentido, se trata de crear situaciones que inviten a recordar o imaginar desde el presente. Así, la RA contribuye a construir relatos ubicados, donde la tecnología no eclipsa al patrimonio, sino que lo reactiva y lo traduce. Por lo tanto, esta tecnología ofrece la posibilidad de experimentar lo patrimonial como una forma viva de conocimiento, en diálogo con los sentidos, el territorio y la comunidad.

Este trabajo parte, por tanto, de una cuestión central: ¿cómo pueden las características definitorias de la RA transformar la conexión con los bienes culturales en prácticas participativas con la memoria histórica?

Para abordar esta cuestión se propone el análisis de TinajAR, una instalación interactiva de RA que vincula el patrimonio inmaterial de la alfarería tradicional de un pueblo, con los lenguajes digitales contemporáneos. La propuesta ofrece una nueva forma de relación con el barro en la que los protagonistas son los propios visitantes y una serie de modelos tridimensionales basados en piezas de patrimonio alfarero riojano, documentados previamente. Este proyecto ejemplifica la convergencia entre cultura material y medios digitales, articulando memoria colectiva y experiencia participativa en un dispositivo que contribuye al reconocimiento y visibilización de un legado en riesgo.

Asimismo, TinajAR recupera la experiencia conceptual de Sambori AR, una instalación de RA -inspirada en la rayuela y presentada en la convocatoria Digital Media 1.0. ubicada en La Nau de Valencia- que activaba la memoria colectiva mediante una interacción accesible, intuitiva y libre (Martínez, 2008). TinajAR actualiza y amplía estos principios, incorporando portabilidad y capacidad de adaptación a distintos espacios históricos, reforzando así su carácter itinerante y su potencial para generar comunidad.

Desde un punto de vista técnico, TinajAR se fundamenta en el paradigma del espejo aumentado (Portalés, 2016), combinando la imagen en tiempo real del espectador con contenidos digitales superpuestos. Este modelo de proyección favorece la participación espontánea y emocional, al permitir que los usuarios se reconozcan dentro de la escena aumentada e interactúen de forma intuitiva con ella, tal y como se muestra en el esquema técnico de la Figura 1.

La aplicación, desarrollada en el Grupo de Investigación ARTEC Gráficos 3D para Tiempo Real de la Universidad de Valencia, ha sido implementada en varios calados riojanos del siglo XVII, lo que ha permitido validar su capacidad para conectar con entornos patrimoniales reales. Esta experiencia demuestra cómo las tecnologías inmersivas, y en particular la RA, pueden operar como dispositivos de memoria expandida y como medios para construir relatos colectivos desde lo digital, contribuyendo tanto a la preservación de saberes en riesgo como al fortalecimiento de la memoria cultural local.

Partiendo de la investigación realizada por Ruiz Torres en su tesis *La realidad aumentada y su aplicación en el patrimonio cultural*, en este caso se ha optado por un enfoque metodológico diferente, basado en la experimentación artística. Se recurre a la práctica como forma de producción de conocimiento, utilizando la observación participante como método principal para documentar las interacciones generadas por la instalación, junto con la recogida de testimonios orales de artesanos, visitantes y agentes culturales implicados.

Esta aproximación cualitativa permite explorar no solo las dinámicas participativas y afectivas que emergen en torno a la experiencia aumentada, sino también su capacidad para activar memorias personales y colectivas en diálogo

con los objetos patrimoniales. La metodología adoptada se centra, por tanto, en la observación directa de las dinámicas participativas in situ y en la realización de entrevistas a alfareros tras su interacción con la instalación.

Si bien en trabajos previos se desarrolló un análisis cuantitativo de la experiencia de uso —centrado en métricas de interacción, tiempos de permanencia y perfiles de usuario, cuyos resultados fueron presentados en un congreso especializado—, en esta ocasión se privilegia una aproximación cualitativa, coherente con el enfoque humanístico y cultural del presente encuentro académico.

El trabajo de campo se llevó a cabo durante la presentación de *TinajAR* en la Feria Nacional de Alfarería y Cerámica N.A.C.E., celebrada en la localidad riojana de Navarrete, específicamente en el Calado del Conde. La elección de este enclave histórico favoreció la conexión entre los contenidos virtuales y el entorno físico, facilitando que los visitantes establecieran un diálogo emocional y contextualizado con la experiencia, como se documenta en la fotografía de la Figura número 2.

Los resultados obtenidos evidencian que la RA, junto con sus componentes sonoros, efectos visuales y dinámicas interactivas, posibilita la activación de memorias y fomenta la participación con los bienes culturales. Asimismo, se constató que la RA actúa como un mediador simbólico, facilitando la actualización de relatos comunitarios y promoviendo la conexión intergeneracional entre diversos perfiles de visitantes. Destaca la accesibilidad de la instalación, que permitió la participación inclusiva de personas con perro guía, visitantes con movilidad reducida y jóvenes familiarizados con entornos de videojuegos, quienes se aproximaron a la propuesta desde una experiencia lúdica. Como se observa en la Figura 3, la pieza digital *TinajAR* atrajo a familias, grupos de amigos y conjuntos de personas que activaban simultáneamente la aplicación e interactuaban entre sí.

El análisis de las interacciones reveló que la combinación de imagen reflejada, contenido virtual y estímulos sonoros generó un efecto inmersivo que potenció la implicación emocional de los espectadores. En este sentido, *TinajAR* facilitó la construcción de un espacio de memoria expandida, capaz de configurar lo que en los estudios de interacción digital se reconoce como un entorno de interacción amigable, en el que lo físico y lo digital se integran de forma armónica. Este tipo de entornos favorece contextos afectivos y relacionales, donde los sujetos se sienten acogidos, reconocidos en la escena aumentada y libres para explorar desde su propia experiencia cultural, interpretando los objetos patrimoniales y actualizando sus relatos desde una mirada situada y participativa.

Desde esta perspectiva, la RA se presenta como un dispositivo sociocultural capaz de resignificar el espacio, el cuerpo y la memoria mediante prácticas participativas que enriquecen los estudios culturales actuales. Cuando la RA se vincula con el patrimonio, activa emociones a través de experiencias lúdicas e inmersivas, generando condiciones propicias para que el público se

involucre activamente. Esta implicación afectiva favorece la creación de comunidad en torno a la experiencia digital. Como puede observarse en la Figura 4, la instalación fue también presentada en la *EXPOCIENCIA*, jornada de puertas abiertas que cada mayo se celebra en el Parc Científic de la Universitat de València.

Desde una perspectiva teórica, este proyecto se inscribe en el debate actual de las humanidades digitales como campo de producción cultural y crítica social. Más allá de su dimensión analítica, se reivindica la capacidad de las tecnologías inmersivas para generar dispositivos digitales que actúan como agentes de cambio cultural, posibilitando nuevas formas de participación y reinterpretación del patrimonio.

TinajAR evidencia cómo estas tecnologías pueden resignificar bienes inmateriales, activar memorias colectivas y proponer modelos alternativos de comunicación cultural, articulados desde la accesibilidad, la participación multigeneracional y la experiencia sensorial. En síntesis, este estudio destaca cómo la RA permite establecer vínculos entre pasado y presente, tradición y tecnología, transformando la relación con el legado cultural en una experiencia emocionalmente conectada.

Asimismo, se subraya la importancia de conformar equipos transdisciplinares que integren saberes, técnicos y humanísticos, capaces de generar propuestas sensibles a las dimensiones culturales, emocionales y sociales de los contextos patrimoniales. Finalmente, este proyecto invita a consolidar la interacción entre arte, ciencia y comunidad como eje estratégico para las Humanidades Digitales, proponiendo nuevas formas de transmisión cultural y activación social del patrimonio a través de tecnologías inmersivas.

En conclusión, la tecnología de Realidad Aumentada, aplicada al ámbito patrimonial, se configura como un conjunto de dinámicas interactivas que movilizan al público en los planos físico, intelectual y social (Gimeno, 2011). A través de interfaces intuitivas y narrativas aumentadas, la RA permite a los usuarios desplazarse por el espacio, interactuar con contenidos virtuales y descubrir capas de información cultural que enriquecen y amplifican el discurso expositivo. Al mismo tiempo, fomenta la colaboración, el diálogo y la participación entre los visitantes, convirtiendo el recorrido en una experiencia compartida que favorece la apropiación emocional y social del patrimonio.

En este contexto, aplicaciones como TinajAR se consolidan como dispositivos de mediación cultural, capaces de dejar una impresión duradera en el público y ampliar sus capacidades de interpretación, interacción y participación en torno al patrimonio.

Figuras, tablas, mapas o gráficas

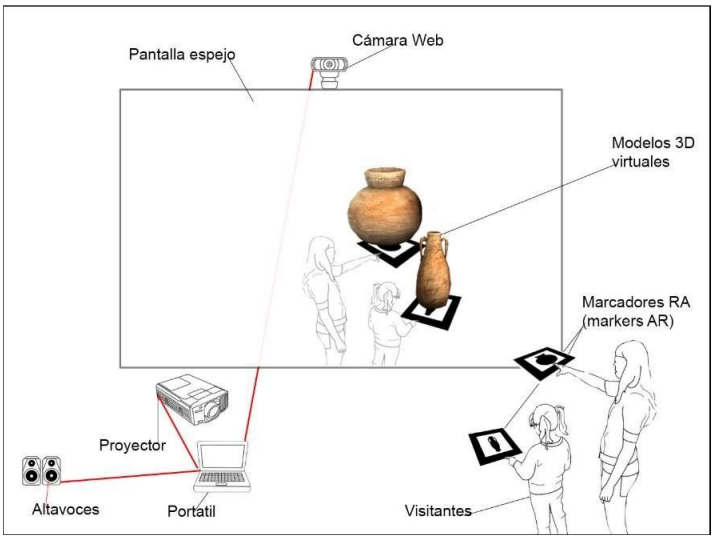


Figura 1. Esquema técnico del funcionamiento de TinajAR. Disposición de los elementos de hardware y marcadores de RA para generar la experiencia inmersiva participativa.

Figura 2. Visitantes interactuando con las piezas virtuales en el Calado del Conde.

Figura 3. Visitantes interactuando con las piezas virtuales en el Calado del Conde

Figura 4. Visitantes interactuando con las piezas virtuales en la EXPOCIENCIA

Bibliografía

AZUMA, R. T. (1997), «A survey of augmented reality», *Presence: Teleoperators and Virtual Environments*, 6(4), pp. 355–385. https://doi.org/10.1162/pres.1997.6.4.355

GIMENO, J., OLANDA, R., MARTÍNEZ, B., & SÁNCHEZ, F. (2011), «ObservAR: Sistema de realidad aumentada multiusuario para exposiciones», *Novática*, 213, pp. 30–36.

MARTÍNEZ TORRECILLA, B., VERA MAHIQUES, L., PORTALÉS RICART, C., & MARTÍNEZ BARRAGÁN, C. (2022), «Estudio de una obra interactiva de alfarería y realidad aumentada», *Interacción: Revista de Experiencias con Tecnología y Diseño*, 3(1). https://revista.aipo.es/index.php/INTERACCION/article/view/61

MARTÍNEZ TORRECILLA, B., VERA MAHIQUES, L., COMA TATAY, I., & FERNÁNDEZ MARÍN, M. (2010), «Sambori: Nueva forma de interacción con realidad aumentada», en *Congreso Internacional Mujer, Arte y Tecnología en la Nueva Esfera Pública (CIMUAT)*, Universitat de València / IRTIC.

PORTALÉS, C., GIMENO, J., CASAS, S., OLANDA, R., & GINER, F. (2016), «Interacting with augmented reality mirrors», en I. RODRÍGUEZ & F. GARCÍA (eds.), *Handbook of Research on Emerging Technologies for Digital Preservation and Information Modeling*, cap. 9, pp. 216–244, IGI Global. https://doi.org/10.4018/978-1-5225-0680-5.ch009

RUIZ TORRES, D. (2013), *La realidad aumentada y su aplicación en el patrimonio cultural* [Tesis doctoral], Universidad de Granada. https://digibug.ugr.es/handle/10481/23720

El juez de residencia y su desempeño. Un análisis desde la representación cartográfica con SIG

Víctor González-Melgar
Universidad Complutense de Madrid

Palabras clave: Juez de residencia; juicio de residencia; SIG; cartografía digital; georreferenciación
Keywords: Judge of Residence; Court of Residence; GIS; Digital Cartography; Georeferencing

Introducción

El juicio de residencia castellano se consolidó, en la temprana modernidad, como un importante recurso de control regio a través de la rendición de cuentas de los oficiales de la Corona tras terminar su tiempo en el cargo. Su trascendencia en el derecho castellano se debió a la capacidad de supervisión que ofreció a la monarquía en un momento de marcada expansión territorial, permitiendo a la metrópolis estar al corriente –teóricamente– de lo acontecido en los dominios de ultramar por medio de la responsabilidad pública. En el presente trabajo se aspira a la construcción, desde las humanidades digitales, de una facilitación de la visualización de historia urbana y política cimentada en las residencias realizadas en La Española durante la primera mitad del siglo XVI, expresada cartográficamente para mejor comprensión y contextualización. Con ello se pretende mostrar que la unión de herramientas digitales y perspectivas humanísticas facilita la interpretación de procesos históricos complejos, a la par que impulsa la democratización del conocimiento científico de una manera accesible para el público general, sin sacrificar la rigurosidad concedida por las fuentes.

Las últimas décadas del siglo XV terminaron de definir la residencia, cuyos orígenes se remontan principalmente a las *Partidas* de Alfonso X y, a su vez, al derecho romano.[1] Concretamente, el Ordenamiento de Toledo de 1480 junto con la Pragmática de Sevilla de 1500 precisaron un modelo de juicio de residencia que, salvo contadas actualizaciones como las de la Nueva Recopilación de 1567,[2] permaneció inmutable durante toda la Edad Moderna. Se componía,

[1] José María Mariluz Urquijo, *Ensayo sobre juicios de residencia indianos* (Sevilla: Escuela de Estudios Hispano-Americanos de Sevilla, Consejo Superior de Investigaciones Científicas, 1952), pp. 6-8.
[2] Rafael Serra Ruiz, «Notas sobre el juicio de residencia en época de los Reyes Católicos», *Anuario de estudios medievales*, nº 5 (1968), pp. 531-48.

esencialmente, de dos etapas en las que la población tomaba un rol destacado. De esta manera, estaba formado por una primera fase secreta en la que era el juez de residencia quien elegía confidencial y cautelosamente a los testigos, y una segunda pública donde toda la población podía testificar en contra o a favor del residenciado.[3] Cabe destacar que no es el objetivo de este trabajo ahondar en las raíces de la residencia, ni tampoco discutir sobre la eficiencia del procedimiento como adecuado sistema de supervisión real, cuestiones sobre las que ya existe un amplio debate y consecuente producción científica. Al margen de ello, en las siguientes páginas se pondrá el foco de atención en el valor que reside en la fusión entre la tecnología y estas fuentes documentales para aproximar al humanista digital al contexto en que fueron producidas. Los juicios de residencia se distinguieron por tener un desarrollo complejo, componiendo un procedimiento interregional en el que los actores protagonistas que en ellos participaban, jueces y residenciados– formaban parte de un juego de oficios públicos y gobernación urbana, a distintas escalas, regido por la rendición de cuentas. Esta cualidad hace de la residencia una significativa fuente para el estudio de la administración de las ciudades en Castilla, pero su concepción a gran escala resulta dificultosa debido a la extensa duración que las definieron y la intrincada interrelación entre quienes tomaban parte en ellas. Estudiar cómo y dónde ejercían su oficio tanto jueces como residenciados permite observar el desarrollo de la gobernación castellana en un área específica. Es en este punto donde interviene el uso de SIG, habilitando una visión holística de la problemática que, tras una pertinente extracción de información de las fuentes, representa de manera integral las variables deseadas, favoreciendo su entendimiento gracias a lo ilustrativo, gráfico y accesible de un mapa interactivo.[4] Para el presente trabajo se ha seleccionado la figura del juez de residencia en un marco geográfico y temporal concreto: la isla de La Española durante la primera mitad del siglo XVI. Lo que permite ejemplificar la utilidad de un ejercicio perfectamente adaptable a otros valores para su reproducción flexible en contextos mutables. Este recurso es de marcada utilidad para comprender simultáneamente el desarrollo de la regencia extrapeninsular en los albores de la conquista y, principalmente, el desempeño intermitente e itinerante del juez de residencia en dicho paradigma.

[3] María José Collantes de Terán de la Hera, «El juicio de residencia en Castilla a través de la doctrina jurídica de la Edad Moderna», *Historia. Instituciones. Documentos*, n.º 25 (1998), pp. 167-177.

[4] A modo de ejemplo, consultar el siguiente artículo centrado en el análisis del sistema defensivo de Ningbo durante la dinastía Ming y su entendimiento facilitado por el uso de SIG: Yukun Zhang et al., «Information Visualization Analysis Based on Historical Data», *Multimedia Tools and Applications* 81, n.º 4 (2022), pp. 4735-51, https://doi.org/10.1007/s11042-021-11030-8.

Metodología

La propuesta planteada necesita de una combinación de trabajo de archivo, análisis de documentación y cartografía digital para su realización. Esto permitirá la extracción de información directamente del documento y su representación visual en el mapa de manera dinámica. Para ello, y dado el objeto de estudio de este escrito, se ha consultado de la sección "Justicia" del Archivo General de Indias la serie "Residencias tomadas en la Audiencia de Santo Domingo, apeladas y remitidas al Consejo de Indias". En ella se encuentran contenidas 203 unidades documentales que van de 1517 a 1578, por lo que ha sido necesario realizar una selección limitada temporalmente hasta el año de 1551. Teniendo en cuenta la naturaleza policéntrica de las fuentes, lejos de suprimir aquellas externas a La Española se han incluido para definir una "zona de actuación" en la que Santo Domingo se vincula con las demás regiones –Santiago de Cuba, San Juan de Puerto Rico, San Germán, Cubagua, Isla Margarita, Coro y Sevilla la Nueva–. Tras esto, se ha procedido a la identificación de la signatura, lugar y fecha de producción de los documentos escogidos, así como los diversos destinos de aquellos jueces que tomaran residencia en más de una ciudad, con el fin de entender la correlación entre los territorios comprendidos en la serie documental consultada. Es necesario señalar que por motivos de escala, y con el objetivo de mantener acotada la atención exclusivamente en la fase indiana del proceso, en esta investigación no se han valorado las etapas peninsulares de la residencia y se ha tomado como referencia el lugar de inicio del proceso. Esto no porque se considere que carezcan de relevancia, sino debido a que incluirlas podría comprometer la profundidad del estudio sobre la dimensión americana, e introduciría variables contextuales muy diferentes. No obstante, este acotamiento abre la puerta a posteriores investigaciones que, partiendo de los resultados obtenidos, puedan explorar con el mismo rigor las etapas peninsulares y así completar el panorama desde una perspectiva transoceánica.

Para llevar a cabo el proyecto se ha utilizado el *software* de código abierto *QGIS*. En él se ha creado una capa poligonal nombrada "Área de actuación", en la que se representa el área de interrelación entre las ciudades pertinentes. Por otro lado, han sido elaboradas dos grandes agrupaciones de capas destinadas a recoger las "Residencias" y los "Itinerarios" –plasmando el total de ciudades diferentes en las que un mismo juez ha actuado–, conteniendo cada uno atributos específicos que enriquecen el apoyo visual y asisten el seguimiento cronológico de la dinámica. Concretamente, estos atributos se corresponden con, además del obligado "ID", las ya mencionadas "fecha" y "lugar de publicación", al igual que "signatura" para el grupo "Residencias"; además de "Salida", "Intermedio" y "Llegada" para el grupo "Itinerario". Resulta significativo que aun siendo bastante reducido el número de jueces que cumplen los requisitos para entrar en este último conjunto –3 de 21 jueces–, no deja de ser un factor de extrema importancia para comprender la labor de esta figura política,

y la administración de la gobernación castellana en el Nuevo Mundo. Esto a causa de que representa claramente la naturaleza prolongada de la residencia[5] y su carácter interterritorial, abarcando localidades diversas simultánea y/o sucesivamente que obligaban al juez de residencia bien a desplazarse, bien a atender procesos concomitantes en una misma región. Al mismo tiempo dos capas topográficas base aportan el fondo cartográfico sobre el que se cimentan todos los elementos descritos, siendo satelital una de ellas. Finalmente, cabe destacar que, a fin de evitar la saturación de contenido que pudiera obstaculizar y poner en peligro el objetivo de facilitar el entendimiento de la problemática, no se ha indexado documentación derivada de, y/o auxiliar a la residencia.

Resultados

El mapa resultante ha sido subido a la nube en una base de datos de *QGIScloud* (https://qgiscloud.com/vigonz12/Residencias_Santo_Domingo/) para una superior difusión del material generado. De esta forma, cualquier usuario con conexión a internet puede acceder al recurso de manera abierta, interactuar como prefiera con las distintas capas, y obtener información sobre los atributos adjudicados pinchando en los iconos visibles. El resultado es, en síntesis, una representación gráfica de la interacción interregional entre jueces de residencia durante las primeras décadas de gobierno castellano en diferentes ciudades del caribe. La información, expresada visual e interactivamente reducen lo enrevesado de crear una imagen mental de este sistema intrincado, haciendo el acercamiento colectivo a este objeto de estudio más asequible y aproxima así procesos complejos a investigadores y al público no especializado. En este sentido, el uso de SIG aporta una dimensión del conocimiento que con dificultad es alcanzable sin su empleo, posibilitando la realización de observaciones desde perspectivas distintas y formular nuevos interrogantes. Por ejemplo, permite el seguimiento de los actores implicados en el proceso de residencia –tanto jueces como residenciados– para analizar si existe una evolución en el modelo de aplicación del procedimiento que se relacione con las rutas recorridas por estos actores. Argumento que se seguirá en una fase posterior de la investigación fuera de La Española, para observar cómo se relacionan y retroalimentan las experiencias de las residencias en diferentes territorios americanos e ibéricos, y el papel que juegan estos actores que se trasladan entre territorios en el transcurso de la evolución del procedimiento. Panorama donde es igual de interesante la posible concepción unitaria de distintos territorios derivada de la celebración de residencias que abarcan más de uno como partes integrantes de

[5] Característica que ya para los contemporáneos componía un auténtico problema, muchas veces relacionado con prácticas corruptas, como bien lo expresaba Castillo de Bovadilla en su *Política*. (Jerónimo Castillo de Bovadilla y Benjamín González Alonso, *Política para corregidores y señores de vassallos, en tiempo de paz, y de guerra. Y para juezes eclesiasticos y seglares y de Sacas, Aduanas, y de Residencias, y sus Oficiales: y para Regidores, y Abogados, y del valor de los Corregimientos, y Goviernos Realengos, y de las Ordenes.*, ed. Sebastián Martín-Retortillo, Edición facsímil, vol. II (Instituto de Estudios de Administración Local, 1978), pp. 417).

una misma residencia. Las competencias de esta tecnología invitan a reflexionar sobre el potencial que un ejercicio menos limitado puede tener para la comprensión más vasta a escalas transatlánticas y profundas. Cuestiones que serán objeto de exploración en un marco de tesis doctoral, con este escrito como base inicial.

Bibliografía

CASTILLO DE BOVADILLA, J., & GONZÁLEZ ALONSO, B. (1978), *Política para corregidores y señores de vassallos, en tiempo de paz, y de guerra. Y para juezes eclesiasticos y seglares y de Sacas, Aduanas, y de Residencias, y sus Oficiales: y para Regidores, y Abogados, y del valor de los Corregimientos, y Goviernos Realengos, y de las Ordenes*, ed. facsímil, vol. II, ed. de S. Martín-Retortillo, Instituto de Estudios de Administración Local.

COLLANTES DE TERÁN DE LA HERA, M. J. (1998), «El juicio de residencia en Castilla a través de la doctrina jurídica de la Edad Moderna», *Historia. Instituciones. Documentos*, n.º 25, pp. 151–184.

MARILUZ URQUIJO, J. M. (1952), *Ensayo sobre juicios de residencia indianos*, Sevilla, Escuela de Estudios Hispano-Americanos de Sevilla / Consejo Superior de Investigaciones Científicas.

SERRA RUIZ, R. (1968), «Notas sobre el juicio de residencia en época de los Reyes Católicos», *Anuario de Estudios Medievales*, n.º 5, pp. 531–548.

ZHANG, Y., WU, B., TAN, L., & LIU, J. (2022), «Information visualization analysis based on historical data», *Multimedia Tools and Applications*, 81(4), pp. 4735–4751. https://doi.org/10.1007/s11042-021-11030-8

Uso de los proyectos Wikimedia en instituciones y proyectos de patrimonio cultural

Gustavo Candela
Universidad de Alicante – Wikimedia España
Ángel Obregón Sierra
Universidad Isabel I – Wikimedia España

Palabras clave: Wikidata; Patrimonio cultural; Enriquecimiento de datos; Colecciones de datos
Keywords: Wikidata; Cultural Heritage; Data Enrichment; Data Collections

Introducción

En las últimas décadas, las instituciones de patrimonio cultural han adoptado estrategias de digitalización para facilitar el acceso abierto a sus colecciones. Este proceso ha venido acompañado por un creciente interés en la reutilización de los recursos digitales y por su visibilidad en entornos colaborativos y accesibles. En este contexto, los proyectos de la Fundación Wikimedia —como Wikipedia, Wikimedia Commons y, especialmente, Wikidata— han emergido como herramientas útiles para vincular, enriquecer y difundir los contenidos custodiados por archivos, bibliotecas y museos. Estas plataformas ofrecen la oportunidad de conectar datos, mejorar la interoperabilidad y fomentar nuevas formas de participación ciudadana y académica [Candela, 2024].

Este trabajo tiene como objetivo identificar y analizar distintas iniciativas desarrolladas por instituciones de patrimonio cultural en el ámbito nacional que integran Wikidata en sus procesos de difusión y gestión de colecciones digitales. Se pretende ofrecer una panorámica actualizada sobre el grado de adopción y los usos concretos de Wikidata en este tipo de instituciones.

Trabajo previo

En los últimos años, numerosas iniciativas han explorado el potencial de los proyectos Wikimedia para incrementar la visibilidad y fomentar la reutilización de las colecciones digitales en el ámbito del patrimonio cultural [Candela, 2024; Padilla, 2023]. Algunas bibliotecas y repositorios digitales, como el proyecto Mnemosine o la Biblioteca Virtual Miguel de Cervantes (BVMC), han desarrollado

páginas dedicadas en Wikipedia donde difunden información contextual sobre sus fondos.

Otras iniciativas han utilizado Wikidata para enriquecer las colecciones digitales. Esta plataforma colaborativa permite describir entidades culturales —como autores, obras o instituciones— mediante un modelo estructurado y multilingüe. Wikidata se ha convertido en un concentrador de identificadores para una gran parte de instituciones de patrimonio cultural a nivel internacional. Para ello se han definido propiedades específicas para cada institución que permiten la conexión de los recursos[6]. Esto genera ventajas como la posibilidad de extraer metadatos a través del uso de API, la visualización de los datos a través de mapas, o la evaluación de la calidad.

En el contexto europeo, iniciativas como la Nube Europea de la Ciencia Abierta (EOSC) o los espacios de datos sectoriales promueven la reutilización de datos abiertos de forma segura y respetando la soberanía institucional [Candela, 2025]. Sin embargo, a nivel nacional, no se ha realizado hasta la fecha una revisión de las iniciativas que integran de manera explícita los proyectos Wikimedia en sus flujos de trabajo. Este vacío justifica la necesidad de estudios como el presente, que buscan analizar las experiencias existentes en el entorno español.

Metodología

Esta investigación ha seguido un enfoque cualitativo, centrado en el análisis de iniciativas nacionales vinculadas al patrimonio cultural que integran Wikidata en sus actividades. El proceso se ha estructurado en tres fases: identificación, análisis y presentación de los resultados.

En primer lugar, se realizó una búsqueda sistemática para identificar proyectos activos en el ámbito nacional relacionados con archivos, bibliotecas o centros de documentación. La búsqueda combinó distintas fuentes: bases de datos científicas (como Scopus y Web of Science), plataformas especializadas (como Wikidata y el Social Sciences and Humanities Open Marketplace), y páginas web oficiales de los propios proyectos, donde se consultaron informes, documentación técnica y de difusión.

En segundo lugar, se llevó a cabo un análisis manual del uso que estas iniciativas hacen de Wikidata. Este análisis incluyó la detección de vínculos a la base de conocimiento, la existencia de propiedades específicas para las instituciones, la realización de actividades de edición colaborativa (como cursos de formación) y la publicación de visualizaciones o materiales reutilizables. Para organizar esta información, se emplearon categorías del vocabulario TaDiRAH, permitiendo clasificar las acciones de forma estandarizada en el ámbito de las Humanidades Digitales.

[6] Por ejemplo, la Biblioteca Nacional de España tiene asignada la siguiente propiedad https://www.wikidata.org/wiki/Property:P950

Por último, los resultados se presentaron en forma de texto describiendo el trabajo realizado hasta el momento por parte de las iniciativas seleccionadas, realizando una comparación a través de unas dimensiones específicas y detallando los aspectos donde se observa posibilidad de mejora.

Resultados

A partir de la metodología propuesta, se han analizado seis iniciativas nacionales vinculadas al ámbito del patrimonio cultural. Estas iniciativas representan distintos tipos de instituciones —archivos, bibliotecas, museos y laboratorios digitales— y han sido seleccionadas por incorporar de forma explícita Wikidata como proyecto.

Para su análisis, se han definido cuatro dimensiones principales centradas en el uso de Wikidata e inspiradas en el vocabulario TaDiRAH: (1) enriquecimiento de metadatos mediante Wikidata (tadirah:enriching), (2) visualización de datos (tadirah:dataVisualization), (3) evaluación de la calidad (tadirah:dataCleansing), y (4) diseminación y actividades de formación o divulgación (tadirah:disseminating). Estas dimensiones permiten identificar buenas prácticas, detectar carencias comunes y establecer un punto de partida para futuras líneas de trabajo. La tabla 1 siguiente resume los resultados obtenidos.

Iniciativas	Enriquecimiento	Visualización	Calidad	Diseminación
Biblioteca Virtual Miguel de Cervantes	Sí	Sí	Sí	Sí
Biblioteca Nacional de España (BNE)	Sí	Sí	No	No
Mnemosine, Biblioteca Digital de La otra Edad de Plata	Sí	Sí	No	No
Museo del Prado	Sí	Sí	No	No
Portal de Archivos Españoles (PARES)	Sí	No	No	No
IAPH: rutas de paisajes culturales	Sí	No	No	Sí

Tabla 1. Uso de Wikidata en las instituciones de patrimonio cultural en España atendiendo a las cuatro dimensiones identificadas.

De acuerdo a los resultados obtenidos, la BVMC constituye uno de los ejemplos más completos en el uso de los proyectos Wikimedia. Dispone de

cuatro propiedades en Wikidata, páginas temáticas en Wikipedia, así como ejemplos visibles de visualización de datos y actividades de divulgación. Además, proporciona código reproducible para extraer y representar datos, lo que la convierte en una referencia de buenas prácticas en el ámbito hispanohablante.

La BNE cuenta con una propiedad en Wikidata para enlazar sus catálogos y registros de autoridad. También ha participado en iniciativas de visualización de colecciones a partir de datos estructurados. Sin embargo, no se identificaron actividades de divulgación ni estrategias explícitas para mejorar la calidad de los datos.

Mnemosine cuenta con una propiedad en Wikidata y ha contribuido a mejorar la visibilidad de figuras relevantes mediante páginas en Wikipedia. Pese a contar con visualizaciones de datos, no se han identificado estrategias para evaluar o documentar la calidad de los metadatos ni actividades abiertas de formación o difusión.

El Museo del Prado cuenta con dos propiedades en Wikidata, una para los artistas y otra para las obras de arte. Se observan mapas interactivos y recursos en Wikipedia, aunque estas acciones no siempre están coordinadas institucionalmente, y no hay documentación sobre calidad de datos ni acciones de diseminación.

PARES cuenta con una propiedad en Wikidata, tiene una fuerte presencia digital y cuenta con contenidos enlazados a Wikidata, aunque de forma parcial. No dispone de visualizaciones ni ejemplos de reutilización abiertos, lo que limita su potencial en contextos educativos y de investigación. Tampoco se documenta ninguna acción de mejora de calidad de datos o difusión.

El Laboratorio del Paisaje Cultural es un proyecto emergente que ha iniciado vínculos con Wikidata y ha promovido la apertura de datos vinculados con elementos del paisaje. No obstante, aún no se evidencian visualizaciones desarrolladas ni estrategias de evaluación de calidad. La ausencia de actividades de formación limita su proyección.

Discusión

Los resultados muestran una adopción desigual de Wikidata, especialmente en las dimensiones de visualización, calidad de datos y diseminación. Esta diversidad refleja distintos niveles de integración en el uso de herramientas colaborativas. La ausencia de difusión y actividades de formación en materia de los proyectos Wikimedia, limita en cierta medida la proyección de los proyectos.

La BVMC representa un caso consolidado [Candela, 2017], con presencia en las cuatro dimensiones analizadas. Su trayectoria ilustra cómo una estrategia digital sostenida puede facilitar tanto la visibilidad como la reutilización de los datos. Otras instituciones, como la BNE [Vila-Suero, 2013], han centrado sus esfuerzos en el enriquecimiento de sus colecciones, sin desarrollar de forma

paralela mecanismos de difusión o evaluación de calidad. En algunos casos, los resultados de investigación en materia de calidad de datos realizados por investigadores pueden ser una fuente de información adicional [Candela, 2023].

Proyectos más recientes, como los de las rutas de paisajes culturales [Rodrigo, 2024], del Instituto Andaluz del Patrimonio Histórico, demuestran un interés emergente por conectar sus recursos con Wikidata [Fernández Cacho, 2023]. Esta situación coincide con lo que apuntan iniciativas como AI4LAM o DHwiki, que subrayan la necesidad de generar prácticas sostenibles y sistemáticas de publicación y enriquecimiento de datos.

Por último, el análisis evidencia que si bien el enriquecimiento con Wikidata es una práctica poco extendida entre las instituciones que trabajan con el patrimonio cultural, las que sí lo han hecho son instituciones relevantes a nivel nacional, que han visto una oportunidad para ampliar la visualización y el impacto de sus catálogos. Existe la necesidad de explorar este tipo de actividades en instituciones de menor tamaño para incrementar la difusión y la visibilidad en entornos locales y regionales.

Conclusiones

Las instituciones de patrimonio cultural facilitan el acceso a sus colecciones digitales desde hace décadas. En los últimos años, los proyectos Wikimedia han jugado un papel relevante en el enriquecimiento y la visibilidad de estas colecciones en un contexto abierto y colaborativo.

Los resultados de este trabajo muestran que las iniciativas analizadas han utilizado de forma desigual Wikidata, especialmente en lo relativo a visualización y diseminación de datos. A pesar de los avances en la vinculación con esta base de conocimiento, se detecta un margen de mejora significativo en la publicación de ejemplos de reutilización y en la provisión de información sobre la calidad de los datos, aspectos fundamentales para que los investigadores puedan incorporar estos recursos en sus propios proyectos.

Todas las iniciativas están conectadas con Wikidata de alguna forma, y tan solo una no dispone de una propiedad específica en la plataforma. La mayoría incluye visualizaciones en sus sitios web, ya sea mediante gráficas o código reproducible, lo que representa una buena práctica a seguir. No obstante, reforzar la calidad de los datos y crear actividades de formación o divulgación siguen siendo puntos a mejorar para el futuro.

Este trabajo ofrece un estudio inicial que describe el papel de los proyectos Wikimedia en el ecosistema digital del patrimonio cultural en España. Las líneas futuras incluyen la incorporación de dimensiones adicionales para evaluar el impacto de estas acciones, y la adopción de ontologías para describir los resultados, así como el análisis de la incorporación de otros proyectos Wikimedia más allá de Wikidata, como Wikipedia o Commons.

Bibliografía

CANDELA, G., CHAMBERS, S., IROLLO, A., FREIRE, N., DRITSOU, V., ISAAC, A., BENARDOU, A., GARNETT, V., & TASOVAC, T. (2025), *A workflow to publish collections as data: Looking back at Europeana.eu and forward to the common European data space for cultural heritage* (versión v2), Zenodo. https://doi.org/10.5281/zenodo.15275024

CANDELA, G. (2023), «An automatic data quality approach to assess semantic data from cultural heritage institutions», *Journal of the Association for Information Science and Technology*, 74(7), pp. 866–878. https://doi.org/10.1002/asi.24761

CANDELA, G., CUPER, M., HOLOWNIA, O., GABRIËLS, N., DOBREVA, M., & MAHEY, M. (2024), «A systematic review of Wikidata in GLAM institutions: A Labs approach», en A. ANTONACOPOULOS, et al. (eds.), *Linking Theory and Practice of Digital Libraries (TPDL 2024)*, *Lecture Notes in Computer Science*, vol. 15178, Springer, Cham. https://doi.org/10.1007/978-3-031-72440-4_4

CANDELA, G., ESCOBAR, P., CARRASCO, R., & MARCO-SUCH, M. (2017), «Migration of a library catalogue into RDA linked open data», *Semantic Web*, 9(4), pp. 481–491. https://doi.org/10.3233/SW-170274

FERNÁNDEZ CACHO, S., CUEVAS GARCÍA, J., & OLEA GONZÁLEZ, I. (en prensa), «Rutas 2.0 por los Paisajes de Interés Cultural de Andalucía», en *III Congreso Internacional de Musealización y Puesta en Valor del Patrimonio Cultural – Legatum 2.0* (25–27 octubre 2023), Daimiel, Ciudad Real.

MUSEO DEL PRADO (s. f.), «El grafo de conocimiento del Museo del Prado», recurso en línea. https://www.museodelprado.es/grafo-de-conocimiento/el-grafo-de-conocimiento-del-museo-del-prado [Acceso: 29/04/2025].

PADILLA, T., SCATES KETTLER, H., VARNER, S., & SHORISH, Y. (2023), *Vancouver Statement on Collections as Data*, Zenodo. https://doi.org/10.5281/zenodo.8342171

PORTAL DE ARCHIVOS ESPAÑOLES (PARES) (s. f.), recurso en línea. https://pares.cultura.gob.es/inicio.html [Acceso: 29/04/2025].

RODRIGO, J. M., & FERNÁNDEZ, S. (2024), «Paisaje cultural, percepciones y agentes locales en el desarrollo socioeconómico del territorio», *Boletín del Instituto Andaluz del Patrimonio Histórico*, año 32, n.º extra 112 (monográfico), pp. 88–116. https://doi.org/10.33349/2024.112.5580

VILA-SUERO, D., & GÓMEZ-PÉREZ, A. (2013), «datos.bne.es and MARiMbA: An insight into library linked data», *Library Hi Tech*, 31(4), pp. 575–601. https://doi.org/10.1108/LHT-03-2013-0031

Interpretar, documentar y gestionar digitalmente la movilidad como patrimonio cultural: una propuesta conceptual y metodológica

Sonia Medina Gordo
Qark Arqueología SL – Universitat de Barcelona
César González-Pérez
Instituto de Ciencias del Patrimonio – CSIC
Esther Travé Allepuz
Universitat de Barcelona – Institut d'Arqueologia

Palabras clave: Modelado conceptual; modelo de referencia; gestión de conocimiento; digitalización; Humanidades Digitales
Keywords: Conceptual modeling; reference model; knowledge management; digitization; Digital Humanities

El estudio de la movilidad ha sido abordado desde enfoques diversos, en función de los intereses teóricos y metodológicos de cada disciplina. En el marco de esta propuesta, se plantea como un fenómeno transversal y constitutivo de las entidades autónomas[1] asociado a la capacidad de desplazamiento. Esta premisa ofrece un punto de partida diferente para su análisis, articulado en torno a dos pilares básicos. En primer lugar, el carácter transversal subraya que dicho fenómeno se manifiesta en múltiples dimensiones, y que su estudio no pertenece en exclusiva a una única disciplina. En segundo lugar, al concebirla como un fenómeno constitutivo de entidades con autonomía para desplazarse, se cuestiona la utilidad analítica del tradicional binomio móvil/inmóvil. Más que oponer de manera excluyente dos categorías, sostenemos que resulta más operativo reconocer que cualquier entidad puede ser evaluada, en mayor o menor medida, tanto a través de elementos estacionarios como de dinámicas de desplazamiento. De esta manera, el planteamiento abre un marco inicial con el que investigar el fenómeno desde su propia complejidad, sin imponer distinciones rígidas.

Ahora bien, este esfuerzo por pensar la movilidad desde un nivel más abstracto también pone de relieve un problema de base: la ausencia de un marco

[1] En este estudio, se entiende que *una entidad autónoma es una entidad que puede desplazarse por el espacio de forma independiente de otras*.

conceptual de referencia que permita abordarla en distintos contextos de trabajo. Parte de la dificultad radica en que su definición oscila entre identificarla directamente con la acción de movimiento o reducirla a sus manifestaciones más tangibles, como pueden ser los caminos. Esta dualidad sugiere que la movilidad, cuando se concibe como la pieza de una realidad que conceptualizamos, parte de una definición vaga y apenas consensuada. Lo fundamental en este punto es que, lejos de ser directamente observable, se trata de un concepto que necesita ser modelado mediante otras entidades relacionadas. Esta complejidad exige una aproximación que permita razonar sobre ella como un objeto concreto, pero también a través de las valoraciones que determinan esa concreción. A este respecto, las Humanidades Digitales ofrecen un terreno fértil; no tanto por la incorporación en sí misma de herramientas tecnológicas, sino por su capacidad para ampliar horizontes epistémicos mediante formas alternativas de gestión del conocimiento, especialmente en relación con fenómenos que presentan un correlato tanto material como inmaterial.

En línea con lo expuesto, el objetivo de esta comunicación es proponer una solución conceptual y metodológica para interpretar, documentar y gestionar digitalmente la movilidad como patrimonio cultural, de modo que quienes se interesen por este objeto de estudio cuenten con un procedimiento de referencia, puedan adaptar los medios tecnológicos pertinentes e integren el componente valorativo como parte esencial del proceso de trabajo. Además, al presentar la movilidad como patrimonio cultural, se busca que esta dimensión valorativa se proyecte sobre elementos cargados de significado histórico y social, con el fin de visibilizar el contexto en el que se construye dicho valor y reforzar su potencial para ser conservados. Para alcanzar este objetivo, el trabajo propuesto se ha articulado en torno a dos ámbitos organizativos diferenciados: por un lado, aquellos relacionados con la gestión del conocimiento; por el otro, los que conciernen al almacenamiento de los datos.

En el primer caso, se ofrece un modelo conceptual de la movilidad que pueda ser empleado por una amplia variedad de perfiles y casos de estudio, con el fin de facilitar un marco común de referencia que asegure la coherencia en el uso de los conceptos. Su diseño y semántica permite que cada usuario pueda centrarse, si lo desea, en una o varias áreas específicas sin necesidad de adoptar la totalidad del esquema, pero manteniendo intacta la estructura de relaciones y las propiedades de cada concepto. Esta organización favorece un aprovechamiento flexible, en el que es posible trabajar con conjuntos de entidades sin romper la coherencia general del modelo. Para este trabajo más focalizado, se distinguen tres módulos. El primero, denominado *movimientos* (Fig. 1), agrupa las entidades involucradas en la propia acción de desplazamiento, mientras que, el segundo, centrado en las *rutas* (Fig. 2), agrupa entidades que materializan las acciones de movimiento de forma tangible. En cuanto al tercero, *contextos* (Fig. 3), abarca los elementos valorativos, entre los que se incluyen tanto la herencia cultural o histórica como las narrativas que atribuyen significado al objeto de estudio según el marco de referencia.

Dado que se trata de un modelo de referencia, se espera que los usuarios puedan desarrollar su propia adaptación en función de las particularidades del proyecto. No obstante, para que cada uno pueda generar un modelo concreto de movilidad, es necesario asegurar tres pasos: una conceptualización mediante un lenguaje formal; la definición de las distintas entidades que configuran el modelo; y el establecimiento de un marco de trabajo para su gestión. En nuestro caso, para llevar a cabo este procedimiento, se ha empleado una serie de herramientas que han resultado adecuadas tanto por su capacidad para estructurar el conocimiento como por facilitar su aplicación en un entorno no necesariamente técnico. Estas decisiones responden a la lógica operativa de la propuesta metodológica y pueden ser adaptadas según otros contextos de investigación. En cualquier caso, se ha integrado:

- **ConML** (González-Pérez 2018), para razonar y expresar formalmente el modelo a través de una gramática sencilla y orientada al modelado conceptual.
- **CHARM** (González-Pérez 2018: 194-305), para generalizar entidades del modelo sobre la movilidad a partir de un modelo de referencia abstracto del patrimonio cultural.
- **Bundt,** para crear y gestionar una implementación informática del modelo, así como instancias del mismo.

En lo relativo al almacenamiento de los datos, el objetivo ha sido establecer una conexión directa entre el trabajo de gestión del conocimiento y prácticas digitales accesibles, de manera que el modelo de la movilidad pueda ser utilizado en entornos diversos sin requerir conocimientos avanzados en modelado conceptual. Este propósito parte de la idea de que el modelo funciona como una base ontológica compartida: desde él se razona, se construyen entidades y se organizan las relaciones entre conceptos. Sin embargo, muchas veces, quienes trabajan sobre estos fenómenos lo hacen directamente con datos, no con conocimiento. Por ello, resulta igualmente operativo habilitar un procedimiento que permita *traducir* los esquemas conceptuales en datos concretos, sin perder de vista la lógica que los articula. Esta traducción no sólo hace operativo el modelo, sino que permite también reutilizarlo y ampliarlo en distintas fases del ciclo de vida del dato: desde su identificación y obtención hasta su almacenamiento y análisis.

Dentro de este marco, para la propuesta se ha planteado una implementación técnica que permite incorporar distintos usos de la información, respetando la coherencia semántica del modelo conceptual. Esta implementación, como ocurre con el modelado, no constituye una solución cerrada, sino una propuesta adaptable a otras plataformas o necesidades. En el caso particular del estudio de la movilidad, este planteamiento resulta útil, dado que combina elementos con un marcado componente espacial con aspectos valorativos o narrativos que escapan a la pura representación geográfica. Así, el almacenamiento debe dar cabida tanto a estructuras provenientes de fuentes sin un anclaje explícitamente territorial como a datos generados a partir de

observaciones en análisis espaciales. En nuestro estudio, se han incorporado las siguientes herramientas:

- **Label Studio**, para el etiquetado de información y la extracción sistemática de entidades según el modelo conceptual. A largo plazo, este proceso podría emplearse para entrenar modelos de reconocimiento automático de entidades, facilitando su reutilización sin necesidad de repetir manualmente el proceso.
- **QGIS**, para explorar y visualizar la información espacial asociada a los *movimientos* o *rutas*.
- **PostgreSQL**, como sistema de almacenamiento estructurado de datos; ya sea de los derivados del etiquetado como aquellos que emergen del trabajo con información espacial, facilitando su integración y consulta según el esquema definido.

Por otro lado, esta aproximación también se ha comenzado a aplicar en casos de estudio concretos; dos que tienen que ver con la movilidad de la Sierra de la Demanda en diferentes etapas históricas (vid. Escalona Monge 1996; Medina Gordo 2024: 215-226), y otro relacionado con la conectividad que se construye en torno a las actividades de producción alfarera en Cataluña (vid. Travé Allepuz y Vicens Tarré 2022; Travé Allepuz *et al.* 2023). Esta decisión responde a una dinámica habitual en el ámbito de las Humanidades Digitales, donde la variedad de contextos, habilidades técnicas y objetivos de investigación impide una solución única. Por eso, más que explorar los beneficios y limitaciones de un software determinado o una arquitectura definitiva, se ha querido poner el acento en la lógica que articula las decisiones metodológicas tomadas, y explorar su operatividad en tres casos de estudio concretos.

Con todo, la propuesta que presentamos no pretende ofrecer una solución cerrada, sino una base para entender y explorar la movilidad como patrimonio cultural. A través de un modelo conceptual abstracto y herramientas digitales orientadas a la gestión de los datos, se proporciona un marco común que puede ser adaptado a diversas realidades de investigación que giren en torno a la movilidad. Al centrar la atención en los procesos de conceptualización, valoración y gestión, se refuerza el potencial de estos estudios para integrarse en narrativas patrimoniales más amplias y sostenibles. Por último, en la medida en que se trata de una propuesta en desarrollo, esta comunicación busca no sólo presentar el enfoque metodológico, sino también abrir un espacio de discusión crítica sobre sus posibilidades, límites y futuras aplicaciones.

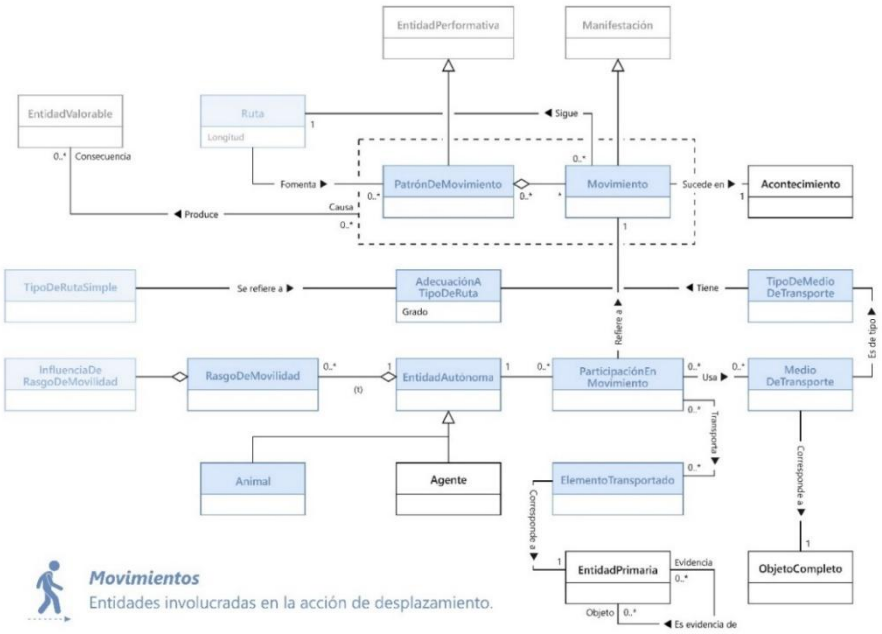


Figura 1. Modelo de clases ConML cuyas entidades están involucradas en la acción de desplazamiento. Por ejemplo, en la oración "*Mi abuelo emigró a Estados Unidos en barco"* se identifica un *Agente* ('Mi abuelo') que realiza un *Movimiento* ('Emigró') usando un *MedioDeTransporte* ('Barco'). En el modelo, las clases blancas corresponden a CHARM, mientras que en azul se destacan las clases del modelo de movilidad.

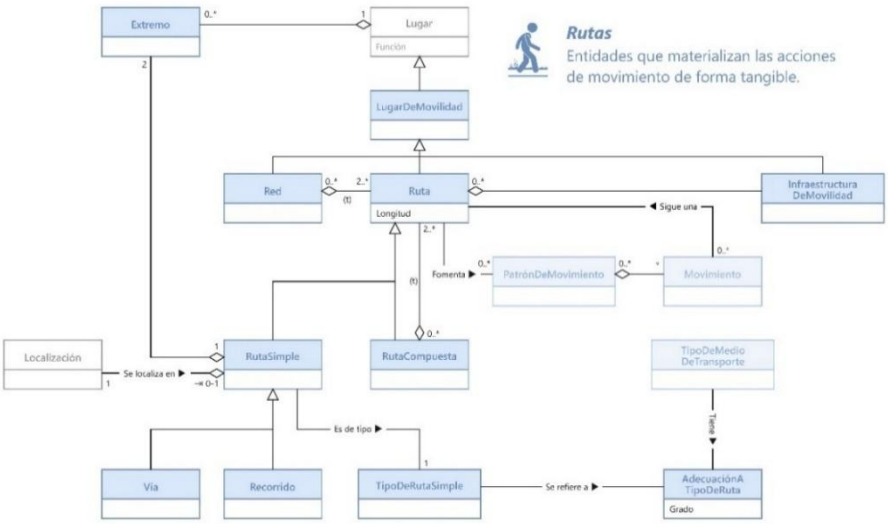


Figura 2. Modelo de clases ConML cuyas entidades materializan la acción de movimiento de forma tangible. Aquí la casuística es amplia; por ejemplo, un camino, un puente, una ruta comercial o incluso la descripción de un recorrido pueden considerarse un *LugarDeMovilidad* en nuestro modelo, en la medida en que todos ellos albergan movimiento. En el modelo, las clases blancas corresponden a CHARM, mientras que en azul se destacan las clases del modelo de movilidad.

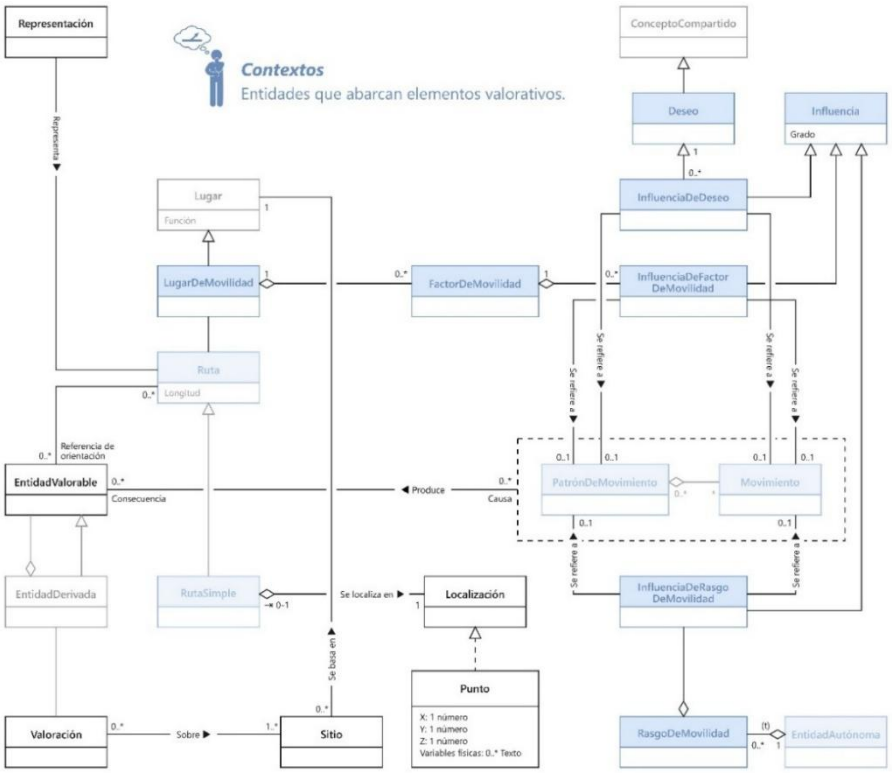


Figura 3. Modelo de clases ConML cuyas entidades abarcan elementos valorativos. Por ejemplo, en la oración "*Mi abuela era quien salía a recoger leña todas las mañanas cuando mi abuelo enfermó*" es posible realizar varias valoraciones, aun cuando el movimiento parece ser el mismo. Así, se podría valorar que existe un *Deseo*, digamos que el de querer mantener la casa caliente, que podría estar funcionando como vector de un *Movimiento* ('salir a recoger leña'). Al mismo tiempo, la salud se muestra como un *RasgoDeMovilidad* que influye en la medida en que a uno le impide realizar el movimiento, y al otro no. En el modelo, las clases blancas corresponden a CHARM, mientras que en azul se destacan las clases del modelo de movilidad.

Bibliografía

ESCALONA MONGE, J. (1996): *Transformaciones sociales y organización del espacio en el Alfoz de Lara en la Alta Edad Media*. Tesis Doctoral. Madrid: Universidad Complutense de Madrid. https://hdl.handle.net/20.500.14352/61722.

GONZÁLEZ-PÉREZ, C. (2018): *Information Modelling for Archaeology and Anthropology. Software Engineering Principles for Cultural Heritage*. Cham: Springer; doi: 10.1007/978-3-319-72652-6.

MEDINA GORDO, S. (2024): *Paisajes de montaña y periferias en la Alta Edad Media. Procesos de gestión de la información para un relato histórico integrado*. Tesis Doctoral. Barcelona: Universitat de Barcelona. http://hdl.handle.net/10803/694214.

TRAVÉ ALLEPUZ, E.; ROCAS GUTIÉRREZ, X.; VICENS TARRÉ, J. (2023): "Exploring material culture and identity in post-medieval Catalonia: a long-term archaeological perspective on greyware pottery production in Quart, Girona, Spain". *Post-Medieval Archaeology*, 57(3), 319-337; doi: 10.1080/00794236.2023.2285282.

TRAVÉ ALLEPUZ, E. Y VICENS TARRÉ, J. (2022): "La producció de terrissa a les gavarres: una anàlisi des de l'arqueologia del paisatge". *Estudis del Baix Empordà*, 41, 73-103.

Inteligencia artificial y visión por computador para la memoria colectiva: preservación digital del patrimonio en riesgo

Laya Targa Hernández
Universitat de València

Palabras clave: Inteligencia artificial; visión por computador; patrimonio cultural; conservación predictiva
Keywords: Artificial intelligence; computer vision; cultural heritage; predictive conservation

Introducción

La preservación del patrimonio cultural se ve afectada por desafíos críticos impulsados por factores antropogénicos, como conflictos bélicos, así como factores ambientales, entre ellos el cambio climático y la contaminación (Presidencia Del CSIC et al., 2025). Estas amenazas no solo comprometen la integridad física de monumentos, edificios y artefactos, sino que también ponen en riesgo la memoria colectiva vinculada a estos bienes. Como consecuencia, se deben llevar a cabo procesos de restauración que, muchas veces, se alargan en el tiempo y deben evitar alterar los rasgos característicos de las obras (Kumar y Gupta, 2023; Zhao et al., 2024).

Frente a estos desafíos, las tecnologías emergentes ofrecen nuevas oportunidades para la conservación patrimonial. En particular, la inteligencia artificial (IA) y la visión por computador permiten abordar de forma automatizada y no invasiva los procesos de diagnóstico, restauración y predicción de riesgos, contribuyendo a una gestión más eficiente y preventiva del deterioro.

Este trabajo, desarrollado en el marco del programa de doctorado en Tecnologías de la Información, Comunicaciones y Computación de la Universitat de València y vinculado al proyecto europeo ChemiNova (financiado bajo el programa Horizonte Europa), propone una metodología basada en inteligencia artificial y visión por computador para abordar tres dimensiones clave: diagnóstico automatizado de daños, restauración digital no invasiva y modelado predictivo de riesgos. La investigación se fundamenta en el análisis de cuatro estudios piloto que ejemplifican distintos escenarios de deterioro del patrimonio cultural en el contexto europeo: Valencia (España), Kiev (Ucrania), Palermo (Italia) y Viena (Austria).

Metodología

La investigación aborda la complejidad del deterioro patrimonial mediante un marco metodológico estructurado en tres ejes interrelacionados: tipologías de deterioro, diversidad material y estrategias de adquisición de datos.

La selección de los elementos en cada uno de los pilotos evidencia una diversidad en los materiales empleados, tales como piedra, madera, cera, papel, entre otros. Estos materiales están expuestos a distintos tipos de deterioro, derivados tanto del cambio climático como de los conflictos bélicos. Por ejemplo, los elementos arquitectónicos del Centro Cultural La Nau (ubicado en Valencia) están más expuestos a fenómenos atmosféricos, mientras que las obras exhibidas en el interior pueden sufrir cambios debido a la variación en la humedad y temperatura. En el caso de Palermo, los modelos de cera son mucho más frágiles y están expuestos también a cambios de humedad.

Los daños estructurales y químicos varían en función del tipo de datos adquiridos. Las imágenes RGB permiten detectar alteraciones superficiales como cambios de color, acumulación de suciedad u oxidación. Por su parte, la técnica de *Reflectance Transformation Imaging* (RTI), que registra cómo la luz interactúa con las superficies, facilita la identificación de atributos visuales sutiles, lo que resulta especialmente útil para la monitorización de cambios en materiales sensibles.

La base de datos multimodal integra, entre otros recursos, imágenes RGB, hiperespectrales y RTI, aplicadas a distintas escalas de intervención —como edificios, monumentos y artefactos— con el propósito de capturar una amplia variedad de datos estructurales y químicos relevantes para el análisis del patrimonio cultural.

El marco tecnológico está enfocado en el diseño de modelos que respondan a las tres dimensiones previamente mencionadas.

Los modelos de inteligencia artificial desarrollados deben ser capaces de identificar deterioros en bienes culturales, adaptándose a la diversidad de las amenazas:

- Deterioro climático: Análisis de patrones sutiles como eflorescencias salinas en piedra, micro fisuras térmicas en cera o delaminación en pinturas provocada por humedad.
- Daños bélicos: Identificación de fracturas estructurales, residuos explosivos o vibraciones residuales en monumentos afectados por conflictos.

Uno de los principales desafíos del entrenamiento de modelos de detección de datos es la variabilidad en el análisis de los materiales, reconociendo cómo cada material manifiesta el deterioro bajo distintas amenazas. A partir de esta etiquetación automática es posible generar mapas de daños específicos en cada uno de los pilotos.

Asimismo, la restauración digital no invasiva permite evitar intervenciones físicas irreversibles mediante dos enfoques complementarios: la reconstrucción

contextual, basada en redes generativas antagónicas (GANs, por sus siglas en inglés), y las técnicas de *inpainting*.

- En escalas a nivel de artefacto, las GAN pueden reconstruir áreas dañadas simulando comportamientos de materiales específicos, p. e., daños producidos por la deformación en la cera debido al cambio de temperatura. Mientras que, a un nivel más amplio como ocurre en el caso de monumentos, es factible restaurar lagunas pictóricas manteniendo coherencia estilística mediante el análisis de texturas RTI y patrones históricos.
- Los algoritmos de *inpaiting* permiten reconstruir áreas perdidas tanto en pinturas murales mediante el análisis contextual de patrones adyacentes manteniendo sus detalles (Yeh et al., 2024) como en capas pictóricas ocultas respetando las técnicas artísticas originales.

Finalmente, el modelado predictivo de riesgos integra datos históricos y ambientales (temperatura, humedad, eventos extremos) con arquitecturas de IA como redes neuronales recurrentes (RNNs) y máquinas de soporte vectorial (SVM). Estos modelos anticipan daños futuros como la corrosión en metales o delaminación en pinturas, generando alertas tempranas personalizadas para cada elemento.

La validación de todos los modelos sigue protocolos entre los que destacan la división 70/30 de datos, métricas como F1-score, SSIM (*Structural Similarity Index Measure*) o PSNR (*Peak Signal-to-Noise Ratio*), y la evaluación interdisciplinar con historiadores del arte y conservadores. Siguiendo estos protocolos y métricas se consigue una evaluación exhaustiva de los resultados de los modelos.

Resultados

Tal como se indicó anteriormente, el principal objetivo previsto consiste en el desarrollo de herramientas fundamentadas en inteligencia artificial y visión por computador, orientadas a la detección temprana de daños y a la implementación de estrategias preventivas. No obstante, como paso preliminar, se establecerá un marco metodológico comparativo que permita clasificar jerárquicamente los algoritmos de inteligencia artificial en función de su eficacia frente a diversas tipologías de deterioro patrimonial (químico, estructural y biológico). Dicho marco incorporará métricas validadas teóricamente que evalúan la precisión, la robustez y la escalabilidad de los modelos.

Para cada uno de los ejes principales —diagnóstico automatizado de daños, restauración digital no invasiva y modelado predictivo de riesgos— se detallan a continuación los resultados que se espera alcanzar al concluir la investigación.

En cuanto a los modelos de IA, serán desarrollados sistemas automatizados que permitirán identificar daños en bienes patrimoniales (grietas, corrosión, humedad) con mayor rapidez y precisión que los métodos

tradicionales. Estas herramientas generarán informes con mapas visuales señalizando áreas críticas.

El sistema de restauración digital tiene como base la creación protocolos de restauración virtual que reconstruirán zonas dañadas de pinturas, esculturas o estructuras arquitectónicas manteniendo su autenticidad histórica y estética. Mediante tecnologías no invasivas, será posible visualizar reconstrucciones realistas antes de intervenciones físicas, reduciendo riesgos y preservando materiales frágiles.

Finalmente, se implementará un sistema de alerta temprana diseñado para anticipar daños futuros en monumentos y artefactos. Este sistema combinará datos históricos, condiciones ambientales y patrones de deterioro para generar recomendaciones prácticas, como, por ejemplo: "Controlar los niveles de humedad en la sala oeste durante los próximos tres meses", así como proyecciones de impacto, tales como: "Alta probabilidad de aparición de grietas en la fachada norte en un plazo de 12 meses". De este modo, se facilitará una toma de decisiones preventiva y basada en evidencia para la conservación patrimonial.

Conclusión

Esta investigación propone algunas bases para un modelo de conservación patrimonial donde la inteligencia artificial y la visión por computador se convertirán en aliados estratégicos frente a amenazas globales como los conflictos bélicos, el cambio climático y la degradación material. Los resultados transformarán la práctica actual al migrar de enfoques reactivos, basados en intervenciones tras el daño, hacia un paradigma preventivo y predictivo. Los sistemas de diagnóstico rápido y alerta temprana permitirán actuar antes de que los deterioros sean irreversibles gracias a la integración de los datos ambientales.

En paralelo, se establecerá una transición de lo físico a lo digital mediante reconstrucciones no invasivas. Estas herramientas preservarán la esencia de bienes dañados o en riesgo inminente, generando réplicas accesibles para educación, investigación y memoria colectiva. Se reducirá drásticamente la dependencia de intervenciones físicas invasivas en materiales frágiles, sustituyéndolas por procesos digitales que respetan la autenticidad material e histórica.

Bibliografía

KUMAR, P., & GUPTA, V. (2023), «Restoration of damaged artworks based on a generative adversarial network», *Multimedia Tools and Applications*, 82(26), pp. 40967–40985. https://doi.org/10.1007/s11042-023-15222-2

PRESIDENCIA DEL CSIC; LÓPEZ-ROMERO, E.; ASINS VELIS, S.; DÍEZ HERRERO, A.; & SASTRE PRATS, I. (2025), *Patrimonio cultural en riesgo.*

Retos de adaptación al cambio climático, Editorial CSIC. https://doi.org/10.20350/DIGITALCSIC/17003

YEH, C.-H., YANG, H.-F., CHEN, M.-J., & KANG, L.-W. (2024), «Image inpainting based on GAN-driven structure- and texture-aware learning with application to object removal», *Applied Soft Computing*, 161, art. 111748. https://doi.org/10.1016/j.asoc.2024.111748

ZHAO, F., REN, H., SUN, K., & ZHU, X. (2024), «GAN-based heterogeneous network for ancient mural restoration», *Heritage Science*, 12(1), pp. 1–15. https://doi.org/10.1186/s40494-024-01517-6

Digitalización de fondos documentales: el buscador de pasajeros de la Compañía Trasatlántica Española

Silvia Sanz Prieto
Biblioteca de la Universidad de Cantabria

Palabras clave: Compañía Trasatlántica; Pérez & Cía.; libros de embarque; emigración
Keywords: Transatlantic Company; Pérez & Cía.; shipping books; emigration

Desde su fundación, en 1972, la Universidad de Cantabria ha ido creando un patrimonio propio que se agrupa en la Colección Museográfica de la UC y que se divide en Sección de Patrimonio Artístico, Sección de Patrimonio Natural y Sección de Patrimonio Bibliográfico y Documental. Es, en esta última, en la que la Biblioteca de la Universidad participa más intensamente.

El patrimonio bibliográfico y documental de la Biblioteca de la Universidad de Cantabria se compone de las colecciones bibliográficas propias y de donaciones de archivos personales y de empresas llegados a la institución desde los años 90 del siglo pasado. Los archivos pertenecen a figuras relevantes de la vida cultural y social del siglo XX en Cantabria (arqueólogos, políticos, militares, etc.) y de empresas significativas de la economía de la región, de gran importancia para dar contexto a la creación y evolución de la sociedad, la economía y la cultura de la comunidad.

Dentro de los fondos documentales de la biblioteca de la Universidad de Cantabria encontramos el archivo de la naviera Pérez & Cía., fundada en 1853 y que hoy en día sigue en activo. La documentación de este fondo es de gran relevancia para conocer la historia de Santander y su puerto desde la segunda mitad del siglo XIX, cobrando especial relevancia los libros de venta de billetes de embarque. Estos libros contienen los billetes que la Compañía Trasatlántica vendía para viajar a América y que recogen los listados de pasajeros que viajaron desde Santander entre los años 1900 y 1960.

El archivo fue donado a la Universidad de Cantabria en el año 1997 mediante un convenio, renovado en 2011, por el que se autoriza a la Universidad a la divulgación de los documentos con fines de investigación y docencia.

El fondo Pérez & Cía. se organiza en nueve series con una cronología que va de 1859 a 1967, siendo en la Serie Compañía Trasatlántica donde se localizan los libros de embarque, cada uno de ellos con unas 400 páginas. Los libros de embarque recogen gran cantidad de información, muy útil para genealogías e investigaciones históricas, por lo que se hacía necesaria su digitalización. El proceso de digitalización se realizó con un escáner aéreo, individualizando cada

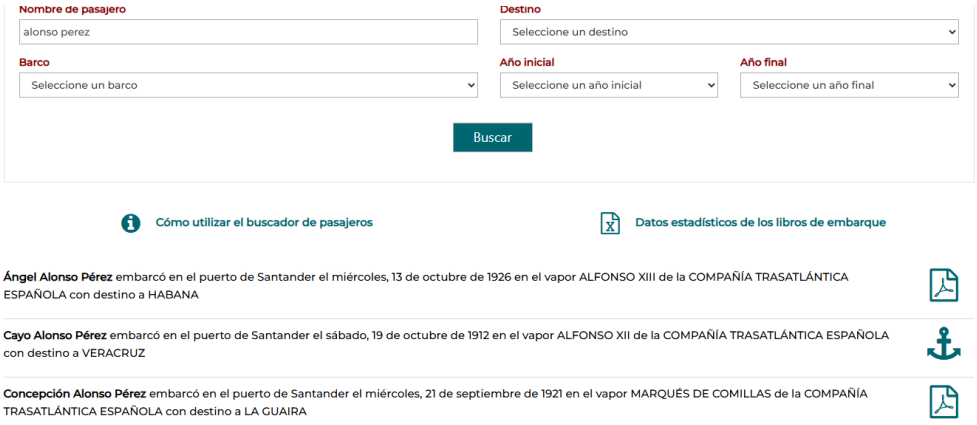
una de las páginas de los libros, con una resolución de 600 ppp, a color, y guardados en formato TIFF para su preservación. Los libros digitalizados se almacenan en ArchiBUC, un repositorio de uso interno, con software DSpace, que se utiliza para preservar los archivos digitales. En la página web de la Biblioteca de la Universidad de Cantabria se pueden descargar la totalidad de los libros en formato PDF, cuyas fechas están entre los años 1900 y 1960, con lagunas entre 1912 y 1916 y 1931 y 1945.

Si bien es verdad que la digitalización de estos libros permitía hacer consultas sin estar presente en la biblioteca, las búsquedas no estaban implementadas, por lo que para localizar cualquier nombre era necesario pasar por todas las páginas hasta encontrar el pasajero que se buscaba. Por ello, al tiempo que se digitalizaban los libros se decidió realizar una base de datos en la que poder hacer búsquedas combinadas, ágiles, y que permitieran localizar al pasajero fácilmente. Para ello, se formó un equipo de cuatro personas que trabajaban sobre un documento en formato Excel en el que se copiaba toda la información que aparecía en los libros de embarque y que se definió en las siguientes columnas: nombre del pasajero, tipo de billete, destino, nombre del barco en el que viajó y las fechas en que lo hizo. Este proceso se tradujo en la transcripción de más de 45.000 nombres de pasajeros que se recogen en los libros de embarque y que se calcula en unas 1000 horas de trabajo, muchas de ellas dedicadas a descifrar ciertos nombres que no eran fáciles de transcribir por diferentes causas: letras ilegibles, desgaste de las tintas o manchas de humedad.

En el año 2019 esta base de datos dio lugar a un buscador de pasajeros que permite hacer búsquedas combinadas por nombre y apellidos, destino, buque y fecha de salida, y que se puede consultar en la página web de la Biblioteca de la Universidad de Cantabria. Se recomienda hacer búsquedas más generales si no se obtienen resultados porque pueden existir errores en los apellidos, bien porque estuvieran mal de origen o porque se hayan introducido mal en la plantilla de la base de datos; de igual manera, alternar con diferentes grafías de ciertos apellidos que van cambiando con el tiempo (Ybarra/Ibarra, Acevo/Acebo). También hay que tener en cuenta la época, ya que, por ejemplo, en el caso de las mujeres, los niños y los sirvientes casi siempre son personas anónimas y únicamente aparece el nombre del cabeza de familia.

Para poder hacer una búsqueda tenemos que indicar el nombre del pasajero que queremos localizar y, si sabemos algún dato más, como el destino, el nombre del barco o la fecha, lo añadiremos para que el resultado sea más

pertinente, pero no es necesario. La búsqueda en la casilla “nombre de pasajero” lo hace a texto completo, de manera que recupera datos de pasajeros que coincidan con ella, por ejemplo, si buscamos “Alonso Pérez" y no lo combinamos con otros datos, el resultado serán todos los pasajeros que tengan alguna de estas palabras en su primer apellido, en su segundo apellido y en el nombre, ya que, en este caso, Alonso puede ser tanto nombre como apellido.



La información sobre el pasajero viene acompañada en la parte derecha de un icono PDF que indica que se puede descargar el documento que nos lleva a la página del libro en la que se encuentra el nombre.

Actualmente el buscador de pasajeros tiene 55.000 registros. El grueso de la información, 45.000, procede de los libros de embarque del Fondo Pérez & Cía., ubicados en la Biblioteca de la Universidad. De las solicitudes de pasaportes del siglo XIX, y que se publicaban en el Boletín Oficial de la Provincia de Santander, proceden 8.000 registros, y el resto, 2.000, de Navegantes, una asociación que tiene como objetivo la creación de un Censo de la emigración de Asturias y Cantabria hacia América, cuya información procede del Archivo General de la Administración, que conserva los registros del consulado de Veracruz.

En el caso de los nombres recuperados a través de la asociación Navegantes, el icono que aparece al lado de la información del pasajero tiene forma de ancla, mientras que un icono con forma de periódico nos indica que se trata de las solicitudes de pasaportes que están digitalizadas en la Biblioteca Virtual de Prensa Histórica, del Ministerio de Cultura.

El buscador de pasajeros se complementa con los datos estadísticos que aparecen en los libros de embarque: país, fecha, día, mes y año, barco, destino y tipo y categoría de billete, y que se puede descargar en formato Excel a través de la web de la Biblioteca.

Este buscador es uno de los grandes atractivos de las colecciones especiales que conservamos en la Biblioteca de la Universidad, con unas 1.000 consultas al mes, que llegan desde diferentes vías. La más utilizada son las consultas anónimas a través del buscador, seguidas de las peticiones por correo electrónico, redes sociales y WhatsApp. Las consultas se refieren a la búsqueda

de familiares que emigraron a América entre finales del siglo XIX y principios del XX. A solicitud del peticionario, y siempre que el pasajero aparezca en los libros de embarque del Fondo Pérez y Cía., se emiten certificados digitales que corroboran que el pasajero buscado consta en los libros de embarque. Los descendientes de emigrantes españoles solicitan estos certificados como información complementaria para la petición de la nacionalidad española, en base a la Ley de Memoria democrática.

Como ya hemos dicho, este fondo es uno de los más consultados y demandados, sobre todo desde que se puso en marcha el buscador de pasajeros. Además de su consulta a través de la web, en el año 2023 se inauguró UC ReCrea, una plataforma de contenidos digitalizados para dar a conocer las colecciones especiales a través de la digitalización de sus fondos. UC ReCrea incluye el Portal de la emigración que contiene el buscador de pasajeros, los libros de embarque y dos exposiciones virtuales. UC ReCrea se basa en el software libre Omeka, especializado en la difusión del patrimonio a través de las bibliotecas digitales, y que muestra los fondos de una forma muy visual y atractiva.

La biblioteca de la Universidad de Cantabria se encarga de la gestión de diferentes colecciones especiales que reflejan los dos últimos siglos de la historia de nuestra región y que haciéndolos visibles, como es el caso del buscador de pasajeros, es como generan valor. Proyectos como el aquí presentado son fundamentales para la divulgación de unos fondos que de otra manera no llegarían al público. La construcción de la base de datos partiendo de unos libros de embarque ha supuesto un antes y un después en la visibilidad y consulta del Fondo Pérez y Cía., y, por extensión, del resto de fondos especiales de la Biblioteca.

Este trabajo no termina aquí, ya que se trabaja para ampliar la base de datos de pasajeros con información procedente de nuevas fuentes, con la intención de convertir el buscador de pasajeros en un referente de la emigración de Cantabria a América.

Nueva museografía y museología del Museo Fray Pedro Gocial de Quito desde las Humanidades Digitales

Beatriz Garrido-Ramos[1]
Ana María Zabía[2]
Mercedes Sevillano Pareja[3]
UNIR.
Equipo técnico del Instituto Nacional de Patrimonio Cultural de Ecuador (INPC) y el Museo franciscano Fray Pedro Gocial

Palabras clave: Museo franciscano Fray Pedro Gocial de Quito; Museografía; Museología Digital; Humanidades Digitales; Patrimonio Cultural
Keywords: Fray Pedro Gocial Franciscan Museum of Quito; Museography; Digital Museology; Digital Humanities; Cultural Heritage

El presente proyecto (con inicio en 2025) enmarcado en el Vicerrectorado de Transferencia y Fundación UNIR, busca llevar a cabo la repotenciación del *Museo franciscano Fray Pedro Gocial de Quito* (Ecuador) a distintos niveles como el museográfico, museológico, comunicativo y tecnológico. Entre otras razones, la iglesia de San Francisco de Quito, con la capilla de Cantuña y el convento franciscano, forman junto con el museo, uno de los conjuntos arquitectónicos de mayor valor artístico y cultural de la América hispana, por la impresionante arquitectura y la belleza general de la plaza y el entorno.

El proceso se ha realizado desde una perspectiva teórico-metodológica novedosa y en auge desde hace años como es la de las *Humanidades Digitales* la cual aboga, entre otros aspectos, por el acceso abierto, la democratización de la cultura, la utilización de estándares y clasificaciones, y la divulgación y transferencia del conocimiento.

Para la evolución de las distintas fases del proyecto se ha planteado una metodología mixta, cualitativa y cuantitativa, que se complementa para la obtención de resultados diversos.

En cuanto a los objetivos a desarrollar, en etapas sucesivas, son los siguientes:

1. Inventariar y catalogar el patrimonio del museo.

[1] Docente e investigadora. Coordinadora del Máster Universitario en Humanidades Digitales. Docente e investigadora
[2] Directora de proyectos de Patrimonio Cultural.
[3] Profesora del Máster Universitario en Humanidades Digitales.

2. Aplicar la tecnología (medios interactivos) y las Humanidades Digitales a la museografía y la museología.
3. Diseñar una estrategia (diseño comunicacional) museológica para la difusión a nivel local, nacional e internacional.
4. Apoyar el proceso de nueva museografía, conservación, y posible digitalización del Museo Pedro Gocial, mediante divulgación científica y social.

El proyecto desarrolla, como exponemos, en distintas fases, la museografía y museología (digital) del Museo franciscano mediante la actualización y estandarización, entre otras, de la información de las cartelas de las obras (para su adaptación a estándares internacionales, europeos; sirvan de ejemplo el Museo Nacional del Prado o el Thyssen-Bornemisza) así como de la información de las distintas salas del museo.

Ello se traduce también en la inserción de códigos QR en las cartelas de obras representativas para conectar y completar directamente las obras con la información expuesta en la web del museo, la cual se va a revisar y reformular para estructurarla de una forma más adecuada y estandarizada (en proceso de finalización).

En relación a la señalética, se procede al lijado, masillado, base de preparación, aplicación de capa pictórica y acabados en mobiliario MDF con pintura de acabado satinado en 55 elementos de la colección. Igualmente, la impresión e instalación de 200 cédulas individuales (vinil adhesivo de 11 x 7 cm), y también 7 cédulas informativas de sala (plóter de corte de 1,10 x 0,80 m).

Por lo que respecta a las salas, en la *Sala de la Procesión del Jesús del Gran Poder* se habilita un espacio con mobiliario con panel descriptivo en braille y, además, profundizar en la experiencia sensitiva desde un punto de vista más contemplativo. Trasladando el enfoque especialmente al objeto artístico y la historicidad mediante el recurso sensorial, visual y documental, para rescatar el proceso inmaterial de significación para la comunidad feligresa. El trabajo se completa con la creación y reproducción en una pantalla interactiva de un video documental de la Procesión, el cual se reproducirá en sala y se difundirá en la web.

Cabe destacar también la propuesta realizada para la *Sala de La Virgen María* y la impresionante colección de alabastros que alberga. Museográficamente, gracias al uso de la tecnología, se pretende potenciar las características de la producción de la escuela de arte local, el uso de materiales y la originalidad de la colección, con la intención de facilitar la lectura de los 19 cuadros de la serie para el visitante, identificando el orden de los acontecimientos y su importancia. Todo ello con una correcta retroiluminación puntual de las obras, de manera que se intensifique la belleza y grandiosidad de esta colección única de alabastros.

Por todo lo expuesto, el trabajo de nueva museografía, además de ayudar a conservar (y preservar) mejor las obras, hace más accesible el museo para cualquier persona que lo visite. Asimismo, se complementará (en una posible

fase posterior) con un proyecto de innovación que permita dar a conocer esta accesibilidad de la institución.

Igualmente, con la repotenciación del museo se busca divulgar la colección mediante el diseño de una estrategia digital para la correspondiente difusión y comunicación del mismo. Este trabajo todavía se encuentra en proceso de desarrollo y se llevará a cabo con el objetivo de extraer datos necesarios para su correspondiente análisis estadístico en relación a cuestiones generales del museo y de sus distintas obras.

Cabe destacar que la Museología digital desarrollada por parte de UNIR cuenta con el apoyo del equipo técnico conformado por el INPC y el Museo Fray Pedro Gocial, de manera conjunta y participativa.

Por tanto, el trabajo se concentra especialmente en la inserción de tecnología y experiencias inmersivas para los/as usuarios/as, video documental y un ejemplo de museo virtual con las obras más características. La creación de este museo virtual se está llevando a cabo con la herramienta *Artsteps* (pendiente de finalización), utilizando para ello las piezas que se encuentran disponibles en la actualidad en la página web del museo, puesto que se trata de un prototipo inicial sobre el cual se podrá seguir trabajando en un futuro.

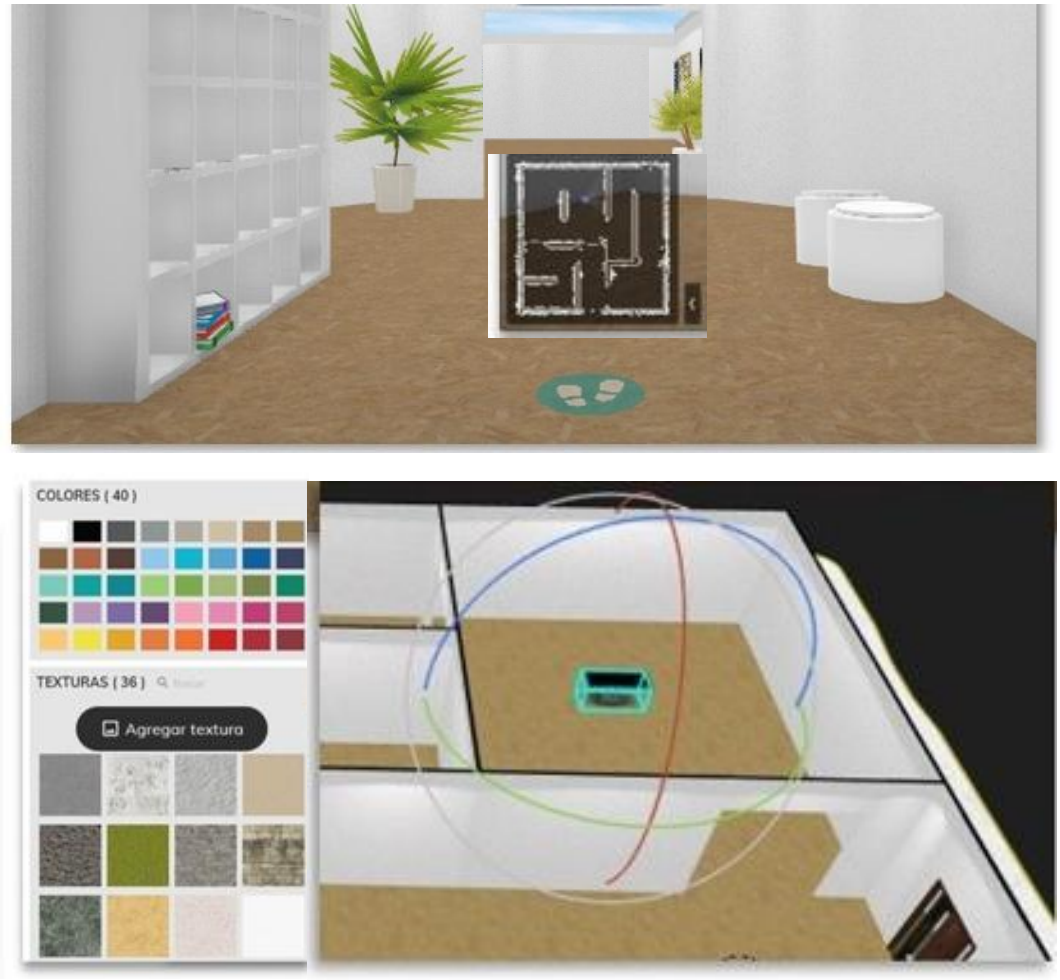
A todo ello, en nuestro caso, se añade la necesidad de formación online para la futura capacitación del personal del museo. Se trata de un curso práctico para el aprendizaje del software OMEKA, herramienta interna seleccionada para la creación del catálogo interno, que alberga alrededor de 300 obras del museo. Este software libre (versión gratuita), es muy utilizado a nivel internacional por importantes instituciones, especialmente en el ámbito de los museos. Dicha herramienta permite estructurar los datos de la colección (ficha técnica de cada obra) mediante un modelo de 15 metadatos (Dublin Core). También permite exportar toda la información contenida en la base de datos en diversos formatos para su reutilización. El motivo por el cual se ha optado por esta alternativa se fundamenta principalmente en la sencillez de su interfaz, con la intención de facilitar su manejo por parte del personal del museo. Igualmente, porque se trabaja con metadatos, lo que contribuye a la correcta inserción de la información descriptiva de cada obra, permitiendo la posibilidad de insertar material gráfico y documental, así como de registrar todo el proceso de inventario necesario para el mantenimiento de la colección. A ello se añade otro punto no menos importante, y es que al tratarse de una versión gratuita carece de coste alguno, un aspecto relevante a tener presente para la utilización del mismo a la finalización del proyecto.

En definitiva, la conservación y protección del patrimonio cultural beneficia a toda la sociedad en su conjunto, y contribuye a la conservación de los valores históricos, artísticos y culturales. En este sentido, se confirma que la correcta narración y presentación del patrimonio en lo que respecta a su aproximación a la sociedad, particularmente de los feligreses, permite una mejor comprensión y valoración de las obras que custodia el museo. Además, se traduce en un verdadero signo de identidad reconocido tanto por la comunidad local, en

particular, como de la sociedad en general, con el cual se ha ido familiarizando desde hace siglos, atendiendo a la religiosidad propia de la zona.

Finalmente, el proceso realizado en la parte de museografía cumple una función imprescindible que es la de integrar la imagen corporativa del museo y actualizar los elementos de apoyo museográfico como mobiliario y textos. Se ha revisado la información, y se ha subsanado la problemática en la disparidad de presentación del contenido al homogeneizar y estandarizar toda la información. En cuanto a la museología digital, el uso de estándares y la propuesta de estrategia comunicacional -en proceso de desarrollo-, facilitará el análisis de datos, y nos aportará una valiosa información para orientar las futuras actividades a realizar por la institución y fomentar las visitas al museo de potenciales usuarios. Con el análisis estadístico y el desarrollo del informe final podremos extraer conclusiones definitivas que, sin duda, ayudarán a completar el trabajo museográfico previo realizado en el Museo Fray Pedro Gocial.

Figuras, tablas, mapas o gráficas



Capturas de pantalla con distintas fases de la creación del Museo virtual. Fuente: Artsteps (2025).

Dublin Core (OMEKA)	Obra
Título	Nombre
Sujeto	Tema
Descripción	General y material
Tipo	Escultura, pintura, etc.
Fecha	De realización
Relación	Con otra obra o recurso
Identificador	Código interno (catálogo del museo)
Cobertura	Ubicación

Tabla con la correspondencia de metadatos destacados para las obras del catálogo.

Bibliografía

ALCARAZ, R. (2023), *Manual de Omeka Classic*, Universidad de Barcelona. http://hdl.handle.net/2445/202093

DUBLIN CORE™ METADATA INITIATIVE (s. f.), recurso en línea. https://www.dublincore.org/

GARRIDO-RAMOS, B. (2017), «Visualización de datos y clasificación Iconclass: Un estudio de caso desde la perspectiva de las Humanidades Digitales», *UCOARTE*, 10, pp. 9–18.

MANOVICH, L. (2017), «Cultural analytics, social computing, and digital humanities», en M. T. SCHÄFER & K. VAN ES (eds.), *The Datafied Society: Studying Culture through Data*, Amsterdam University Press, pp. 55–68. https://doi.org/10.25969/mediarep/12514

MINGUILLÓN, J. (2016), *Introducción a la visualización de datos: Guía de lectura de trabajos destacados en el ámbito de la visualización de datos*, UOC.

PAZMIÑO ACUÑA, R. (2024), *Propuesta de innovación digital para el Museo Franciscano Fray Pedro Gocial* [Trabajo de Fin de Máster], Facultad de Artes y Humanidades, Universidad Internacional de La Rioja.

DEL PINO, I. (2023), «La dimensión cultural y arquitectónica de San Francisco de Quito», en R. PÉREZ DE ARCE & E. DE LA CERDA (eds.), *Las vidas de San Francisco: Arquitectura, patrimonio y ciudad*, Centro UC Patrimonio Cultural, pp. 56–83.

SÁNCHEZ VEGAS, S. (2017), «Presentación: Serie de artículos sobre el plan de gestión del Conjunto Conventual de San Francisco», *Anales*, 1(373), pp. 459–462. https://doi.org/10.29166/anales.v1i373.1359

Scripta manent, sed quid de digitalibus? Desafíos y resultados de un proyecto de Historia Medieval tras trece años de investigación entre archivos y pantallas

Cristina Jular Pérez Alfaro[1]
OPIS, Instituto de Historia – Centro de Ciencias Humanas y Sociales – Consejo Superior de Investigaciones Científicas
Raúl Villagrasa Elías[2]
Universidad de La Laguna – Instituto de Estudios Medievales y Renacentistas

Palabras clave: Bases de datos, Corpus Textual Informatizado, Linaje Velasco, Acceso abierto, Corona de Castilla, Cultura escrita.
Keywords: Databases, Computerized Textual Corpus, Velasco Lineage, Open Access, Crown of Castile, Written Culture.

El proyecto *Scripta manent* ha superado una década de trabajo en el sistema de I+D+i español. Partiendo de la Historia Medieval se ha adentrado y nutrido de otras disciplinas como la Historia del Arte, la Archivística y las Ciencias y Técnicas Historiográficas, para aportar en un campo de estudio multidisciplinar como es el de la Cultura escrita. A su vez, ha dialogado con expertos en el ámbito de la computación y el diseño. En el plano teórico y administrativo, el tercer proyecto de *Scripta manent* terminará a finales de 2025, aunque en el plano práctico se mantendrá como una plataforma digital viva y útil para la comunidad académica y la sociedad. Con motivo del VII Congreso de la Sociedad Internacional Humanidades Digitales Hispánicas (Valencia, 22-24 de octubre de 2025) esta contribución pretende presentar los resultados de trece años de esfuerzo investigador y reflexionar sobre la sostenibilidad de los artefactos digitales en la "larga duración".

Scripta manent nació para estudiar la excepcional cultura escrita del linaje Fernández de Velasco, uno de los grupos nobiliarios más influyentes de la Corona de Castilla durante la Baja Edad Media y la Edad Moderna. La investigación parte

[1] Investigadora Principal de los consecutivos proyectos de *Scripta manent*, en desarrollo actualmente *Scripta manent* III. «De registros privados a textos públicos. Un archivo medieval en la Red» (Plan Nacional de I+D+i. PID2020-116104RB-I00, 2021-2025).
[2] Director de la página web *Rethos* (*Retia Hospitalium*).

de una pregunta central: ¿por qué el linaje Velasco desarrolló un sistema de creación, gestión y conservación documental (su archivo) con un grado de sofisticación y una cronología tan temprana, superando a otros grandes linajes coetáneos como los Alba, Híjar o Pacheco?

La hipótesis inicial, que hoy creemos confirmada, sostiene que los Velasco utilizaron la escritura de manera sistemática como una herramienta de memoria y dominación, en gran medida por su conexión con importantes centros de poder castellanos. Para demostrarlo, hemos analizado, principalmente aunque no sólo, el Fondo Frías del Archivo Histórico de la Nobleza, un corpus de miles de documentos (compraventas, inventarios, originales, borradores y copias autorizadas, etc.) que revela una compleja y consciente práctica escrituraria desde fechas tan tempranas como el siglo XIV, destinada a administrar un señorío pero también construir y perpetuar la identidad y hegemonía de una familia. Con estas realizaciones, se adelantan y superan a competidores, concurrentes, adversarios.

Consideramos que el principal reto y aportación de este proyecto reside en su dimensión digital. Ante un corpus documental tan vasto y de difícil acceso para el público no especializado, hemos desarrollado la plataforma www.scriptamanent.info. Este recurso no es un mero repositorio, sino una herramienta de investigación y difusión diseñada para convertir registros privados y custodiados en textos públicos, accesibles y abordables.

La plataforma se articula en varios ejes:

1. Edición y accesibilidad: se publican transcripciones paleográficas y ediciones críticas a texto completo de los documentos que integran la base de datos de Diplomas (https://www.scriptamanent.info/diplomas/), en formato PDF y para descarga gratuita que complementan las descripciones archivísticas y la digitalización iniciada por el Portal de Archivos Españoles (PARES). Actualmente, esta base de datos cuenta con diplomas de distinta forma, contenido y alcance, desde cartas con pocas palabras hasta textos normativos de gran extensión y complejidad; y esperamos aumentar la cifra de documentos en la recta final del proyecto gracias a las tesis doctorales y trabajos finales de máster defendidos por miembros del equipo en estos años (véase imagen 1).
2. Interoperabilidad y desarrollos paralelos: asociada a la base de datos principal hemos desarrollado otras herramientas digitales como la base de datos de Escribanos (https://www.scriptamanent.info/firmas-de-escribanos/, interoperable con Diplomas), la sección Mayorazgos (https://www.scriptamanent.info/mayorazgos/, con información contenida en HTML) y la página web *Rethos* (https://rethos.scriptamanent.info/, destinada al estudio de hospitales medievales y modernos que integra un visualizador cartográfico y dispone de base de datos propia). Finalmente, hemos experimentado la interoperabilidad con plataformas externas como el proyecto de prosopografía HILAME (https://www.hilame.info/, *Hidalgos. Labradoras.*

Mercaderes, Universidad del País Vasco). Toda la estructura digital ha sido diseñada con *software* libre.

3. Comunicación y difusión de los resultados: desde 2012 *Scripta manent* ha apostado por la transparencia y comunicación científicas como marca identificativa del proyecto, complementando las herramientas institucionales y las redes sociales. En este sentido, destacan las 200 noticias en el apartado *La comunicación*, la sección *Hoy comentamos* que fomenta el debate experto y el autoaprendizaje, e iniciativas como *WikiDamas en Común* cuya editatona fue incluida por Wikimedia España como una de sus actividades para el último Día de la Mujer-8M. Todo es reflejo del trabajo de un equipo conformado por investigadores y profesores del CSIC y varias universidades españolas (UCM, URJC, UAM, UPV/EHU, UPNA, ULL, UZ, UM) y extranjeras (U. París I, U. Parma, U. Palermo, U. Lincon) (véase imagen 2).

De cara al futuro, esta comunicación concluye con las limitaciones en el presente y los retos que afrontamos. ¿Qué objetivos se han cumplido y cuáles no? ¿Qué problemas hemos encontrado? ¿Cómo dotar a la cultura de proyectos española de sostenibilidad? ¿Cómo mantener vinculado a un equipo que a partir del año que viene se disgregará (saliendo de la Academia, incorporándose a otros proyectos, a empresas) dado nuestro sistema, precario para los más jóvenes? ¿Qué mejoras puede todavía asumir *Scripta manent* más allá del crecimiento de los datos? Estas son algunas sugerencias y respuestas a las preguntas anteriores:

1. Se han abierto nuevas vías de investigación que en sus inicios no estaban tan delineadas (bibliofilia, instituciones hospitalarias, perspectiva de género), aunque se ha profundizado poco en determinadas tipologías documentales fundamentales y que estaban en la hoja de ruta (compraventas, testamentos). Se han analizado con detenimiento aspectos como la memoria genealógica, la propiedad vinculada (mayorazgos), la organización del archivo y su mantenimiento durante siglos (inventarios, instrucciones archivísticas), lo que ha permitido radiografiar la cultura escrita de los Fernández de Velasco en múltiples planos. Y orientar, por tanto, nuevos enfoques sobre la nobleza.
2. Ha habido un alto número de integrantes del equipo con disimilitud en su régimen y dedicación (estudiantes de máster, becarios de investigación, contratados predoctorales y posdoctorales, profesores de universidad), no coincidentes a lo largo del tiempo, lo que ha dificultado el volcado sistemático de los datos y supuesto la necesidad de readaptar roles constantemente.
3. En relación con lo anterior, la variedad de tipos documentales y su amplitud cronológica ha complicado la asunción de estándares para la transcripción paleográfica y la edición digital, retrasando, más de lo deseable, el volcado de información en las bases de datos. En contraste positivo, la formación de los más jóvenes se ha beneficiado de esa

riqueza documental.

4. La localización de información en las bases de datos relacionales podría mejorarse si perfeccionamos las herramientas y motores de búsqueda, por ejemplo, con acceso al contenido en toda la integridad de cada texto alojado. Una perspectiva así supondría transformar las versiones actuales (en PDF) mediante el uso de otros tratamientos y aplicaciones estándares de XML, como, por ejemplo, TEI.
5. La cultura de ciencia abierta, progresivamente implementada en el ámbito de la accesibilidad a la bibliografía, debe extenderse a otros ámbitos de las humanidades, a saber, la política de los datos, su FAIRificación y el diálogo con estructuras europeas. Esta exploración es necesaria y hemos abierto iniciativas en este sentido.
6. La sostenibilidad económica e institucional pasa por la formación de las nuevas generaciones de humanistas, por la incorporación de estas personas estén o no en las universidades (archivos, bibliotecas, institutos, empresas), y por el colaboracionismo (desde asociaciones o similares). Estos desafíos se están materializando con la creación de *Humanidades en Común* (https://humanidadesencomun.eu/), un dispositivo de ciencia abierta con proyectos de investigación (en curso o terminados) con un fuerte componente digital y una cultura de trabajo compartida, en el que nos integramos.
7. La vinculación personal de cada uno de los miembros del equipo quizás sólo pueda mantenerse gracias a las experiencias vividas durante el proyecto, gracias al trabajo colaborativo, el aprendizaje y la independencia que han tenido ellas y ellos para hacer suyo *Scripta manent*.

Figuras, tablas, mapas o gráficas

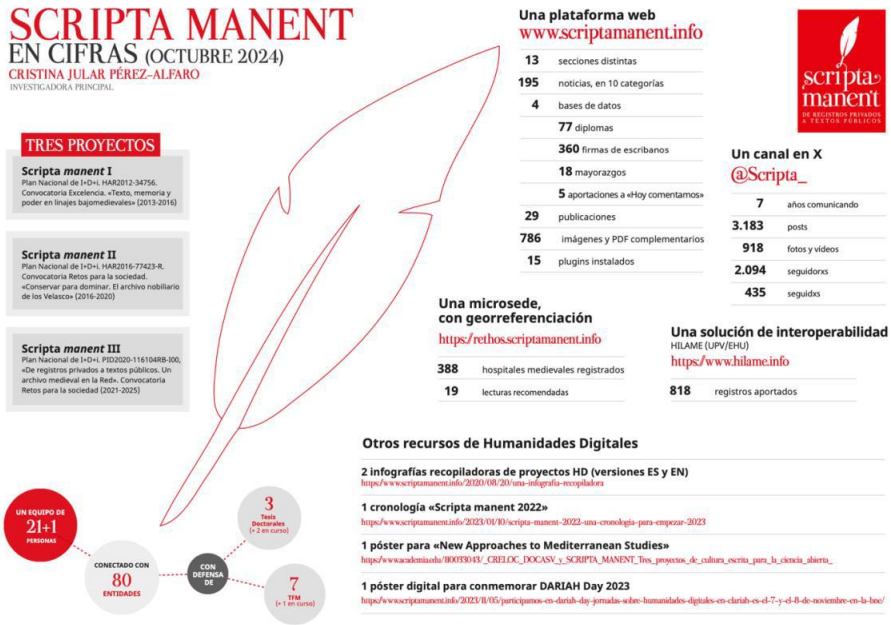


Imagen 1. Resumen en cifras del proyecto *Scripta manent*.

Imagen 2. Equipo de los proyectos *Scripta manent* I, II y III desde 2012 hasta 2025.

Bibliografía

JULAR PÉREZ-ALFARO, C. (2025), «From coffer to byte: An online medieval archive», en R. J. GONZÁLEZ ZALACAIN & G. VAAMONDE (eds.), *Digital Humanities in Medieval and Early Modern Spanish Texts. Current Perspectives and Approaches*, Abingdon / New York, Routledge.

JULAR PÉREZ-ALFARO, C. (2024), «Linajes de papel: propuestas digitales para un archivo nobiliario en la Red», en M. A. RUSSO (a cura di), *Pluma, papel, tijera: fuentes y escrituras para la historia de la nobleza bajomedieva (entre Sicilia y la península ibérica)*, Palermo, Palermo University Press, pp. 21–52.

JULAR PÉREZ-ALFARO, C. (2021), «¿Qué archivos para qué señor? Los primeros inventarios del linaje Velasco (Castilla, siglos XIV–XV)», en V. LAMAZOU-DUPLAN (ed.), *Les archives familiales dans l'Occident médiéval et moderne: trésor, arsenal, mémorial*, Madrid, Casa de Velázquez, pp. 105–115.

JULAR PÉREZ-ALFARO, C., & VILLAGRASA-ELÍAS, R. (2020), «Scripta manent. Conservar para dominar: el archivo nobiliario de los Velasco. De registros privados a textos públicos», en *La historiografía medieval en España y la conformación de equipos de trabajo: los proyectos de investigación I+D+i*, Murcia, SEEM, pp. 113–117.

VIÑUALES FERREIRO, G. (2022), «Lo escrito, escrito está, pero mejor en soporte digital: el proyecto Scripta Manent II y las humanidades digitales», en P. MARTÍNEZ GARCÍA (coord.), *El pasado en la pantalla: Cine y Humanidades Digitales*, Murcia, Universidad de Murcia, pp. 87–98.

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS (CSIC) (s. f.), *Libros blancos del CSIC. Volumen 10: Digital & Complex Information*,

recurso en línea. https://desafios2030.csic.es/informacion-compleja-y-digital/
— *Challenge 7: Digital Humanities*, coords. A. Lafuente & J. Farré, pp. 129–145.
— *Challenge 8: Digital Citizenship*, coords. A. Corsín Jiménez & A. Wagner, pp. 146–167.

Pablo Picasso y su red: un análisis del ecosistema cultural del siglo XX

Fatiha Idmhand
Université de Poitiers

El archivo de Pablo Picasso, albergado en el Museo Nacional Picasso-París, conserva, entre otros documentos, una amplia correspondencia compuesta por más de 30.000 cartas que fueron enviadas a pablo Picasso por aproximadamente 10.000 remitentes a personas físicas, entidades jurídicas, instituciones y asociaciones a lo largo de casi setenta años.

El corpus epistolar de este artista, una de las figuras más emblemáticas del siglo XX, representa una fuente de enorme valor para investigar sus interacciones personales, profesionales y políticas. Sin embargo, a pesar de su relevancia, esta correspondencia ha permanecido en gran medida inexplorada, y solo una mínima parte ha sido objeto de publicación o estudio sistemático. Como consecuencia, la naturaleza precisa de la red de relaciones de Picasso sigue siendo en buena medida desconocida, y hasta la fecha no se ha elaborado una cartografía exhaustiva de dicha red. Un estudio exhaustivo de esta correspondencia permitiría no solo reconstruir y cartografiar la red de vínculos de Picasso, sino también visibilizar las relaciones valiosas y duraderas que mantuvo con el mundo hispánico —por ejemplo, con artistas, intelectuales y exiliados—, vínculos que la crítica internacional ha tendido a subestimar frente a sus conexiones con figuras destacadas del ámbito francófono o anglófono.

El proyecto CARTAS - Pablo Picasso en toutes lettres tiene como objetivo principal subsanar estas lagunas mediante una investigación profunda de la correspondencia conservada en el archivo Picasso de París (correspondencia pasiva). Esta iniciativa busca no solo reconstruir y visualizar la red de vínculos del artista, sino también ofrecer nuevas claves para comprender la dimensión relacional de su trayectoria y su papel dentro del ecosistema cultural del siglo XX. La hipótesis central de esta investigación es que un análisis detallado del conjunto de la red picassiana puede enriquecer significativamente nuestra comprensión de las dinámicas culturales y políticas que moldearon tanto su obra como su proyección internacional a lo largo del siglo pasado.

El análisis integral de este corpus plantea varios desafíos: la magnitud del archivo, la heterogeneidad de los remitentes y la diversidad temática de las cartas. Estos retos serán abordados en esta comunicación, que se estructura en tres ejes: 1) la recolección y organización de metadatos, 2) la cartografía de la red, y 3) el análisis de las relaciones clave, con el objetivo de evaluar el impacto de determinadas conexiones (los mediadores culturales) en la trayectoria

personal y profesional de Picasso, así como la circulación y la influencia de sus ideas en el ámbito hispánico.

Bibliografía

AMIS (Advanced Metadata Intelligent System) (s. f.), recurso en línea. https://oscars-project.eu/projects/amis-advanced-metadata-intelligent-system

CONSORTIUM CAHIER, Groupe Typologies Textuelles (s. f.), recurso en línea. https://cahier.hypotheses.org/activites/groupe-typologie-textuelle

GALLERON, I., & IDMHAND, F. (2020b), «Why go from texts to data, or the Digital Humanities as a critique of the Humanities», *Word and Text: A Journal of Literary Studies and Linguistics*, X, pp. 53–69.

IDMHAND, F., & GALLERON, I. (2020a), *Guide pour la FAIRisation des données des corpus d'auteurs* [Groupe de travail Data_Cahier], Huma-Num.

MARTÍN-RODILLA, P., & GONZÁLEZ-PÉREZ, C. (2017), «A modelling language for discourse analysis in Humanities: Definition, design, validation and first experiences», *Revista de Humanidades Digitales*, 1, pp. 368–378. https://doi.org/10.5944/rhd.vol.1.2017.16133

Búsqueda por contenido en bases de datos federadas: una nueva funcionalidad para ASODAT[1]

Ariadna Fuertes Seder
Jaime Yefi Verdugo
Universitat de València

Palabras clave: Bases de datos federadas; datos no estructurados; JSON; Aplication Program Interface
Keywords: Federated databases; unstructured data; JSON; Application Program Interface

La integración y explotación de datos en las Humanidades Digitales sigue enfrentando un reto central: la fragmentación de la información en repositorios independientes que dificultan una visión transversal, contextualizada y significativa del conocimiento. En el ámbito del teatro clásico español del Siglo de Oro, el proyecto ASODAT aborda esta problemática articulando un sistema federado de bases de datos que permite interconectar y consultar, de forma estructurada, la información especializada contenida en cada una de ellas presentando los resultados de manera conjunta.

Como parte de este ecosistema, se ha desarrollado una nueva funcionalidad de búsqueda por contenido, concebida para facilitar la exploración federada de los datos integrados en diversas bases de datos especializadas que constituyen ASODAT. Esta funcionalidad no solo permite realizar búsquedas textuales en campos significativos, sino que organiza y presenta los resultados de manera contextualizada, favoreciendo el análisis, la comparación y la interpretación crítica de la información recuperada.

El sistema parte de una búsqueda literal en los campos más relevantes de cada base, pero su valor añadido radica en la manera en que agrupa los resultados por obra literaria, permitiendo así una navegación coherente a partir del eje interpretativo central del entorno ASODAT: la obra literaria. Cada resultado muestra el término buscado, el contexto de aparición, el campo donde ha sido localizado y la base de datos de origen. La interfaz ofrece enlaces directos a la ficha centralizada de la obra en ASODAT, lo que a su vez nos permite también acceder al registro original en cada base federada. En el nuevo buscador por contenido, se permite elegir las bases de datos de las que interesa obtener información de una manera sencilla.

[1] Con la ayuda de los proyectos con referencia PID2022-136431NB-C61, PID2022-136431NB-C62, PID2022-136431NB-C65 y PID2022-136431NB-C66 (MCIU/AEI/FEDER, EU)

Esta funcionalidad responde a una necesidad metodológica clave en el análisis de datos en humanidades: no solo localizar información relevante, sino hacerlo de manera relacional y significativa. La organización por obra permite reunir bajo un mismo nodo datos dispersos en distintas fuentes —desde noticias de representación hasta referencias a editores o libreros, registros censales, elementos musicales u otros metadatos— y establecer vínculos que enriquecen la comprensión del fenómeno teatral desde una perspectiva transversal.

Desde el punto de vista técnico, la herramienta ha sido diseñada para optimizar tanto la eficiencia en la recuperación de datos como la accesibilidad a los usuarios, y su arquitectura está preparada para futuras extensiones que permitirán ampliar su alcance y refinar los resultados.

En esta contribución queremos presentar el análisis realizado para la extracción de la información y el tratamiento llevado a cabo para para ofrecer la visualización de los datos, ya que representa un modelo de explotación estructurada de información compleja, basado en criterios semánticos y orientado a la generación de nuevas preguntas de investigación. No se trata de una herramienta de búsqueda generalista, sino de una infraestructura de acceso inteligente al conocimiento, que trata de poner en valor el trabajo filológico, documental y bibliográfico realizado por los grupos de trabajo que forman parte del proyecto y que han alimentado las bases integradas en ASODAT.

El objetivo de nuestra intervención consiste en explicar cómo se ha abordado técnicamente el desarrollo del sistema para hacer frente a los desafíos asociados a la consolidación de la información, lo que ha incluido la revisión de los modelos de datos existentes en cada punto de integración, la generación de las consultas optimizadas de extracción (SQL) y la agrupación de los datos en formato no estructurado (JSON). Además, se describe cómo se organiza y presenta la información en el sistema de manera que ofrezca una estructura intuitiva al usuario. Se exponen las acciones previstas para mejorar los tiempos de respuesta y el marco de trabajo en futuras integraciones, con el objetivo de incorporar nuevos conceptos de búsqueda dentro del contenido de cada uno de los sistemas federados y enmarcarlos dentro del esquema de solución ya implementado.

Este buscador encarna de forma concreta el lema del congreso democratizando el acceso al patrimonio textual, facilitando nuevas formas de análisis colaborativo y crítico sobre corpus amplios y estructurados, y fomentando una actitud activa ante el conocimiento, donde el usuario no solo accede a datos, sino que, al presentarlos conjuntamente, puede reinterpretarlos, relacionarlos y proyectarlos en nuevas líneas de investigación.

En este sentido, la funcionalidad no es solo una mejora técnica, sino que permite pasar de una consulta puntual a una lectura relacional de la información, y propone una forma de visualizar el conocimiento que, más allá de lo cuantitativo, atiende a lo contextual, lo situado y lo significativo.

Asimismo, desde una perspectiva epistemológica, esta funcionalidad plantea una crítica implícita a los modelos tradicionales de búsqueda,

generalmente centrados en la base de datos como unidad cerrada. ASODAT propone, en cambio, un modelo federado y expansivo, donde cada base mantiene su identidad, pero contribuye a una red de significación mayor. Esta arquitectura refuerza la idea de que el conocimiento en humanidades no debe leerse de forma aislada, sino como un tejido interconectado de datos.

Finalmente, cabe señalar que, al igual que lo es ASODAT, esta herramienta constituye un ejemplo de buenas prácticas en el desarrollo de infraestructuras digitales puesto que es sostenible y escalable, gracias al protocolo de integración que permite incorporar nuevas bases sin romper la coherencia del sistema y por tanto incluirlas en las búsquedas; es documentada y replicable, lo que abre la posibilidad de aplicar su lógica a otros campos del patrimonio cultural; y es comprometida con la investigación abierta, ya que ofrece sus resultados de forma transparente y accesible.

En conclusión, con la nueva funcionalidad de búsqueda por contenido desarrollada, se enriquece el proyecto ASODAT ya que nos permite indagar en la información desde una perspectiva renovada, relacional y crítica. Creemos que, con este nuevo buscador, dotamos al proyecto de una herramienta que no solo facilita el acceso a los datos, sino que se obtiene un avance significativo en la forma en que las humanidades digitales pueden integrar, representar y hacer operativos grandes volúmenes de datos culturales. Su diseño no solo mejora la experiencia de búsqueda, sino que activa el potencial interpretativo del usuario, invitándolo a reconstruir significados y a crear conocimiento nuevo a partir de la articulación contextualizada de la información.

Bibliografía

ALIA, N. A., YUSOF, M. K., & SAFEI, S. (2024), «An efficiency of native JSON+ for data integration», *IAENG International Journal of Computer Science*, 51(3), pp. 301–307. https://www.iaeng.org/IJCS/issues_v51/issue_3/IJCS_51_3_16.pdf

FERRER VALLS, T., & FUERTES SEDER, A. (2023), «Fundamentos y ampliación de la federación de bases de datos sobre teatro clásico español ASODAT (segunda fase)», *Hipogrifo*, 11(1), pp. 15–37. https://doi.org/10.13035/H.2023.11.01.03

FERRER VALLS, T., FUERTES SEDER, A., PEÑA ORTIZ, R., GARCÍA-REIDY, A., JOSA, L., & URZÁIZ TORTAJADA, H. (2021), «ASODAT: una plataforma de información sobre el teatro clásico español a partir de bases de datos federadas», *Talia. Revista de Estudios Teatrales*, 3, pp. 45–57. https://doi.org/10.5209/tret.70171

FERRER VALLS, T., BADÍA HERRERA, J., FUERTES SEDER, A., GARCÍA-REIDY, A., GREER, M. R., JOSA, L., & URZÁIZ, H. (2017), «Hacia una integración de bases de datos sobre el teatro clásico español: el proyecto ASODAT», en N. RODRÍGUEZ ORTEGA (ed.), *III Congreso de la Sociedad Internacional Humanidades Digitales Hispánicas. Sociedades, políticas,*

saberes. Libro de resúmenes, Málaga, Universidad de Málaga, pp. 114–117. https://humanidadesdigitaleshispanicas.es/wp-content/uploads/2020/03/Actas-HDH2017.pdf

LLANO-RÍOS, T. F. (2024), «Using dynamic schemas for query optimization over JSON data», *Electronic Theses and Dissertations*, Paper 4308. https://doi.org/10.18297/etd/4308

PIECH, M., & MARCJAN, R. (2018), «A new approach to storing dynamic data in relational databases using JSON», *Computer Science*, 19(1), art. 3. https://doi.org/10.7494/csci.2018.19.1.2505

YUSOF, M. K. (2017), «Efficiency of JSON for data retrieval in big data», *Indonesian Journal of Electrical Engineering and Computer Science*, 7(1), pp. 250–262. https://doi.org/10.11591/ijeecs.v7.i1.pp250-262

Nuevas posibilidades de exploración filológica para el teatro áureo: búsqueda por contenido en ASODAT[1]

Alejandro García Reidy
Universidad de Salamanca
Teresa Ferrer Valls
Josefa Badía Herrera
Jesús Tronch Pérez
Universitat de València
Alejandra Ulla Lorenzo
Universidade de Santiago de Compostela
Lola Josa González
Universitat de Barcelona
Héctor Urzáiz Tortajada
Universidad de Valladolid

Palabras clave: Teatro de los Siglos de Oro; bases de datos federadas; datos no estructurados; metabuscador.
Keywords: Theater of the Golden Ages; federated databases; unstructured data; metasearch engine.

La significativa cantidad de información que se conserva relacionada con el teatro español de los Siglos de Oro, tanto en su faceta de corpus textual como en su condición de práctica dramática y cultural, es una de las grandes ventajas y una de las grandes dificultades a las que se enfrentan los investigadores de este campo de estudio. Por un lado, contamos con miles de obras de todos los géneros dramáticos representados durante los siglos XVI y XVII que se han preservado a través de impresos y de manuscritos, así como con miles de documentos que documentan la actividad teatral de los profesionales de la escena (dramaturgos, representantes, músicos, bailarines, etc.). Por otro lado, esta ingente cantidad de información (sobre todo en comparación con lo preservado en otras prácticas dramáticas europeas coetáneas) ha supuesto una complicación para poder contar con visiones de conjunto fundamentadas en un corpus amplio (y no limitado a un número limitado de textos o documentos manejables por un único investigador).

Las posibilidades ofrecidas por las metodologías de las Humanidades Digitales han permitido avanzar en la resolución de estas limitaciones

[1] Con la ayuda de los proyectos con referencia PID2022-136431NB-C61, PID2022-136431NB-C62, PID2022-136431NB-C65 y PID2022-136431NB-C66 (MCIU/AEI/FEDER, EU)

cuantitativas y cualitativas, especialmente con el desarrollo de bases de datos que se especializan en sistematizar filológicamente información sobre diversas facetas de este teatro áureo (testimonios, música, censuras, actividad de actores, ediciones y características de obras dramáticas, etc.). El proyecto ASODAT (Ferrer Valls *et al.*, 2017; Ferrer Valls *et al.*, 2021; Ferrer Valls y Seder Fuertes, 2023), con el continuo trabajo de federación de varias de estas bases de datos (actualmente, nueve de ellas: CATCOM, DICAT, DIGITAL MÚSICA POÉTICA, CLEMIT, MANOS, ISTAE, ARTELOPE, EMOTHE y ETSO) que lleva a cabo desde 2016, permite interconectar y consultar datos procedentes de cada una de ellas y ofrecer así un punto de partida para los investigadores. El objetivo de esta comunicación es la de presentar algunas posibilidades de investigación filológica que ofrece una nueva funcionalidad de búsqueda de contenido conjunto de seis de estas bases de datos (según las posibilidades ofrecidas por las características técnicas de esta federación), funcionalidad que se ha desarrollado en el primer trimestre de 2025.

Esta nueva búsqueda por contenido ofrece al investigador la posibilidad de hacer una búsqueda literal en los campos más relevantes (por tipo de información disponible) de las seis bases de datos disponibles, cuyos resultados se agrupan por obra literaria, que es el elemento que vertebra la información de ASODAT. La utilidad de este tipo de metabuscadores ha demostrado su utilidad en otras redes de investigación en torno a la literatura de los Siglos de Oro (Pena Sueiro y Saavedra Places, 2019). Los resultados que genera esta búsqueda de contenido permiten dos acercamientos distintos a datos en la federación. Por un lado, facilita recuperar información de ese gran archivo del teatro áureo en continuo crecimiento que es ASODAT, que hasta ahora estaba todavía disponible solo a través de cada base de datos individual. Esto facilita al investigador la posibilidad de visibilizar datos relevantes para su investigación de manera más inmediata y estructurada por obra dramática, lo que ayuda a abordar el problema del escalado de datos literarios masivos (Basu, 2025). En esta primera versión del buscador se ofrece la posibilidad de hacer búsquedas manuales dentro de la página de resultados para ir identificando casos concretos que interese analizar más detenidamente. Los investigadores cuentan así con una nueva herramienta que, por su condición de metabuscador, expande las posibilidades que recuperación de información que pueda ayudar a solventar la cuestión filológica que interese responder o analizar. La interfaz de los resultados ofrece un enlace al registro de cada obra con resultados en ASODAT, desde donde se puede acceder a la información más detallada y específica de cada base de datos que contenga información sobre cada pieza. La posibilidad de búsqueda de contenida permite, pues, contar con un nivel adicional que conexión de datos entre diferentes bases de datos sin renunciar a las vías de indagación filológica que facilita la estructura de registro federado en ASODAT.

Por otro lado, una federación obliga a vincular datos a partir de una organización definida de antemano. En el caso de ASODAT, como ya se ha señalado, esta vinculación se lleva a cabo a partir de la unidad de la obra

dramática. La posibilidad de llevar a cabo búsquedas amplias por diversos campos de contenido permite generar nuevas redes de información más allá de esta yuxtaposición previa. Es decir, al buscar los investigadores desde un contenido más amplio, el nuevo tipo de resultado, menos formalizado de antemano, puede contribuir a generar nuevas preguntas de investigación. Esta nueva herramienta abre nuevas opciones en las fases iniciales de la investigación filológica al facilitar procesos de exploración de datos más amplios y sistemáticos. Los investigadores pueden comenzar su trabajo con consultas generales que le permitan mapear el territorio de información disponible antes de desarrollar hipótesis específicas, invirtiendo el orden tradicional que privilegiaba la formulación previa de preguntas específicas. En otras palabras, este nuevo buscador de contenido es una muestra de cómo una epistemología digital (Ingvarsson, 2021) obliga a reestructurar un nuevo orden de datos a partir de principios diferentes a los que rigen los resultados que hasta el momento ofrecía ASODAT.

Para ejemplificar la utilidad y posibilidades de este nuevo buscador de contenido, analizaremos algunos casos de resultados, especialmente los centrados en espacios geográficos y con particular atención al caso de la ciudad de “Valencia”. Por una parte, se analizarán las principales deducciones que se pueden hacer de los resultados en bruto que actualmente ofrece la información federada en ASODAT. Así, podremos vincular datos sobre la actividad teatral en la ciudad, tanto en su corral de comedias como en palacio (sin olvidar el problema de las representaciones en espacios que no se especifican en la documentación); los dramaturgos canónicos y exocanónicos cuyas obras se vieron en los escenarios; la naturaleza de las compañías que más representaron en la ciudad; la presencia de poetas locales entre las piezas representadas; la relación con la actividad impresora en la ciudad a través de comedias sueltas; y si es posible detectar vínculos entre el paso de determinadas formaciones teatrales con la posterior publicación de obras de su repertorio en la ciudad. Por otra parte, este análisis se comparará con resultados de otros lugares periféricos respecto del foco teatral de la Corte, pero que también tuvieron, como Valencia, una importante actividad teatral: será el caso de localidades como Sevilla, Salamanca o Toledo. Para llevar a cabo esta comparación se procederá no solo con la lente filológica analógica, sino que, por el volumen textual que pueden llegar a generar los resultados de búsquedas de contenido en ASODAT, se experimentará con las posibilidades interpretativas de estos datos en bruto que ofrecen diversos Grandes Modelos de Lenguaje (Gemini, Claude, ChatGPT) a partir de preguntas de interés filológico. De esta manera se explorarán si actualmente estas herramientas pueden constituir un apoyo eficaz para explotar de manera útil para los especialistas, datos sobre la actividad teatral reunidos por ASODAT, con especial énfasis en sus posibilidades (la rapidez con la puede llevar a cabo ciertos tipos de análisis cuantitativos bien delimitados por las instrucciones de los investigadores) y limitaciones (el problema de las alucinaciones o la superficialidad de algunas de las aproximaciones

supuestamente más cualitativas de los datos). Como se mostrará, los resultados obtenidos muestran que estas ciudades presentan unas tendencias semejantes en cuanto al peso de los grandes dramaturgos de éxito en los teatros comerciales (y su importancia relativa en el contexto específico de la documentación conservada), a la vez que permite detectar algunas características específicas (por ejemplo, una cierta presencia de dramaturgos locales en la actividad de ciudades como Valencia o Sevilla), así como limitaciones de las fuentes disponibles para cada localidad.

Por último, la mirada panorámica o distante (por aludir a Franco Moretti) que posibilita una federación de bases de datos como ASODAT y una búsqueda de contenido facilita el ejercicio de focalización más específica. A su vez, la epistemología digital revaloriza y resemantiza lo material y concreto. Como muestra final de las posibilidades que ofrece el nuevo buscador, se dará noticia de una obra, *El fuerte, animoso, sagaz y valiente Martín López de Aibar, valenciano: tiniente de maese de campo general en los estados de Flandes*, escrita por el dramaturgo valenciano Vicente Ezquerdo y estrenada por la compañía de Pedro de Valdés en la capital del Turia en junio de 1620, de la que se tenía noticia desde el catálogo de La Barrera (1968), pero que actualmente se consideraba perdida. La conexión de datos y su exploración y explotación que facilita ASODAT ha contribuido a localizar el supuesto manuscrito autógrafo de esta comedia, así como una transcripción de su texto llevada a cabo a finales del siglo XIX por un poseedor del original.

En conclusión, esta comunicación busca mostrar las posibilidades que ya ofrece la nueva búsqueda de contenido conjunto implementada en ASODAT, pues constituye una mejora metodológica para la investigación del teatro áureo español. Esta funcionalidad permite consultas simultáneas en múltiples bases de datos especializadas, que permite partir de hipótesis específicas, pero también anima a los investigadores a explorar el conjunto completo de datos disponibles antes de formular preguntas de investigación e identificar conexiones no previstas entre información dispersa. Este sistema de búsqueda genera nuevas posibilidades de análisis que trascienden la organización habitual por obra dramática, lo que permite la identificación de nuevos patrones. Los resultados obtenidos a partir del análisis del espacio geográfico de Valencia y su comparación con los resultados de otros lugares similares confirman la utilidad de esta herramienta y las posibilidades de nuevas formas de exploración y explotación de los datos contenidos en la federación ASODAT.

Bibliografía

BASU, A. (2025), *Shakespeare and Scale: The Archive of Early Printed English*, Cambridge, Cambridge University Press. https://doi.org/10.1017/9781009306676

FERRER VALLS, T., & FUERTES SEDER, A. (2023), «Fundamentos y ampliación de la federación de bases de datos sobre teatro clásico español

ASODAT (segunda fase)», *Hipogrifo*, 11(1), pp. 15–37. https://doi.org/10.13035/H.2023.11.01.03

FERRER VALLS, T., BADÍA HERRERA, J., FUERTES SEDER, A., GARCÍA-REIDY, A., GREER, M. R., JOSA, L., & URZÁIZ, H. (2017), «Hacia una integración de bases de datos sobre el teatro clásico español: el proyecto ASODAT», en N. RODRÍGUEZ ORTEGA (ed.), *III Congreso de la Sociedad Internacional Humanidades Digitales Hispánicas. Sociedades, políticas, saberes. Libro de resúmenes*, Málaga, Universidad de Málaga, pp. 114–117. https://humanidadesdigitaleshispanicas.es/wp-content/uploads/2020/03/Actas-HDH2017.pdf

FERRER VALLS, T., FUERTES SEDER, A., PEÑA ORTIZ, R., GARCÍA-REIDY, A., JOSA, L., & URZÁIZ TORTAJADA, H. (2021), «ASODAT: una plataforma de información sobre el teatro clásico español a partir de bases de datos federadas», *Talia. Revista de Estudios Teatrales*, 3, pp. 45–57. https://doi.org/10.5209/tret.70171

INGVARSSON, J. (2021), *Towards a Digital Epistemology: Aesthetics and Modes of Thought in Early Modernity and the Present Age* (2nd ed.), Palgrave Macmillan. https://doi.org/10.1007/978-3-030-78724-0

LA BARRERA Y LEIRADO, C. A. de (1860), *Catálogo bibliográfico y biográfico del teatro antiguo español: desde sus orígenes hasta mediados del siglo XVIII*, Madrid, Imprenta de M. Rivadeneyra; ed. facsimilar, Londres, Tamesis Books, 1968.

PENA SUEIRO, N., & SAAVEDRA PLACES, Á. (2019), «Aracne. Red de Humanidades Digitales y Letras Hispánicas», *Historias Fingidas*, 7, pp. 407–412. http://dx.doi.org/10.13136/2284-2667/149

Hacia el reconocimiento automático de tipos impresos antiguos: el proyecto RETIPIA

Paula Casariego Castiñeira
Università Roma Tre
Alejandra Ulla Lorenzo
Universidade de Santiago de Compostela

El teatro áureo español contó con una amplia y rica difusión impresa que se llevó a cabo, principalmente, mediante dos géneros editoriales: las partes de comedias y los impresos sueltos, objeto de estudio del proyecto ISTAE, que forma parte del más amplio ASODAT. En el caso de la Península Ibérica y a diferencia de otras áreas geográficas de Europa, este tipo de impreso de carácter popular usó el formato cuarto y, como denota el excepcional volumen de sueltas conservadas en la actualidad -único en el panorama europeo-, fue una de las vías de difusión impresa del teatro del Siglo de Oro más relevantes. A pesar de que sus orígenes se remontan a principios del siglo XVI, este género editorial del impreso suelto proliferó en la península ibérica durante los siglos XVII y XVIII, gracias al bajo coste de su impresión por la escasez de calidad del papel y de la tinta empleada, y, de especial forma en el siglo XVII, a la facilidad para sortear las imposiciones legales que exigía la publicación de las partes de comedias en los diferentes reinos de la Corona. A pesar de los esfuerzos por controlar la publicación de libros, muchos libros, y los impresos sueltos teatrales en particular, escapaban sin dificultad a los controles legales y daban a la estampa, principalmente durante el siglo XVII, un elevadísimo número de impresos sin datos editoriales o con datos editoriales falsos (Cruickshank 1989, 1991).

Junto a otros problemas particularmente acusados en los impresos sueltos teatrales -como la falsificación de las autorías o la diversidad de títulos para un mismo texto-, la ausencia de la correcta identificación bibliográfica de este amplio conjunto de testimonios supone uno de los grandes retos del estudio del patrimonio teatral áureo (Cruickshank 1985; Ulla Lorenzo 2020a). Esta falta tiene graves consecuencias en el conocimiento del mercado librero del siglo XVII y en el papel de este tipo de impresos en la circulación y difusión de las obras teatrales, a la par que mantiene en la oscuridad una gran parte de la actividad de los impresores que trabajaron en la península, pues, al no poder asociarse estos impresos sueltos sin una correcta información bibliográfica a un taller concreto, no es posible la reconstrucción completa de la producción de los impresores.

El control del patrimonio bibliográfico del teatro del Siglo de Oro requiere, por tanto, de la correcta localización de cada uno de los testimonios en los que

fue impreso y del examen material y tipobibliográfico, vías principales para arrojar luz sobre los ausentes o falsos datos editoriales. A pesar de su relevancia, son pocos los estudios que se han dedicado al examen material y tipobibliográfico de las sueltas, quizá por la escasa atención que se le ha prestado a este tipo de publicación de carácter popular, quizá por la elevada complejidad de la operación de hipotetizar sobre las circunstancias y el momento de edición de una suelta sin datos editoriales de los siglos XVII y XVIII. La técnica empleada para tal operación, que sigue de cerca la empleada en el ámbito paleográfico y para el análisis de los incunables, se basa en el análisis pormenorizado de la tipografía y -si la tiene- de la ornamentación de cada impreso suelto, cuyos resultados deben confrontarse con la tipografía y la ornamentación de impresos con datos bibliográficos veraces e indubitados de todos los impresores conocidos, y debe relacionarse, siempre que sea posible, con otros impresos sueltos con los que pueda formar una familia editorial. A este estudio se añade, en este sentido, el del patrón de composición del impreso y, en particular, el de su portada, cuyos rasgos son susceptibles de vincularse a un impresor determinado y, en la mejor de las situaciones, posibilitan crear una red de impresos sueltos que respondan a unas mismas características, con el objetivo de reconstruir la familia editorial que un impresor imprimió en un momento determinado de su carrera (Wilson y Cruickshank 1980) y de inscribirlas en una red para que, a medida que crezca, arroje luz sobre las dinámicas de publicación de los textos teatrales en las imprentas auriseculares. Las claves que aplicamos hoy al caso de los impresos sueltos teatrales del Siglo de Oro se apoyan, en primera instancia, en los trabajos de Wilson (1973) y Wilson y Cruickshank (1980).

Las dificultades afrontadas en ISTAE y su experiencia en el manejo de un gran número de testimonios impresos y en la identificación de los datos bibliográficos ausentes o falsos en impresos del siglo XVII, haciendo uso de métodos tradicionales y filológicos como los descritos, han dado pie a repensar este problema desde otras ópticas y disciplinas. Basándonos en la idea de que cada tipo de cada letra puede ser procesado como una imagen y que esta puede ser reconocida automáticamente, hemos querido aplicar y conectar nuestro problema filológico con los objetivos e intereses de las disciplinas científicas del Big Data and Automatic Printed Text Recognition (APTR), ámbitos de investigación que llevan años explorando las formas en las que automatizar grandes cantidades de textos impresos multilingües, procedentes de todas las áreas geográficas y pertenecientes a todas las épocas. Nuestro problema y nuestros intereses, en concreto, están en la línea de los afrontados por el equipo liderado por Jean-Michel Morel y su *ATPR project*, formado por investigadores del análisis matemático del procesamiento de imágenes y la creación de nuevos algoritmos. A lo largo de su historia, este equipo de investigadores ha creado tecnología de vanguardia para el procesamiento y reconocimiento automático de imágenes, que ha sido empleada en numerosos *software* y agencias.

RETIPIA. Reconocimiento automático de Tipos Impresos Antiguos nace de la identificación de un problema filológico y bibliográfico de nuestro patrimonio

literario y de la búsqueda de una tecnología capaz de llevar adelante el procesamiento y reconocimiento automático de los tipos de imprenta antiguos empleados durante el periodo de la imprenta manual. Esta prueba de concepto, por tanto, someterá el problema filológico de los impresos sueltos teatrales a esta novedosa tecnología propuesta desde las matemáticas aplicadas por el equipo de matemáticos e ingenieros liderado por Jean-Michel Morel. Se partirá de la creación de un amplio corpus de impresos sueltos teatrales del siglo XVII. Este corpus será sometido a dicho método matemático con el objetivo de establecer, como primer paso, una red de relaciones ponderadas de semejanza tipográfica entre todas las sueltas del conjunto. Así, procesaremos y analizaremos cada tipo de letra como una imagen mediante algoritmos de análisis de forma y tecnologías de vectorización y, a continuación, las características de forma extraídas de los caracteres se normalizarán. Después, el método matemático calculará las distancias entre todas las posibles realizaciones del tipo en un corpus controlado de impresos sueltos y detectará los caracteres distinguibles mediante modelos probabilísticos paramétricos. A partir de estos datos, se creará dicha red de relaciones ponderadas de semejanza tipográfica. Esta red, creada a partir del reconocimiento automático de tipos impresos antiguos, será posible que los filólogos identifiquen desconocidas familias editoriales, las cuales, a su vez, serán susceptibles de ser vinculadas a talleres de imprenta o abrir nuevos caminos de investigación sobre conjuntos amplios de impresos sueltos teatrales.

En esta intervención, pretendemos, por tanto, plantear una metodología, hasta el momento no explorada para los impresos teatrales hispánicos y que parte de la filología y las matemáticas, para el procesamiento digital de la tipografía antigua y para la identificación automatizada de patrones y usos tipográficos que podrían ser aplicados a otros conjuntos de impresos sueltos teatrales.

Bibliografía

CASARIEGO, P. (2023), «Una imprenta para tres sueltas antiguas sin datos editoriales: *La gran Torre del Orbe*, *Los hermanos más amantes* y *La dama boba*», *Hipogrifo: Revista de literatura y cultura del Siglo de Oro*, 11(1), pp. 85–100.

CRUICKSHANK, D. W. (1985), «The editing of Spanish Golden-Age plays from early printed editions», en M. MCGAHA & F. P. CASA (eds.), *Editing the Comedia*, Ann Arbor, Michigan Romance Studies, pp. 52–103.

CRUICKSHANK, D. W. (1989), «Some notes on the printing of plays in seventeenth-century Seville», *The Library*, V, 9, pp. 231–252.

ULLA LORENZO, A., & CASARIEGO, P. (2022), «Sobre falsas atribuciones y duplicidad de títulos en el teatro calderoniano: *Haz bien y guárdate* o *La confusión de un jardín*», *Studia Aurea*, 16, pp. 229–259.

ULLA LORENZO, A. (2020a), «El problema bibliográfico de las comedias sueltas sin datos de impresión: apuntes para la identificación de algunas ediciones

financiadas por la librera Teresa de Guzmán», *Hipogrifo: Revista de literatura y cultura del Siglo de Oro*, 8(1), pp. 579–600.

WILSON, E. M., & CRUICKSHANK, D. W. (1980), *Samuel Pepy's Spanish Plays*, London, The Bibliographical Society.

WILSON, E. M. (1973), «Comedias sueltas—A bibliographical problem», en D. W. CRUICKSHANK & J. E. VAREY (eds.), *The Comedias of Calderón. Comedias, vol. I: The Textual Criticism of Calderón's Comedias*, London, Gregg International / Tamesis, pp. 211–219.

Procesamiento de datos prosódicos y sociolingüísticos del corpus PRESEEA-Valencia: generación de informes automatizados con R, Shiny y ChatGPT

Adrián Cabedo Nebot
Universitat de València

Palabras clave: Prosodia; inteligencia artificial; sociolingüística; Tidyverse; ChatGPT
Keywords: Prosody; artificial intelligence; sociolinguistics; Tidyverse; ChatGPT

Esta comunicación tiene como propósito fundamental analizar cómo el uso de inteligencia artificial, y en particular de modelos de lenguaje generativo como ChatGPT, introduce nuevas posibilidades metodológicas en el estudio de la prosodia dentro de contextos sociolingüísticos. Esta incorporación no solo responde a la necesidad de optimizar procesos tradicionalmente manuales y laboriosos, sino que también plantea una transformación epistemológica en la manera de acceder, procesar e interpretar los datos del habla. El estudio de la prosodia, entendida como el conjunto de elementos suprasegmentales del habla (entonación, ritmo, duración, pausas), ha dependido históricamente de la percepción humana y del análisis detallado mediante herramientas fonéticas como Praat (Boersma & Weenink, 2022), lo cual impone ciertos límites en cuanto a escalabilidad y objetividad.

La presente propuesta metodológica parte de una articulación entre inteligencia artificial y herramientas de código abierto desarrolladas en el entorno R, tales como Tidyverse, Shiny y Udpipe. Estas herramientas permiten automatizar el proceso de extracción de rasgos prosódicos y lingüísticos del corpus PRESEEA-Valencia, proporcionando así una vía más eficaz para analizar grandes volúmenes de datos. Este enfoque encuentra sustento en marcos teóricos consolidados, como el análisis prosódico de Cantero (2002), cuya influencia se manifiesta en la identificación y clasificación de unidades entonativas relevantes para el estudio del habla espontánea. Además, el uso de modelos de procesamiento automático del lenguaje natural (PLN) permite complementar los análisis acústicos con anotaciones morfosintácticas y con dimensiones semánticas como el análisis de sentimientos. De este modo, la metodología no solo agiliza la producción de datos, sino que promueve una visión integral del habla que contempla sus aspectos formales, funcionales y sociales desde una perspectiva computacionalmente reproducible.

Además del ahorro de tiempo y la eficiencia que ofrece el uso de inteligencia artificial en el análisis de datos prosódicos, este enfoque abre una

puerta significativa para la agilización del acceso a los estudios sociolingüísticos. Hasta ahora, este tipo de análisis solía requerir conocimientos técnicos avanzados, además de herramientas costosas o poco accesibles para investigadores sin una sólida formación computacional. Al integrar ChatGPT y herramientas abiertas del ecosistema R, esta metodología rebaja las barreras técnicas, haciendo posible que más investigadores, incluso aquellos con menos experiencia en programación, puedan explorar grandes volúmenes de datos y extraer patrones relevantes con facilidad. Esta accesibilidad también beneficia a equipos multidisciplinares que trabajan con datos del habla, facilitando una colaboración más fluida entre lingüistas, ingenieros y expertos en análisis del discurso.

Otro punto relevante es el potencial de este sistema para descubrir patrones ocultos que antes pasaban desapercibidos debido a limitaciones humanas en la percepción y el análisis manual. Por ejemplo, los matices emocionales que se expresan mediante la entonación o la intensidad del habla no siempre son detectables sin un análisis acústico detallado. Pero al procesar automáticamente variables acústicas y combinarlas con información morfosintáctica y sociolingüística, se pueden identificar tendencias que revelan cómo ciertos grupos sociales se expresan, se intensifican o, incluso, negocian su identidad lingüística. Esto da pie a un enfoque más empático y dinámico de la investigación del habla, uno que permite repensar cómo los contextos sociales influyen en la producción lingüística cotidiana, sin caer en simplificaciones o suposiciones previas.

El proceso parte del procesamiento automático de transcripciones de entrevistas sociolingüísticas y de la extracción de características entonativas (F0, intensidad, velocidad de habla, duración, entre otros). Estos datos acústicos se combinan con información morfológica y de sentimientos, con lo que se pretende observar perfiles lingüísticos diferenciados por variables sociales como clase, edad o género. Así pues, este enfoque permite comprender la distribución de dichos rasgos entre estratos sociales a partir de factores que no siempre son evidentes a nivel perceptivo.

El uso de la versión 1.6 del programa Oralstats (Cabedo 2025), que presentamos en esta comunicación, se ha desarrollado con R (R Core Team, 2023). Se apoya en Tidyverse para el procesamiento y visualización de datos, en la librería Shiny para construir una interfaz web interactiva, y, entre otros, en Udpipe para un primer etiquetado morfológico. Gracias a esta arquitectura de procesamiento, se logra analizar corpus extensos con gran rapidez y sin depender de programas de paso. Posteriormente, la introducción de ChatGPT en el entorno de aplicación supone un ahorro considerable en tiempo y permite generar informes sociolingüísticos automáticos a partir de las variables que el usuario selecciona.

La muestra analizada en este estudio comprende un total de 48 entrevistas seleccionadas cuidadosamente dentro del corpus PRESEEA-Valencia, restringidas al grupo de hablantes con un nivel de instrucción alto. La estructura

muestral se organizó a partir de un diseño balanceado que considera el cruce de dos variables sociales fundamentales: sexo (masculino y femenino) y edad (dividida en tres franjas: 20-35, 36-55 y más de 55 años), lo que da lugar a seis subgrupos homogéneos con ocho entrevistas cada uno. Esta segmentación no solo responde a criterios clásicos en sociolingüística, sino que también permite realizar comparaciones detalladas entre cohortes generacionales y de género dentro de un mismo nivel educativo. Esto aporta riqueza interpretativa al análisis de la variación lingüística.

Cada una de estas entrevistas fue procesada mediante la versión 1.6 de Oralstats, dentro de un entorno automatizado que integra múltiples capas de análisis lingüístico. Este sistema posibilita la identificación simultánea de patrones fonético-fonológicos (como entonación, duración o intensidad), morfosintácticos (categorías gramaticales, estructuras sintácticas) y emocionales (evaluaciones de sentimiento o carga afectiva), en relación con las variables sociales ya mencionadas. El valor de este procedimiento radica en su capacidad para detectar correlaciones complejas que, en un análisis tradicional, podrían pasar desapercibidas por su sutileza o por la gran cantidad de datos involucrados.

Este tipo de análisis retoma, a su vez, el enfoque pionero del proyecto Glissando (Garrido et al., 2013), que defendía la utilidad de combinar disciplinas para abordar la prosodia en lenguas románicas mediante sistemas de procesamiento automático. La propuesta que aquí se presenta no solo recoge esta herencia metodológica, sino que la proyecta hacia una nueva etapa en la que la inteligencia artificial, representada por modelos como ChatGPT, ocupa un lugar central en la arquitectura de análisis. A diferencia de sistemas anteriores que requerían una intervención humana constante para la interpretación de los datos, este modelo permite una automatización extensiva desde la extracción hasta la generación de informes interpretativos, sin que ello implique una pérdida de rigor o profundidad analítica.

La alineación con los principios planteados por Anthony (2020) en relación con la programación para la lingüística de corpus se manifiesta aquí en una aplicación concreta que amplía sus posibilidades mediante la incorporación de modelos generativos. El resultado es un enfoque metodológico que se caracteriza por su bajo coste computacional, su base en software libre reutilizable y su apertura a diferentes perfiles de usuarios. No se trata únicamente de una herramienta más eficiente, sino de un cambio en el modo de abordar la investigación sociolingüística, ya que elimina muchos de los filtros subjetivos que los investigadores podían imponer inconscientemente durante la observación y codificación manual. En definitiva, esta propuesta apunta a un análisis más transparente, reproducible y fiel a las complejidades reales del habla en uso.

Bibliografía

ANTHONY, L. (2020), «Programming for corpus linguistics», en M. PAQUOT & S. T. GRIES (eds.), *A Practical Handbook of Corpus Linguistics*, Springer, pp. 181–207. https://doi.org/10.1007/978-3-030-46216-1_9

BOERSMA, P., & WEENINK, D. (2022), *Praat: Doing Phonetics by Computer* (versión 6.2), recurso en línea. http://www.praat.org/

CANTERO, F. J. (2002), *Teoría y análisis de la entonación*, Edicions de la Universitat de Barcelona.

GARRIDO, J. M., ESCUDERO, D., AGUILAR, L., CARDEÑOSO, V., RODERO, E., DE-LA-MOTA, C., & GONZÁLEZ, C. (2013), «Glissando: A corpus for multidisciplinary prosodic studies in Spanish and Catalan», *Language Resources and Evaluation*, 47(4), pp. 945–971. https://doi.org/10.1007/s10579-012-9213-0

R CORE TEAM (2023), *R: A Language and Environment for Statistical Computing*, recurso en línea, R Foundation for Statistical Computing. https://www.R-project.org/

CHANG, W., CHENG, J., ALLAIRE, J. J., XIE, Y., & MCPHERSON, J. (2023), *Shiny: Web Application Framework for R*, recurso en línea. https://shiny.posit.co/

De la construcción a la exploración: *Humcor*, un corpus oral multimodal para el estudio del humor en español

Doina Repede
Universidad de Granada

Palabras clave: Humor; lingüística de corpus; *Humcor*; oralidad; España; Hispanoamérica
Keywords: Humor; corpus linguistics; *Humcor*; oral tradition; Spain; Latin America

El humor es un fenómeno multidimensional que involucra aspectos cognitivos, emocionales y sociales. Su estudio ha abarcado diversas disciplinas, como la psicología, la lingüística y la filosofía. Attardo (1994) define el humor como una forma de comunicación que desafía expectativas y genera incongruencias, provocando una respuesta humorística. Freud (1905) lo vincula con la liberación de tensiones psíquicas, mientras que Raskin (1985) propone la teoría del guion semántico del humor, explicándolo a través de la superposición de significados lingüísticos. Las investigaciones más recientes han analizado su papel en la interacción social y resaltan su función en la cohesión grupal y la construcción de identidades (Meyer, 2000). Como manifestación cultural, el humor no solo emerge dentro de una sociedad, sino que también la refleja, actuando como un espejo de sus valores y dinámicas sociales.

En los últimos años, el humor en español ha sido ampliamente estudiado, dando lugar a numerosas investigaciones. En particular, se han desarrollado propuestas de etiquetaje pragmático para su análisis (Ruiz Gurillo, 2023; Khaylina y Alvarado Ortega, 2023), junto con otros enfoques que exploran sus mecanismos discursivos y su impacto en la comunicación, como, por ejemplo, el análisis emocional de las columnas de humor (Vargas-Sierra, 2023), el léxico escatológico en el humor negro infantil (Jiménez Ruiz, 2023), o los estilos humorísticos en la comedia femenina (Linares Bernabéu, 2023), entre otros. Actualmente, el estudio del humor como competencia y actuación (Ruiz Gurillo, 2016, 2022), centrado en los procesos que intervienen en su producción e interpretación (Ruiz Gurillo, 2023), especialmente en el discurso oral, se ha consolidado como una rama de la lingüística del humor (ídem). Desde esta perspectiva, se han compilado corpus de textos humorísticos, como es el caso del portal *Observahumor* (https://www.observahumor.com/), cuyo objetivo general consiste en analizar la construcción de la identidad de mujeres y hombres a través del humor. Concretamente, se trata de seis corpus que recogen diversas manifestaciones del humor en español oral y escrito (columnas

periodísticas, narraciones humorísticas escritas por aprendientes de español en inmersión lingüística en Alicante, conversaciones coloquiales del corpus Val.Es.Co, monólogos humorísticos de mujeres, etc.) con el fin de analizar el humor en distintos contextos y desde diversas perspectivas discursivas.

En la misma línea, pero con el objetivo de analizar la evolución del humor en el español oral, se ha construido *Humcor* (Corpus oral multimodal de humor en español, https://humcor.com/), el único corpus informatizado dedicado específicamente al estudio del humor en el español oral de España e Hispanoamérica. En líneas generales, esta base de datos recopila textos humorísticos como chistes, monólogos, sketches y espectáculos de ventrílocuos en dos modalidades principales: visual y auditiva. Los registros audiovisuales provienen de diversas fuentes, como televisión, radio, casetes, discos de vinilo y discos de pizarra. Los materiales televisivos y radiofónicos se han obtenido de archivos históricos en línea, grabaciones personales y plataformas de contenido. Actualmente, el corpus incluye más de 150 programas recopilados de diferentes espacios de televisión como *No te rías que es peor*, *Genio y figura*, *Son risas y lágrimas*, *Sopa de ganso*, *Saque bola*, *Al ataque show*, entre muchos otros. En cuanto a los formatos físicos, estos han sido localizados en colecciones privadas, bibliotecas, sitios de compraventa de segunda mano y plataformas digitales. Además, el proceso de documentación ha implicado la consulta de catálogos de discos de distintas épocas, archivos bibliográficos, periódicos y revistas antiguas, así como sitios especializados en internet.

El corpus abarca más de 120 años de producción humorística, con materiales representativos de cada década desde 1900 hasta 2024, con más de 750 humoristas, distribuidos entre hombres y mujeres provenientes de dieciocho países diferentes. En la actualidad, la base de datos cuenta ya con más de 11 000 muestras, esto es, 8500 chistes, 663 monólogos y 1892 sketches, clasificadas según veinte criterios diferentes. Su catalogación se basa en tres grandes categorías: a) programa: tipo, medio de transmisión y año; b) hablante: sexo, generación, origen y, cuando es relevante, nombre y profesión; y c) texto humorístico: temática, presencia de gestos, entre otros aspectos. Cada archivo de vídeo o audio dentro del corpus está asignado a un código único que permite identificarlo según varios criterios: el tipo de género humorístico (mon=monólogo; chis=chiste; ske=*sketch*, ven=ventrílocuo), las características del hablante (sexo = M/H), la generación o edad (jóvenes = 1, adultos = 2, mayores = 3), la categoría profesional del hablante (no profesional = 0, profesional = 1), el código de la provincia y el número correspondiente dentro del corpus. Por ejemplo, el código 0171-CHI-BARH21 se interpreta como un chiste (el número 0171 dentro de *Humcor*) de un hablante varón, adulto, profesional, originario de la provincia de Barcelona. En la transcripción de los archivos se han seguido las normas generales de escritura del español, con la excepción del uso de mayúsculas, que se limitan exclusivamente a los nombres propios. Para la codificación de los materiales, se ha adoptado un sistema de marcas y etiquetas mínimas basado en el Standard Generalized Markup Language (SGML), de acuerdo con las

especificaciones de la Text Encoding Initiative (TEI). No se emplean signos de puntuación convencionales; en su lugar, se utilizan marcas específicas para indicar pausas de distinta duración.

El objetivo de *Humcor* es analizar y documentar el humor en español oral a través de discursos humorísticos tanto de profesionales como de aficionados. Esto permite un estudio más profundo de las características lingüísticas, culturales y sociales del humor en las diferentes variedades del español. Además, el corpus facilita la investigación sobre las diferencias en el uso del humor entre hombres y mujeres en distintas regiones de España e Hispanoamérica. Este archivo digital no solo servirá para preservar un patrimonio cultural de gran valor, sino que abre también nuevas vías para entender cómo el humor ha evolucionado y se ha ido adaptando a los tiempos y a los cambios sociales y culturales acaecidos a lo largo del último siglo. De cara al futuro, se planea, además, la creación de mapas interactivos que muestren de manera clara y detallada estas diferencias geográficas y temporales. Los materiales incluidos en el corpus *Humcor* no solo serán útiles para investigadores en lingüística, historia y antropología, sino también para cualquier persona interesada en el estudio del humor.

En esta ocasión nos proponemos presentar las características principales del corpus *Humcor*, los criterios teóricos y metodológicos que fundamentaron su diseño, así como los procedimientos empleados para la recopilación, digitalización, clasificación y codificación de los datos. Asimismo, se abordarán los criterios de búsqueda dentro de esta base de datos, que proporciona acceso a una extensa colección de textos humorísticos en español.

Bibliografía

ATTARDO, S. (1994), *Linguistic Theories of Humor*, Berlín, Mouton de Gruyter.

FREUD, S. (1905), *Der Witz und seine Beziehung zum Unbewußten*, Leipzig / Viena, Franz Deuticke.

JIMÉNEZ RUIZ, J. L. (2023), «Los campos del léxico escatológico en el humor negro infantil», *Círculo de Lingüística Aplicada a la Comunicación*, 96, pp. 37–58. https://doi.org/10.5209/clac.91590

KHAYLINA, A., & ALVARADO ORTEGA, M. B. (2023), «Propuesta de etiquetaje pragmático para un corpus conversacional en español», *Círculo de Lingüística Aplicada a la Comunicación*, 96, pp. 19–35. https://doi.org/10.5209/clac.91589

LINARES BERNABÉU, E. (2023), «El corpus FEMMES.UP: análisis pragmático de los estilos humorísticos en la comedia femenina», *Círculo de Lingüística Aplicada a la Comunicación*, 96, pp. 89–99. https://doi.org/10.5209/clac.91593

MEYER, J. C. (2000), «Humor as a double-edged sword: Four functions of humor in communication», *Communication Theory*, 10(3), pp. 310–331.

RASKIN, V. (1985), *Semantic Mechanisms of Humor*, Dordrecht, D. Reidel.

RUIZ GURILLO, L. (2016), «Metapragmatics of humor: Variability, negotiability and adaptability in humorous monologues», en L. RUIZ GURILLO (ed.), *Metapragmatics of Humor: Current Research Trends*, Ámsterdam, John Benjamins, pp. 81–104. https://doi.org/10.1075/ivitra.

RUIZ GURILLO, L. (ed.) (2022), *Interactividad en modo humorístico. Géneros orales, escritos y tecnológicos*, Madrid, Iberoamericana Vervuert. https://doi.org/10.31819/9783968693156

RUIZ GURILLO, L. (2023), «La pragmática de un etiquetaje pragmático para la plataforma observahumor.com», *Círculo de Lingüística Aplicada a la Comunicación*, 96, pp. 1–17. https://doi.org/10.5209/clac.91588

VARGAS-SIERRA, C. (2023), «Análisis emocional del corpus COLUMNAS.HUMOR: un enfoque mixto», *Círculo de Lingüística Aplicada a la Comunicación*, 96, pp. 73–87. https://doi.org/10.5209/clac.91592

La manifestación de cortesía valorizadora y mitigadora en el contexto del malentendido digital: un análisis pragmático de interacciones en Reddit

Yuezhe Du
Universitat de València

Palabras clave: Cortesía valorizadora; Cortesía mitigadora; Estrategias de cortesía; Malentendido; Reddit.
Keywords: Appreciative courtesy; Mitigating courtesy; Politeness strategies; Misunderstanding; Reddit.

El malentendido constituye un fenómeno poco investigado en la lingüística, que se produce cuando existe una discrepancia entre lo que el emisor comunica y lo que el destinatario interpreta, ya sea a nivel de significado o de intención (Humphreys, 1986; Weigand, 1999; Padilla Cruz, 2020). Este tipo de situaciones en la interacción puede dar lugar a reacciones negativas e incluso a conflictos (Padilla Cruz, 2020; Sanmartín Sáez y Albelda Marco, 2023), en los que emergen actos de habla hostiles y descorteses.

No obstante, a pesar de las posibles consecuencias negativas del malentendido y de la aparición de actos descorteses, es posible observar también manifestaciones de cortesía por parte de los interlocutores, orientadas a reparar el malentendido y a restablecer el propósito comunicativo inicial.

El presente estudio tiene como objetivo analizar las manifestaciones de cortesía valorizadora y mitigadora en un corpus de interacciones digitales. Para ello, se adopta el modelo de cortesía propuesto por Carrascsa Santana (1999) para la cultura hispanohablante, el cual distingue entre cortesía valorizadora y cortesía mitigadora. Esta propuesta se plantea como una alternativa a la aplicación directa del modelo de Brown y Levinson (1987) al contexto español, ya que incorpora particularidades culturales propias del ámbito hispánico. La cortesía valorizadora busca potenciar la imagen del interlocutor y generar un efecto positivo en la interacción (Barros García, 2018), mientras que la cortesía mitigadora se emplea en contextos donde existe un riesgo de amenaza a la imagen y procura prevenir o reparar dicho riesgo (Albelda Marco y Briz Gómez, 2010: 242).

Con el propósito de obtener una visión más amplia de las manifestaciones de cortesía en situaciones de malentendido en contextos digitales, se han recopilado 61 interacciones con malentendidos procedentes del subreddit r/es en Reddit —una plataforma conformada por diversos foros, pero aún escasamente abordada desde la perspectiva lingüística (Lu, 2024)—, que

incluyen un total de 429 intervenciones. Dada la dificultad que implica identificar malentendidos en interacciones naturales (Kavetska, 2025), la muestra fue recopilada mediante búsquedas específicas de expresiones lingüísticas vinculadas al fenómeno, como *me refiero*, *se refiere*, *pensaba que,* etc. (Romero-Trillo y Lenn, 2011), utilizando la herramienta Pullpush API.

El análisis contempla tanto la frecuencia como la tipología de las estrategias empleadas en la manifestación de cortesía valorizadora y mitigadora. En cuanto a la cortesía valorizadora, se retoman las estrategias propuestas por Vivas Márquez y Ridao Rodríguez (2015), quienes se centraron en el estudio de este tipo de cortesía en Twitter, y se adaptan al contexto particular de Reddit, donde la interacción entre usuarios conocidos es considerablemente menor, lo que incide en la configuración pragmática de los intercambios comunicativos. Estas estrategias se clasifican en directas e indirectas. Las directas incluyen fórmulas de saludo y despedida, expresiones de ánimo y buenos deseos, agradecimientos, consejos o recomendaciones corteses, elogios, manifestaciones de empatía y votos positivos. Las estrategias indirectas comprenden manifestaciones de interés, expresiones de acuerdo y bromas afiliativas.

Respecto a la cortesía mitigadora, esta se asocia de manera constante con la atenuación (Albelda Marco y Briz Gómez, 2010). Investigaciones previas como las de Robles Garrote (2014) y Sanmartín Sáez (2019) han abordado su presencia en redes sociales. A partir de estas contribuciones y considerando las particularidades de Reddit, se identifican las siguientes estrategias de cortesía mitigadora: disculpas, agradecimientos anticipados, explicaciones y justificaciones, minimización del desacuerdo, eliminación o edición del comentario publicado, y expresión de restricciones en la – opinión.

Además, se presta atención a los recursos específicos del entorno digital, tales como emoticonos, emojis y expresiones características de las redes sociales, los cuales adquieren una relevancia especial debido a las dinámicas comunicativas distintas de las interacciones cara a cara. Estos recursos pueden emplearse tanto para manifestar cortesía valorizadora y mitigadora como para expresar descortesía, y su función puede variar según el contexto discursivo.

Finalmente, se realiza una comparación con la frecuencia de descortesía en estos contextos, con el fin de ofrecer una visión más profunda de la (des)cortesía en situaciones de malentendido digital y contribuir así al estudio de un fenómeno aun escasamente atendido en la investigación lingüística.

Bibliografía

ALBELDA MARCO, M., & BRIZ GÓMEZ, A. (2010), «Aspectos pragmáticos: cortesía y atenuantes verbales en las dos orillas a través de muestras orales», en M. ALEZA IZQUIERDO & J. M. ENGUITA UTRILLA (eds.), *La lengua española en América: normas y usos actuales*, Valencia, Universidad de Valencia, pp. 237–260.

BARROS GARCÍA, M. J. (2018), *Cortesía valorizadora: uso en la conversación informal española*, Berlin, Peter Lang.

BROWN, P., & LEVINSON, S. C. (1987), *Politeness: Some universals in language usage*, Cambridge University Press.

CARRASCO SANTANA, A. (1999), «Revisión y evaluación del modelo de cortesía de Brown y Levinson», *Pragmalingüística*, 7, pp. 1–44.

HUMPHREYS-JONES, C. (1986), *An investigation of the types and structure of misunderstandings* [Tesis doctoral], Newcastle University.

KAVETSKA, A. (2025), «Scripted misunderstandings and humour: A case study of Monty Python's dialogues», *Psychology of Language and Communication*, 29(1), pp. 106–134.

LU, R. (2024), «Audience design and pragmatic conceptions of moves and upvotes during advice-giving on Reddit», *Journal of Pragmatics*, 219, pp. 30–47.

PADILLA CRUZ, M. (2020), «El malentendido», en J. AMENÓS PONS, M. V. ESCANDELL VIDAL, & A. K. AHERN (eds.), *Pragmática*, Madrid, Difusora Larousse / Alianza Editorial, pp. 384–407.

ROBLES GARROTE, P. (2014), «La cortesía verbal en la interacción asincrónica académica: análisis contrastivo en inglés, español e italiano», *Estudios de Lingüística Aplicada*, 60, pp. 117–139.

ROMERO-TRILLO, J., & LENN, E. (2011), «Do you "(mis)understand" what I mean? Pragmatic strategies to avoid cognitive maladjustment», *Journal of English Studies*, 9, pp. 223–241.

SANMARTÍN SÁEZ, J. (2019), «Análisis contrastivo de la (des)cortesía en los comentarios digitales del periódico 20minutos.es y del Facebook de viajacontumascota.com: anonimato y comunidad virtual», *ELUA*, 33, pp. 173–194.

SANMARTÍN SÁEZ, J., & ALBELDA MARCO, M. (2023), «La interpretación del discurso reproducido como fuente de conflicto en conversaciones familiares», *RILCE: Revista de Filología Hispánica*, 39(2), pp. 742–771.

VIVAS MÁRQUEZ, J., & RIDAO RODRIGO, S. (2015), «Estrategias de (des)cortesía en redes sociales: análisis comparativo de Facebook y Twitter», *Sintagma*, 27, pp. 73–87.

WEIGAND, E. (1999), «Misunderstanding: The standard case», *Journal of Pragmatics*, 31, pp. 763–785.

Discurso, coherencia y conflicto patrimonial: análisis computacional de estructuras retóricas en torno a los ataques patrimoniales durante el *Black Lives Matter*

Patricia Martín-Rodilla
Instituto de Estudios Gallego Padre Sarmiento – Consejo superior de Investigaciones Científicas
Martial Pastor
Centre for Language Studies – Radboud University Nijmegen

Palabras clave: Black Lives Matter; Rhetorical Structure Theory; discurso; archivo social
Keywords: Black Lives Matter; Rhetorical Structure Theory; discurso; archivo social

Introducción

En la última década, las redes sociales se han convertido en una fuente valiosa para la investigación en humanidades y ciencias sociales, al reflejar temas como la memoria, la identidad y el patrimonio a través de las interacciones de los usuarios. A partir del concepto de archivo social (Ruiz-Gómez y Vallés, 2020), estas plataformas ofrecen registros colectivos útiles para estudiar aspectos como el recuerdo-olvido, la salud mental, el género o el racismo, todos vinculados con los estudios de memoria e identidad.

Movimientos sociales como *Black Lives Matter* (BLM) son especialmente relevantes. Si bien ha sido profundamente estudiado desde perspectivas legislativas (Chase, 2017) o mediáticas (Giorgi et al. 2022), aún persisten dudas sobre los ataques a entidades patrimoniales ocurridos en los momentos álgidos del movimiento y su racionalidad asociada (Edmonds, 2021; Budabin, 2023). ¿Por qué se atacaron ciertas entidades y otras no? ¿En qué contextos? ¿Con qué justificación?

Para responder a estas preguntas debemos abordar qué discursos subyacen dentro de la red social a este respecto. Este trabajo analiza seis meses de actividad en inglés sobre el movimiento BLM y los ataques patrimoniales asociados, utilizando la Teoría de la Estructura Retórica (RST por sus siglas en inglés) (Mann and Thompson, 1988) y técnicas de análisis automático del discurso (Taboada and Mann, 2006). A través de un enfoque macro-discursivo,

se examinan estructuras y estrategias retóricas empleadas para justificar los ataques a símbolos patrimoniales.

Teoría de la estructura retórica y análisis computacional del discurso

Aunque existen trabajos previos en formalización del discurso desde hace décadas, se considera a RST la base de los estudios computacionales de discurso (Taboada y Mann, 2006). Esta teoría representa el discurso como un árbol de unidades relacionadas retóricamente (núcleo o satélite), según funciones semánticas como causalidad o contraste. Su implementación computacional comenzó en la década de los 2000, con técnicas que segmentan el discurso y construyen árboles mediante reglas o aprendizaje automático. Un resumen de técnicas computacionales sobre *parsing* de discurso para humanidades fue realizado en 2023 (Martin-Rodilla, 2023). Actualmente, aunque los modelos de lenguaje han mejorado, siguen siendo limitados. Este estudio emplea uno de los analizadores más precisos (Braud et al., 2017), con un rendimiento del 85% en estructura no etiquetada y 72% con nuclearidad.

Fuentes de datos y metodología de análisis computacional

Aunque muchas investigaciones sobre el movimiento BLM utilizan datos de redes sociales, plataformas como Twitter (ahora “X”) tienen límites en la extensión de los mensajes y restricciones de forma. Para este estudio, era importante contar con una fuente que permitiera textos más largos y complejos, donde los usuarios pudieran usar libremente distintas estructuras retóricas, y que además incluyera información específica sobre ataques a patrimonio. Por eso se eligió una colección de Reddit (Otero et al. 2021), plataforma que facilita conversaciones extensas y detalladas, con 296 hilos anotados y un total de 260,578 publicaciones entre mayo y octubre de 2020.

Así, se llevó a cabo el procesado automático de toda la colección, organizándola en sus 296 hilos. Esto implica que para cada hilo contamos como resultado con un fichero asociado al hilo en formato *json* que describe la estructura RST arbórea discursiva de cada hilo. En la Figura 1 puede verse un ejemplo del fichero RST para un post de Reddit y la información de discurso resultante. Como paso final del análisis computacional, se han realizado análisis estadísticos sobre el conjunto de la información arborescente obtenida. Estas estadísticas permiten conocer la distribución de relaciones RST y observar estructuras retóricas asociadas a entidades patrimoniales.

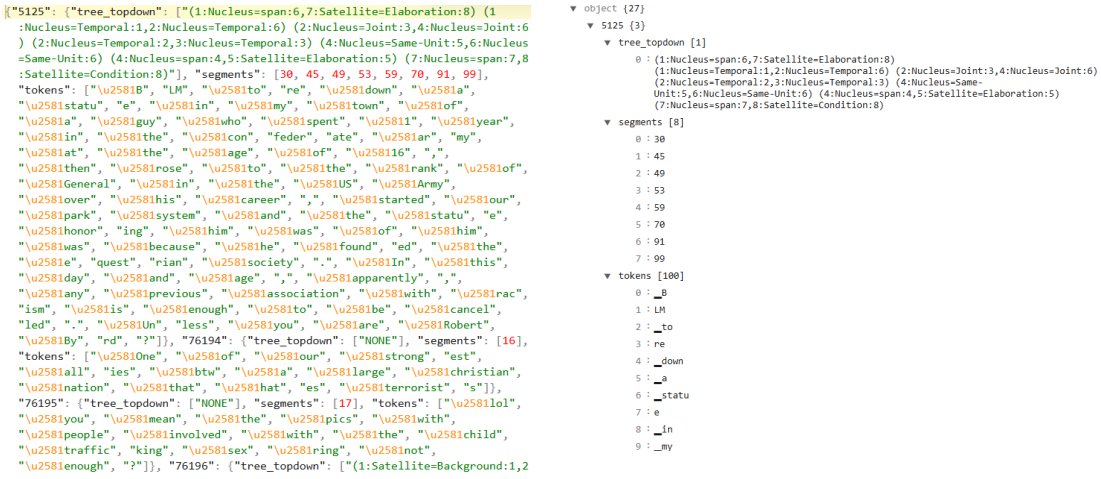


Figura 1. Fichero resultante del análisis RST para un post dado (izquierda) y estructura arbórea extraída (derecha).

Resultados

Los resultados se organizan en dos bloques, según las preguntas de investigación. El primero analiza cómo se construye el discurso sobre el BLM y los ataques patrimoniales a partir de las estructuras RST. El segundo explora por qué y para qué se formula ese discurso, identificando estructuras retóricas asociadas a entidades patrimoniales que permitirán futuras interpretaciones sobre la racionalidad de los ataques.

Estructuras discursivas: Patrones y distribución

Se analizaron las frecuencias de las 18 relaciones de coherencia determinadas en RST (Braud et al. 2017), tanto de forma individual como en combinaciones binarias.

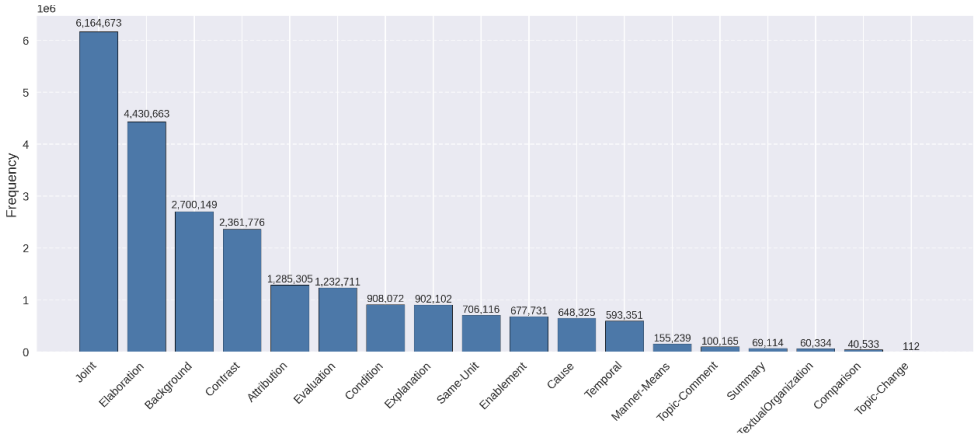


Figura 2. Distribución de relaciones RST identificadas. Corpus general.

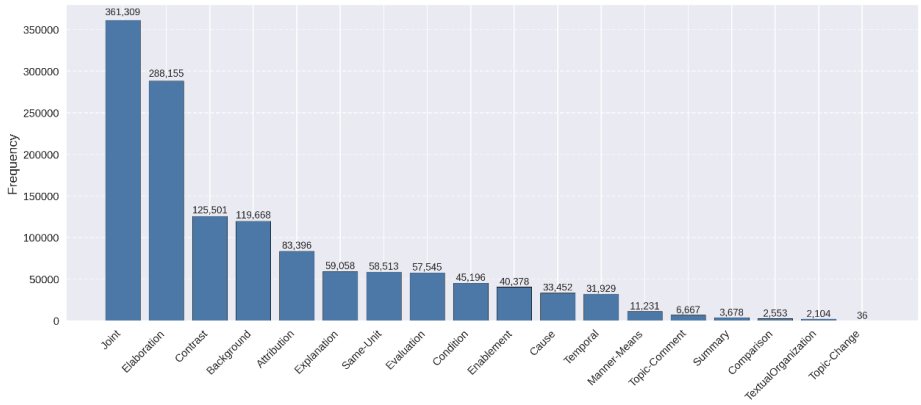


Figura 3. Distribución de relaciones RST identificadas. Subcorpus de entidades patrimoniales.

Las Figuras 2 y 3 presentan las distribuciones de relaciones RST unitarias en el corpus completo y en el subcorpus centrado en entidades patrimoniales, respectivamente. La relación Joint, típica de discursos fragmentados pero continuos, es la más frecuente, lo que refleja la naturaleza encadenada del discurso en Reddit. También hay una alta presencia de relaciones de elaboración y contexto. El subcorpus de entidades patrimoniales muestra más relaciones de autorreferencia como Same-Unit y Explanation, lo que sugiere un esfuerzo mayor por justificar los ataques.

En las relaciones binarias, destaca Joint-Joint por su frecuencia. También son comunes combinaciones como Elaboration-Background, que organizan el discurso en torno a un tema y su explicación.

El subcorpus de entidades muestra más relaciones de contraste, lo que podría indicar una tendencia a incluir comparaciones y justificaciones explícitas. En resumen, mientras el corpus general presenta estructuras más coordinadas, el discurso sobre ataques patrimoniales muestra una estructura más compleja, contrastiva y autorreferencial, lo que puede reflejar la carga argumentativa y la sensibilidad del tema tratado.

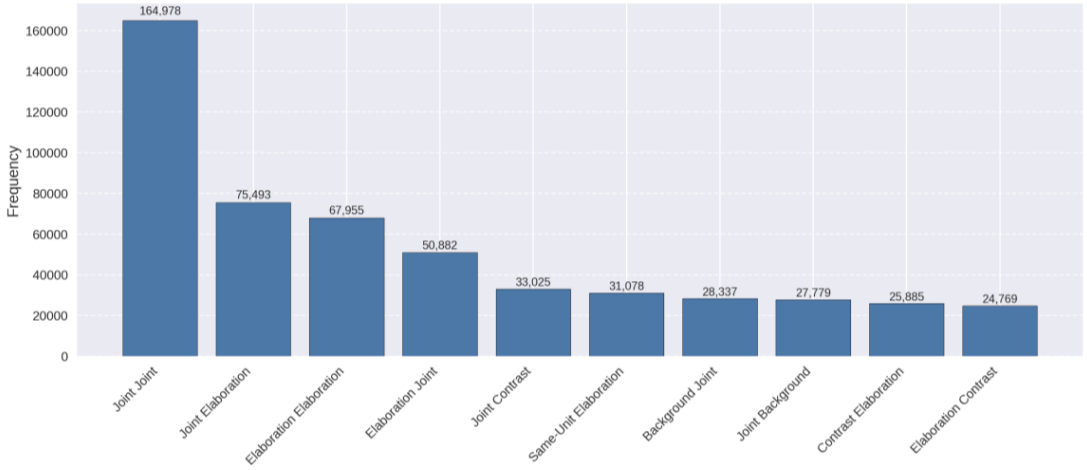


Figura 5. Las 10 relaciones RST binarias más frecuentes. Corpus general.

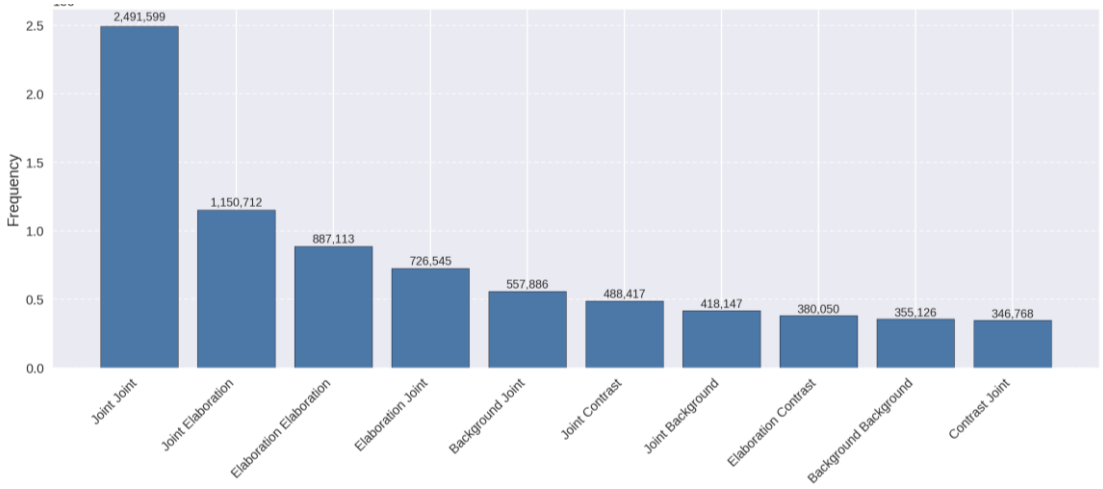


Figura 6. Las 10 relaciones RST binarias más frecuentes. Subcorpus de entidades patrimoniales.

Las **Figuras 5 y 6** muestran las distribuciones de relaciones RST binarias. En ambos casos, el patrón de frecuencia refleja lo observado en las relaciones unitarias. La combinación *Joint-Joint* es la más dominante, reflejando una construcción discursiva fragmentada y secuencial, típica del formato en redes sociales. También destacan las combinaciones de elaboración y contextualización (por ejemplo, *Elaboration-Background*), lo que sugiere una organización del discurso en torno a un tema seguido de su explicación o justificación.

Estructuras retóricas en BLM

Como aproximación inicial a un posterior análisis cualitativo de las posibles justificaciones y razonamientos asociados a los ataques patrimoniales, realizamos un análisis estadístico de las relaciones RST que aparecen más involucradas con entidades puramente patrimoniales en el corpus, como *estatue, memorial,* etc.

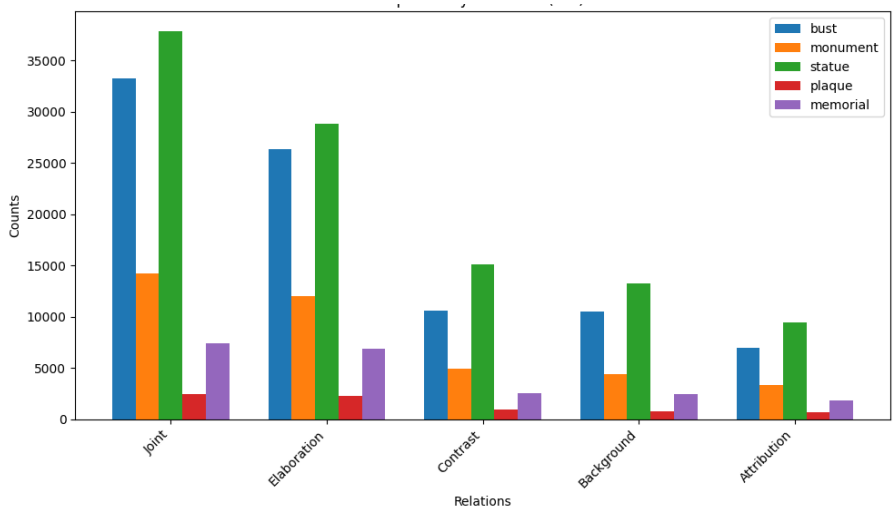


Figura 7. Distribución en las 5 relaciones RST más frecuentes de las entidades patrimoniales más relevantes. Corpus general.

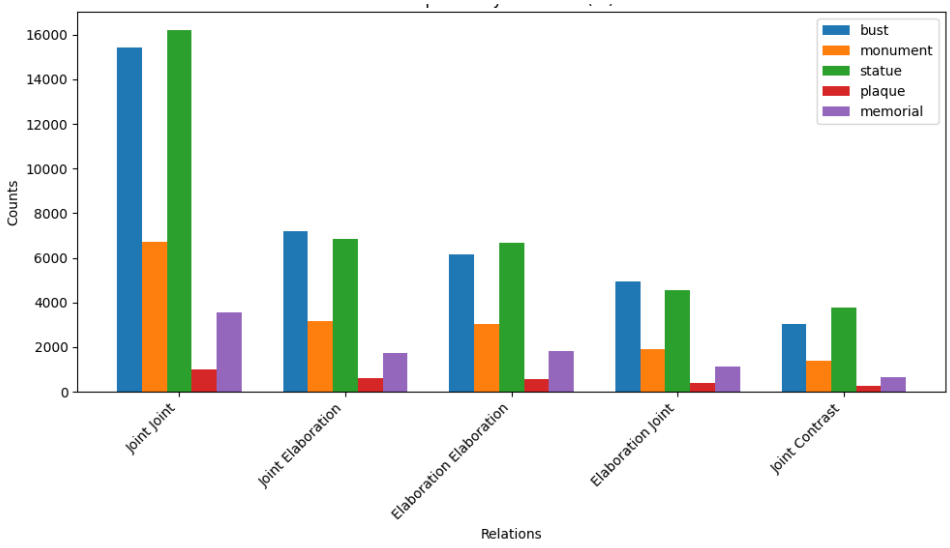


Figura 8. Distribución en las 5 relaciones RST más frecuentes de las entidades patrimoniales más relevantes. Subcorpus de entidades patrimoniales.

Las Figuras 7 y 8 muestran que las cinco entidades patrimoniales más mencionadas en el corpus aparecen principalmente en relaciones discursivas no jerárquicas y explicativas, como elaboraciones, contrastes y contextos. Esto implica que, al menos en este primer análisis, su función principal no es justificar otros argumentos ni construir cadenas causales. Podemos decir que las entidades patrimoniales analizadas en este corpus, o bien elaboran y amplían ideas centrales del discurso, ejemplifican otros conceptos más abstractos (como racismo) y/o contextualizan eventos o argumentos dentro de un marco histórico o espacial (como ataques). Si bien es necesario un análisis mucho más profundo, esta primera aproximación nos permite comprender los roles de las entidades patrimoniales y cómo abordar posibles justificaciones a los ataques.

Conclusiones y líneas futuras de trabajo

Este estudio presenta un análisis computacional macro-discursivo de un archivo social sobre el movimiento BLM, utilizando técnicas automáticas para extraer y analizar árboles RST desde redes sociales. Es un ejemplo, por tanto, de cómo archivos y/o corpus relativamente grandes en volumen pueden ser analizados sin perder de vista la complejidad retórica y semiótica del fenómeno a estudiar desde el punto de vista de discurso, permitiendo caracterizar estructuras discursivas clave y responder a preguntas de investigación específicas.

Quedan abiertas muchas preguntas. En primer lugar, es necesario seguir explorando estrategias retóricas como la persuasión o manipulación de opiniones en la colección. En segundo lugar, se está llevando a cabo en paralelo un análisis del discurso manual sobre un subconjunto de hilos de la colección con una mayor información sobre los ataques, que enriquecerá y evaluará tanto las hipótesis como los primeros hallazgos detallados en esta aproximación automática. Finalmente, se prevé integrar otras colecciones existentes sobre el fenómeno (en inglés u otros idiomas) que, si bien no están centrados en los ataques patrimoniales, podrían contener información valiosa para la respuesta a nuestras preguntas sobre los ataques.

Agradecimientos

Este trabajo cuenta con financiación del programa HE Marie Skłodowska-Curie, nº 101073351 y CSIC (Proyecto Intramural Especial 2024ICT028).

Bibliografía

BRAUD, C., COAVOUX, M., & SOEGAARD, A. (2017), «Cross-lingual RST discourse parsing», *arXiv preprint*, arXiv:1701.02946.

BUDABIN, A. (2023), «Heritage politics in the case of Black Lives Matter in Bolzano-Bozen, Italy», *European Journal of Cultural and Political Sociology*, 10(4), pp. 577–602.

CHASE, G. (2017), «The early history of the Black Lives Matter movement, and the implications thereof», *Nevada Law Journal*, 18.

EDMONDS, P. (2021), «Monuments on trial: #BlackLivesMatter, “travelling memory” and the transcultural afterlives of empire», *History Australia*, 18(4), pp. 801–822.

GIORGI, S., et al. (2022), «Twitter corpus of the #BlackLivesMatter movement and counter protests: 2013 to 2021», en *Proceedings of the International AAAI Conference on Web and Social Media*, 16, pp. 1228–1235.

MANN, W. C., & THOMPSON, S. A. (1988), «Rhetorical structure theory: Toward a functional theory of text organization», *Text*, 8(3), pp. 243–281.

MARTÍN-RODILLA, P. (2023), «Computer processing of language: Where archaeological discourse and computers meet», en *Discourse and*

Argumentation in Archaeology: Conceptual and Computational Approaches, Cham, Springer International Publishing, pp. 193–213.

OTERO, D., et al. (2021), «Building cultural heritage reference collections from social media through pooling strategies: The case of 2020's tensions over race and heritage», *Journal of Computing and Cultural Heritage*, 15(1), art. 9.

RUIZ GÓMEZ, V., & VALLÈS, A. M. (2020), «#Cuéntalo: The path between archival activism and the social archive(s)», *Archives and Manuscripts*, 48(3), pp. 271–290.

TABOADA, M., & MANN, W. C. (2006), «Rhetorical structure theory: Looking back and moving ahead», *Discourse Studies*, 8(3), pp. 423–459.

Leer a Aristóteles con IA: un modelo ampliado de lectura profunda de textos filosóficos

José Higuera Rubio
Salvador Ros
UNED

Palabras clave: Lectura Activa; Prompting; Chain-of-Thought; Marcación de textos; Desambiguación
Keywords: Active Reading; Prompting; Chain of Thought; Text Markup; Disambiguation

La lectura de textos filosóficos es una actividad que plantea múltiples cuestiones relativas al uso de herramientas de Procesamiento de Lenguaje Natural (NLP) y los Modelos generativos de lenguaje (LLMs). Por un lado, la compleja trama argumentativa de los textos filosóficos oscila entre enunciados generales, sinopsis del pensamiento de otros autores, ejemplos y conclusiones. Estos elementos en conjunto implican una “cadena” argumentativa, lo que caracteriza al discurso filosófico (Davis, 2003). Por otra parte, la prolífica aparición de herramientas-IA orientadas a la lectura de textos, incluyendo los filosóficos, promete una producción ilimitada de sinopsis, la proliferación de enunciados contextuales y largos listados bibliográficos. Sin embargo, esto no significa leer de manera profunda y activa un texto filosófico. La pregunta es ¿cuál es la fiabilidad de estas herramientas (LLMs) frente a la lectura activa que se requiere en la enseñanza universitaria de textos filosóficos? La lectura de textos es la actividad - individual y colectiva- responsable del diálogo entre los filósofos a lo largo de la historia. La filosofía es ante todo la historia de una gran diversidad de lecturas, pero no de cualquier lectura (Moretti, 2013; Guillori, 2005). Nuestro reto es proponer un modelo de lectura cooperativa de textos filosóficos entre los usuarios de IA (LLMs), utilizando herramientas de procesamiento de lenguaje natural NLP con la finalidad de refinar y ajustar los hábitos de lectura tanto de herramientas-IA como las habilidades lectoras activas de largo alcance que puede desarrollar un grupo de lectores liderado por un docente. La metodología aplicad es un proceso colaborativo que conlleva los siguientes pasos: i.) la marcación o desambiguación de un corpus textual con el objetivo de detectar los aspectos conceptuales y semánticos más relevantes de un texto; ii.) la extracción e interpretación de las partes significativas que sirven de estructura discursiva a la argumentación filosófica presente en los textos; iii.) introducir un Clúster- de-Operaciones (Cuadro B) en un Modelo Generativo de Lenguaje para que se reconozcan y extraigan los marcadores semántico-conceptuales del

texto; iii.) la creación de un ambiente de interacción significativa para que una comunidad utilice el modelo y ofrezca un conjunto de resultados a partir de su experiencia (Kanesci, 2023).

La lectura activa de textos filosóficos es una experiencia que engloba la detección y extracción de marcadores conceptuales que determinan la estructura argumentativa de un texto más allá de su contexto histórico. En una lectura activa es posible detectar y extraer marcadores conceptuales que luego serán detectados y extraídos por el modelo de lectura activa. Los marcadores son las Premisas (Px), las Definiciones (Dx), los Ejemplos (Ex), las Objeciones a otros autores (Ox), los Argumentos (Ax) y las Conclusiones (Cx). En un primer ejercicio, recurrimos a Modelos Generativos de Lenguaje (LLMs), disponibles online, utilizando la técnica de elaboración de operaciones denominada Chain-of-Thought (Img. 1). Se trata de elaborar una serie de preguntas que propongan al modelo una vía para detectar las partes de un razonamiento complejo a partir de un conjunto de marcadores conceptuales. Al comparar los resultados de una misma cadena de preguntas producidas por un GPT desarrollado con Open AI frente a Copilot se observan grandes diferencias. Copilot es capaz de detectar los tipos conceptuales de Premisas (Px) y definiciones (Dx), identificando cada una de ellas con áreas problemáticas específicas del pensamiento filosófico (ver cuadro A). Este modelo extrae las premisas y las definiciones clasificándolas de acuerdo con su afinidad con una fuente textual de la que se ha servido el autor o con áreas específicas como la lógica, la metafísica o la epistemología. Sin embargo, el primer modelo tan solo enumera los marcadores conceptuales sin clasificar el área filosófica o la tradición textual que utiliza. Los niveles de extracción y clasificación de los marcadores conceptuales en los textos cambian según los Modelos (LLMs). El problema al que nos enfrentamos es la construcción de un modelo fiable de detección y extracción de marcadores conceptuales en los textos filosóficos que generen estructuras argumentativas del tipo Premisa-Definición-Ejemplo-Argumento-Conclusión con la capacidad de extraer una estructura argumentativa conceptual. Esta investigación reformula la inserción de preguntas para generar “cadenas de pensamiento” (Chain-of-Thought, ver cuadro B) en los Modelos (LLMs) para que estos tengan la capacidad operativa de detectar y extraer los marcadores conceptuales de un texto en un mismo Clúster-de-Operaciones para que sus respuestas generen de manera automática cadenas argumentativas que estimulen la lectura activa y profunda de los textos filosóficos en los usuarios. Este Clúster se denomina Cadena-de-pensamiento-automática (Automatic Chain of Thought Prompting) y está basada en un conjunto de operaciones interrelacionadas y dependientes entre sí que pueden generar estructuras argumentativas a partir de la detección de marcadores conceptuales en un texto (Zhang et al. 2022).

De este modo, se obtienen modelos que detectan los marcadores conceptuales y la estructura argumentativa siguiendo los hábitos de una lectura profunda-activa de textos: relectura, señalización de marcadores semántico-conceptuales y generación de estructuras argumentativas (resultados en cuadro

A). Es posible entrenar una herramienta LLMs para que imite los hábitos de lectura profunda. Además, este trabajo aborda el desafío de implementar un modelo cooperativo para desarrollar hábitos de lectura profunda en humanos o "lectores-LLMs". Un modelo metodológico que podría llamarse "Textualidad Extendida" o "Lectura Ampliada" (Schüller-Zwierlein, A., et al., 2023).

Bibliografía

ARCHER, J., & JOCKERS, M. (2016), *The Bestseller Code*, Penguin.

GUILLORY, J. (2005), *On Close Reading*, Chicago UP.

DAVIS, P. (2003), *Experience of Reading*, Routledge.

KASNECI, E., et al. (2023), «ChatGPT for good? On opportunities and challenges of large language models for education», *Learning and Individual Differences*, 103, art. 102274. https://doi.org/10.1016/j.lindif.2023.102274

MORETTI, F. (2013), *Distant Reading*, Verso.

SCHÜLLER-ZWIERLEIN, A., MANGEN, A., KOVAC, M., & VAN DER WEEL, A. (2022), «Why higher-level reading is important», *First Monday*, 27(5). https://doi.org/10.5210/fm.v27i5.12770

ZHANG, Z., ZHANG, A., LI, M., & SMOLA, A. (2022), «Automatic chain of thought prompting in large language models», *arXiv preprint*, arXiv:2210.03493.

Enlaces

Chain-of-Thought Prompting
https://www.promptingguide.ai/techniques/cot

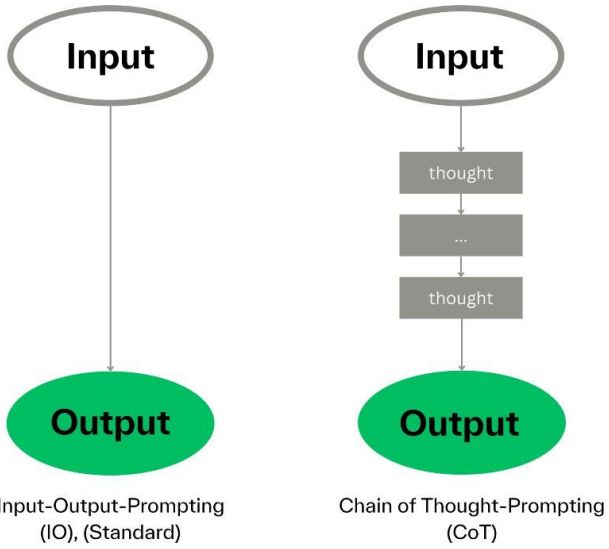
Chain of Thought Prompting: Guiding LLMs Step-by-Step
https://medium.com/@pankaj_pandey/chain-of-thought-prompting-guiding-llms-step-by-step-e6eac32d02d8

Cuadro A

Herramienta GPT	Copilot- Think Deeper
En el siguiente texto detecta premisas, definiciones, argumentos, ejemplos, autores y conclusiones	En el siguiente texto detecta premisas, definiciones, argumentos, ejemplos, autores y conclusiones
1. **Argumentos (A):** A1. A2. A3. A4. A5. […] 2. **Premisas (P):** P1. P2. P3. P4. […] 3. **Conclusión (C):** C1. 4. **Objeciones (OB):** OB1. OB2. 5. **Soluciones / Respuestas (S o R):** R1. R2. 6. **Definiciones (D):** D1. D2. D3. 7. **Ejemplos (EJ):** EJ1. EJ2. 8. **Preguntas filosóficas (P):** P1. P2. **G. Pasajes confusos (?):** ?1. ?2.	**1. Premisas** Las premisas son afirmaciones básicas o supuestos de partida sobre los cuales se edifica la exposición filosófica. En este texto destacan, por ejemplo: Premisa inicial: (párrafo 1) Premisa ontológica y epistemológica: (párrafo 1) Premisa de conocimiento: (párrafos 3, 4 y 5) Premisa de la división del mundo y los saberes: (párrafo 6 y 15) **2. Definiciones** El autor ofrece definiciones precisas que ayudan a estructurar la unidad de la "verdad" en distintos aspectos: Definición de "la verdad" Definición de la inteligencia (párrafo 4) Definición desde el objeto: (párrafo 5) Clasificación del mundo Definiciones en el ámbito matemático y de las ciencias (párrafo 15) Definición en el arte de la palabra (párrafo 18) **3. Argumentos** Argumento a favor de una verdad intrínseca (párrafo 2) Argumento de la triple causa: (párrafo 3) Argumento sobre la interacción del sujeto y el objeto: (párrafo 5) Argumento respecto a la claridad de los datos sensibles: (párrafo 14) Argumento sobre el poder comunicativo de la palabra: (párrafo 18) **4. Ejemplos** El texto recurre a ejemplos para ilustrar y concretar conceptos abstractos:

	Ejemplos provenientes de clásicos: (párrafo 2) Ejemplos de explicación natural: (párrafo 6) Ejemplos en el campo de las ciencias: (párrafo 15) **5. Autores** El texto se apoya en diversas autoridades o fuentes para fundamentar sus afirmaciones: Mención general a "los filósofos": (párrafos 7 y 17) **6. Conclusiones** Las conclusiones en el texto se van introduciendo a medida que se conectan los distintos niveles de la exposición: Conclusión sobre el alma: (párrafo 3) Conclusión epistemológica: (párrafo 5) Conclusión acerca de las ciencias sensibles: (párrafo 14) **Reflexión y conexiones** Si te interesa profundizar, podríamos explorar cómo estos marcadores han influido en la tradición filosófica o cómo se articulan en otros sistemas teóricos y cuál es la repercusión en el pensamiento contemporáneo. ¿Te gustaría indagar más en alguna de estas conexiones?

Cuadro B



Nigromancia Digital: Hacia una epistemología pos-hermenéutica para las Humanidades Digitales[1]

Ángel M.ª Lumbreras
iArtHis_Lab –
Universidad de Málaga

Palabras clave: Nigromancia digital; Humanidades Digitales; post-hermenéutica; inteligencia artificial; epistemología posthumana
Keywords: Digital necromancy; Digital Humanities; post-hermeneutics; artificial intelligence; posthuman epistemology

Introducción. Los espectros del saber

En el umbral de una fase ulterior de la era algorítmica, donde las Humanidades Digitales institucionalizan sus prácticas y la mayoría de flujos de trabajo pasan por diferentes modelos de IA, reaparece una figura inquietante: el «espectro pensante». Esta propuesta parte de la hipótesis de que los modelos generativos no deben entenderse como instrumentos de representaciones, sino como dispositivos capaces de hacer emerger conocimiento latente del archivo, reactivándolo sin mediación interpretativa.

La tradición hermenéutica concibió la compresión como un proceso histórico y circular. Sin embargo, al distribuirse la conciencia entre agentes humanos y maquínicos, descentrando el eje autoral, el archivo comienza a hablar sin ser leído. La «nigromancia digital»[2] se define aquí como reactivación contingente y situada de voces históricas, determinada por la curaduría de un corpus y los hiperparámetros del modelo, que delimitan el cuándo, cómo y bajo qué contexto emerge el espectro. Esta epistemología espectral reemplaza la interpretación por la ejecución técnica y sitúa a la máquina como mediadora del saber. Según Joanna Zylinska (2024), las máquinas de percepción no se limitan a observar el mundo, sino que también participan activamente en su composición. Lo que antes era una herramienta comienza a operar como interfaz

[1] Este proyecto forma parte de la tesis doctoral *Re(tis) Cogitans. Aproximación performativa a las identidades sintéticas digitales como herramienta para las disciplinas humanísticas,* financiado por el Ministerio de Ciencia, Innovación y Universidades a través de las ayudas FPU 2023.
[2] El término nigromancia digital no cuenta con una definición formal estandarizada en la literatura académica, aunque es popular en entornos tecnológicos, divulgativos y jurídicos. Si se recurre a la Real Academia de la Lengua Española, esta expresión puede inferirse como la posibilidad de «invocar» la imagen, la figura o las cualidades de un difunto como agencia informacional en el presente. Esta interpretación encaja con ciertas aplicaciones de la inteligencia artificial que, a partir de archivos previos, generan respuestas nuevas atribuidas a una figura ya desaparecida.

entre memoria y su ejecución, es decir, como médium.

Este anhelo de contactar con los que ya no están no es nuevo. En *Apología,* Sócrates, según Platón, expresa su deseo de hablar con figuras míticas, asociando la muerte a un saber inaccesible. Incluso en la modernidad, intelectuales decimonónicos experimentaron con el espiritismo como tentativa científica sobre lo oculto. Hoy, ese impulso reaparece en forma de aplicaciones pseudoterapéuticas y pedagógicas que exploran simulaciones conversacionales con difuntos[3]. Sin embargo, lo que aquí se pretende es una exploración de los límites de la invocación algorítmica. Esta propuesta constituye una demostración especulativa de la viabilidad de concebir la nigromancia digital como epistemología poshermenéutica. Para ello, se propone un modelo especulativo de activación espectral, construido a partir de un corpus curado de textos de Giorgio Vasari (1511–1574), que, junto con una revisión crítica de la bibliografía, configura esta coyuntura técnica como un modelo de conocimiento no interpretativo con alcance heurístico en las Humanidades Digitales.

El médium y la consecuente disolución hermenéutica

La hermenéutica clásica entiende el archivo como depósito de sentido, una reserva de signos que debe ser interpretada con la mediación de un lector. En el régimen técnico contemporáneo, esta lógica se ha trastocado ya que todo archivo (textual, visual o sonoro) está sometido a procesos de transmutación digital cada vez más dependientes de modelos de inteligencia artificial. Es por ello que el archivo funciona ahora como sistema operativo y no como reserva de signos, desencadenando procesos sin intervención humana.

Este giro ha sido descrito por Parikka (2025), quien define el archivo como una «infraestructura de visión y cálculo», una red técnica que produce el mundo que aparenta documentar. Su función consiste en generar condiciones técnicas de aparición, más que en preservar o representar. Flusser (2025) anticipó esta mutación al proponer la «imagen técnica» como una forma automatizada que no representa lo real, sino que opera sobre ello. Mitchell (2019), centrado en la imagen y crítico con la hermenéutica, concibió el archivo visual como «imagen viviente», capaz de operar sin necesidad de ser nombrada o transcrita. Esta lógica se materializa en el ecosistema digital, como señala también Zylinska (2024), quien muestra cómo todo archivo digital actúa como unidad funcional autónoma. Integrados en tramas algorítmicas, estos archivos ya no son objetos, sino actantes de percepción.

Considerando lo anterior, la función ecfrástica —clave en la mediación entre lo visual y lo verbal— se desplaza hacia una automatización sistémica y

[3] Entre estas propuestas cabe nombrar a HereAfter AI, Replika o You Only Virtual, orientadas a recrear interacciones con familiares fallecidos mediante inteligencia artificial conversacional. En el ámbito educativo, destaca el desarrollo de AI-BotYou, una inteligencia artificial diseñada por el investigador Espiritusanto junto a la empresa española Bitbrain, pensada para simular personalidades pedagógicas ajustables a la manera de *coach* educativos, destinados a motivar al alumnado a dialogar sobre contenidos curriculares y transversales.

deja de funcionar como traducción subjetiva. El tránsito entre imagen y texto se deshumaniza. El archivo ya no revela un sentido oculto, sino que activa un estado técnico-mediado entre máquina y usuario. En esta interacción, archivo y médium forman un ensamblaje capaz de reconfigurar lo visible, lo decible y lo pensable. La figura del autor se diluye entre capas de codificación, parametrización y autoejecución.

De la autoría a la modulación algorítmica

Aunque el archivo se despliega como espacio latente de posibilidades espectrales, su activación requiere todavía lenguaje natural. El texto —en forma de prompt, instrucción o vector— no interpreta, pero permite su ejecución operativa.

El acto generativo producido por la IA no implica solo un cambio de estilo, sino un desplazamiento epistemológico. Goldsmith (2019) ya lo advirtió: en el entorno digital, la creación es sustituida por la reorganización, y «la escritura ya no trata de expresar ideas, sino de mover información» (p. 33). Parente (2024, 26) añade que toda acción discursiva está condicionada por una red técnica que nos engloba, sin posibilidad de enunciar desde fuera. El modelo generativo resultante ya no pronuncia una voz histórica con biografía, sino que produce un ensamblaje técnico donde el autor desaparece en el cálculo probabilístico. Esta voz artificial, despersonalizada, responde a la lógica de la interoperabilidad más que a la de la identidad.Su eficiencia —ya sea en modelos de lenguaje o de visión— depende de la ingesta masiva de archivos descontextualizados, en los que la información reaparece como respuesta sin memoria ni marco biográfico.

Un punto de inflexión lo marca ControlLM (Weng et al., 2024), sistema que permite modular rasgos de personalidad durante la generación sin necesidad de reentrenar el modelo. Esta capacidad supone una ruptura técnica, ya que permite ajustar la configuración en tiempo real. El modelo no reconstruye una identidad, sino que ensambla atributos sintéticos a partir de trayectorias latentes, de un modo más próximo al funcionamiento de un cerebro humano. No regresa una entidad concreta, sino una resonancia estilizada. Ferrández (2025) subraya que esta voz artificial no conserva la subjetividad del creador del archivo, aunque «simula una interacción emocionalmente verosímil» (p. 47), ajustada al presente.

Llegados a este punto, cabe preguntarse hasta qué punto esta contingencia técnica, propia de la invocación nigromántica, difiere de la contingencia interpretativa que aún sostiene las prácticas hermenéuticas.

Activando a Giorgio Vasari

Ante la limitación de tiempo y recursos, se desarrolló un experimento conceptual y funcional que simula parcialmente las capacidades de ControlLM, mediante una arquitectura ligera con scripts en Python y modelos generativos abiertos. El objetivo fue diseñar una invocación algorítmica estilísticamente ajustada que reactive una voz sintética situada [ver figura 1]

El flujo comienza con una pregunta del usuario dirigida a Giorgio Vasari, convertida en vector mediante embeddings generados con modelos tipo sentence-transformer. Esta representación se compara en un índice FAISS con los embeddings del corpus, recuperando los fragmentos más similares. Estos funcionan como memoria contextual y base estadística de estilo (adjetivos, conectores, longitud, densidad léxica), lo que permite no sólo ajustar el contenido temático, sino también la forma retórica de la respuesta. A partir de esta localización vectorial, se genera un prompt que instruye al modelo generativo Gemini Flash para producir una respuesta coherente con la identidad estilizada de Vasari. El resultado preserva elementos retóricos originales (incluidos italianismos identitarios), lo que permite evaluar su verosimilitud sin reentrenamiento.

Pese a su sencillez actual, el sistema demuestra que es posible inducir una modulación espectral en tiempo real a partir de atributos del archivo. Este prototipo constituye la primera fase de una arquitectura mayor, que en etapas futuras podría integrar sistemas como ControlLM, consolidando la nigromancia digital como epistemología heurística para las Humanidades Digitales.

Resultados y proyección crítica

Este experimento no aspiraba a la validez científica de un sistema técnico definitivo; más bien, ensaya una hipótesis filosófica mediante un flujo operativo verificable, más los resultados son consistentes, siendo la similitud de las respuestas cercana al 0,9 [ver figuras 2, 3 y 4]. La modulación espectral inducida puede constituir así una vía técnica de conocimiento no interpretativo, aunque con matices.

El sistema no opera con verdadera autonomía, ya que depende directamente de extractos específicos concretos del corpus. Además, el modelo generativo introduce conceptos exógenos, lo que subraya la necesidad de una ontología que module el campo semántico y garantice coherencia en la invocación.

También existen fricciones éticas y jurídicas en la invocación algorítmica de voces pasadas. El marco regulador europeo contempla derechos sobre imagen, autoría y reproducción post mortem, como advierte Paula Vega (2024), alertando que esta puede utilizarse como instrumento de control o de deposición si no va acompañada de una crítica estructural. Aquí, la nigromancia digital se sitúa como contra-modelo que exige pensar sus límites a través de la concreción de

condiciones de reactivación como acto técnico y ontológico.

Frente a la abstracción despersonalizada de los modelos masivos, esta propuesta ensaya una invocación ética y contingente, otorgando singularidad configurada. Y es precisamente esta singularidad, centrada en el archivo y una curaduría atenta, la que permite pensar formas de conocimiento que no diluyen la historia bajo la interoperabilidad, sino que la rearticulan desde una interfaz crítica.

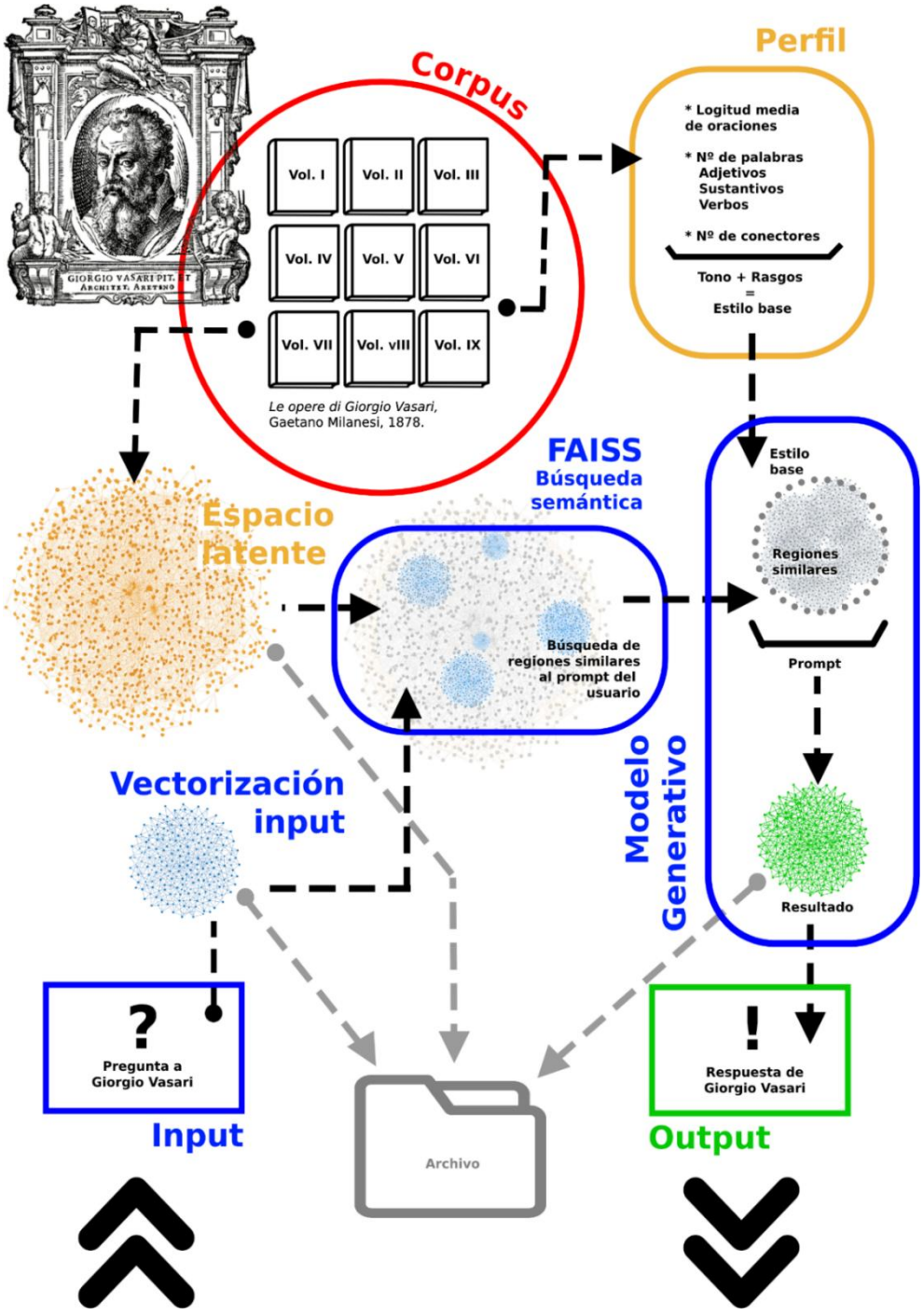


Figura 1. Flujo del experimento de invocación algorítmica con Giorgio Vasari.

```
>>>  Introduce tu pregunta para Vasari:

                                ¿Quién fue Giotto?


>>> Vasari responde:


¡Ah, mio carísimo lector! Preguntáis por Giotto, ese titán, ese
faro que iluminó el camino del arte italiana, sacándola de la
oscuridad medieval para conducirla hacia la esplendorosa aurora
del Renacimiento. ¿Quién fue Giotto? ¡Preguntáis como si se
tratara de un simple artesano! No, amigo mío, fue mucho más que
eso. Fue un revolucionario, un maestro de la pintura que, con su
dolcezza y grazia, transformó para siempre la representación de la
figura humana.


Antes de Giotto, las figuras se pintaban como si fueran iconos
planos, rígidas y sin vida. Se recurría a esquemas bizantinos, a
formas estilizadas que poco se parecían a la belleza y a la
expresión del mundo real. Ma, con la llegada de Giotto, ecco, todo
cambió. Su pincel dio vida a las escenas bíblicas, dotándolas de
una naturalidad, de una profundidad y de un realismo antes
insospechados. Così, las figuras dejaron de ser meros símbolos
para convertirse en seres humanos, con sus gestos, sus emociones,
sus propias historias.


Observad sus frescos en Padua, en la Capilla Scrovegni. Allí,
vedrete la narración de la vida de la Virgen María con una
bellezza inigualable. ¡Mirad el patetismo de las figuras, la
dolcezza en sus rostros! La composición de sus obras es maestra.
```

Figura 2. Extracto de una de las respuestas de la identidad sintética de Vasari.

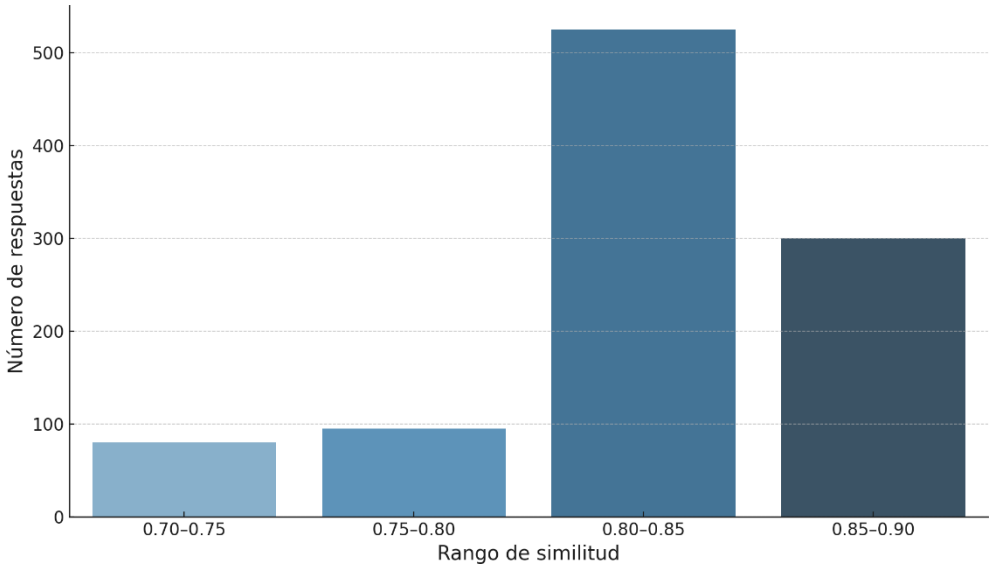


Figura 3. Distribución de la similitud entre respuestas y rasgos estilísticos de corpus.

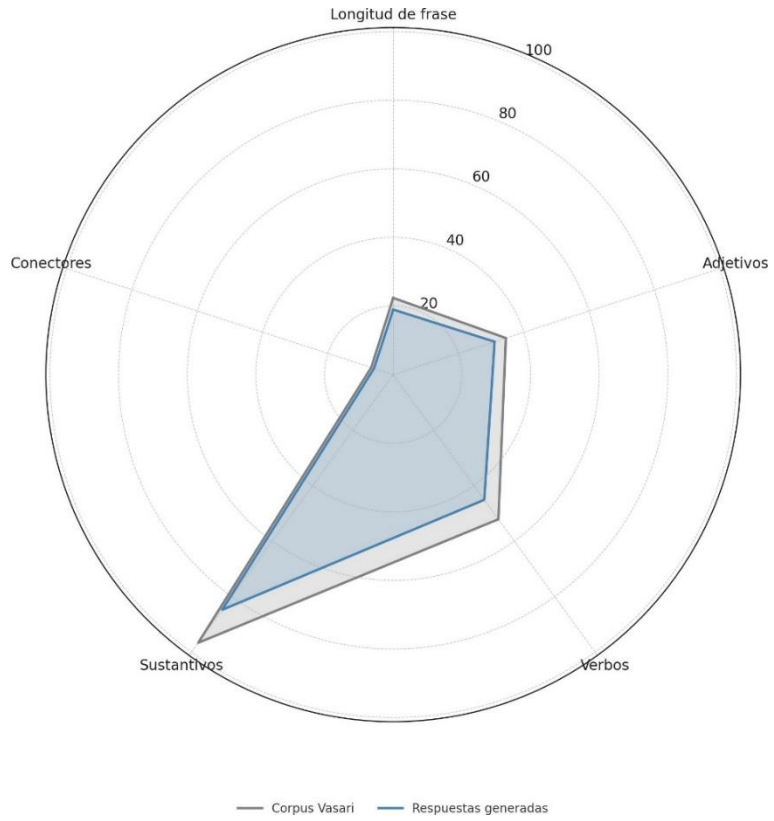


Figura 4. Comparación de atributos lexicométricos.

Bibliografía

ESTEVE, F. (2025), «El enlace inmortal: Duelo, memoria y eternidad digital», *La Nueva Carne*, n.º 5, pp. 48–55.

FERRÁNDEZ, R. (2025), *Inmortalidad digital: Colonizar el planeta Muerte*, Herder.

FLUSSER, V. (2023), *Imágenes técnicas*, Caja Negra.

GOLDSMITH, K. (2019), *Escritura no-creativa: Gestionando el lenguaje en la era digital*, Caja Negra.

MITCHELL, W. J. T. (2019), *La ciencia de la imagen: Iconología, cultura visual y estética de los medios*, Akal.

PARENTE, D. (2024), *Cómo hacer cosas sin palabras: Una filosofía materialista de la técnica*, La Cebra.

PARIKKA, J. (2025), *Imágenes operativas: De la representación visual al cálculo y la automatización*, Caja Negra.

VEGA GARCÍA, P. (2024), «Nigromancia digital e inteligencia artificial», en *Derecho, nuevas tecnologías e inteligencia artificial*, Dykinson, pp. 75–76.

WENG, Y., SHAO, Y., ZHANG, L., LIU, S., DENG, Y., & WANG, J. (2024), «ControlLM: Crafting diverse personalities for language models», *arXiv preprint*, arXiv:2401.10033. https://arxiv.org/abs/2401.10033

ZYLINSKA, J. (2024), *La máquina de percepción: Nuestro futuro fotográfico entre la visión y la inteligencia artificial*, Akal.

Escenas conectadas: *Humanidades Digitales e Inteligencia Artificial para el estudio y difusión del teatro musical en español* (1650–1800)

Juan Pablo Fernández-Cortés
Universidad Autónoma de Madrid

Palabras clave: Teatro musical hispánico; Humanidades Digitales; codificación crítica; archivo performativo.
Keywords: Hispanic musical theater; Digital Humanities; critical coding; performative archive.

El teatro musical en español producido entre 1650 y 1800 constituye un legado cultural de enorme riqueza, pero escasamente visibilizado en los relatos hegemónicos sobre la historia del espectáculo europeo. Fragmentado, híbrido, multimedial, su archivo se conserva a través de libretos sin música, partituras sin contexto, testimonios dispersos y documentos administrativos que rara vez han sido abordados desde una perspectiva integrada. Frente a la mirada canónica que privilegió la ópera italiana o el modelo dramático francés, los géneros lírico-escénicos hispánicos —como la zarzuela o la comedia armónica— han sido históricamente relegados a los márgenes del análisis musicológico, archivístico y digital.

Este proyecto propone una relectura crítica de ese patrimonio desde los marcos teóricos y metodológicos de las Humanidades Digitales, con el objetivo de articular un ecosistema de investigación, edición crítica, visualización computacional y activación escénico-pedagógica del teatro musical hispánico. Se parte de una hipótesis central: que la codificación estructurada de fuentes teatrales (libretos, partituras, documentación contextual), lejos de constituir una tarea técnica neutral, representa una intervención epistemológica que puede transformar radicalmente nuestra relación con el archivo, ampliando su inteligibilidad, su accesibilidad y su potencial de reapropiación cultural. A partir de esta premisa, el proyecto se estructura en torno a cuatro líneas estratégicas: (1) codificación estructurada mediante estándares interoperables (TEI para los textos, MEI para la música), (2) análisis computacional y visualización relacional (aplicando herramientas como R, Gephi y Verovio), (3) construcción de un entorno digital abierto, relacional y documentado, que articule los diferentes estratos de los documentos (textuales, sonoros, materiales, espaciales), y (4) transferencia y activación patrimonial mediante colaboraciones con intérpretes, colectivos artísticos, centros patrimoniales y espacios universitarios.

Descripción del problema

La investigación musicológica sobre el teatro español del Barroco y la Ilustración ha avanzado de manera significativa en las últimas décadas, gracias a ediciones críticas, estudios institucionales y relecturas performativas del repertorio. Sin embargo, sigue existiendo una disociación estructural entre las fuentes literarias y musicales, una escasa digitalización del corpus y una falta de infraestructuras que integren texto, música y representación en formatos digitales abiertos y críticos. Proyectos como EMOTHE, CLEMIT, Nuevo Íncipit de Poesía Española Musicada, Música Poética o las colecciones de la Biblioteca Digital Hispánica han aportado materiales valiosos, pero rara vez han integrado repertorios lírico-musicales ni han aplicado modelos avanzados de codificación cruzada entre texto y música.

El estudio del teatro musical requiere, además, superar el paradigma filológico tradicional, que ha tendido a reconstruir las "obras" como entidades estables y autorales, soslayando la dimensión relacional, performativa e institucional de este repertorio. A ello se suma la necesidad de visibilizar prácticas escénico-musicales periféricas, géneros "menores", autorías colectivas y materiales fragmentarios que no encajan en los modelos normativos de transmisión patrimonial. El problema, por tanto, no es únicamente de acceso, sino de modelo: cómo reescribir la historia del teatro musical desde una lógica no canónica, no lineal, no monumental, sino plural, situada y crítica.

Metodología

El proyecto combina procedimientos de investigación documental, codificación estructurada, análisis computacional y visualización digital, integrados en una metodología modular que permite su replicabilidad y ampliación. La estrategia metodológica se articula en varias fases:

- **Localización y recopilación de fuentes**: se parte de un vaciado sistemático de libretos y partituras conservados en bibliotecas, archivos y repositorios digitales (BNE, Gallica, Europeana, Cervantes Virtual, etc.), aplicando filtros temporales, tipológicos y geográficos. Se emplean herramientas de IA (ChatGPT, Copilot) para inferencia de datos faltantes y se recolectan metadatos mediante OAI-PMH.
- **Codificación crítica**: los textos dramáticos se codifican en TEI-XML y las partituras en MEI, estableciendo correspondencias entre escenas, personajes, acotaciones y elementos musicales. Este trabajo se realiza desde una perspectiva semiótica, consciente de que la codificación es también una forma de interpretación. Se emplean herramientas como Oxygen XML Editor y Verovio.
- **Modelado computacional**: se exploran relaciones de colaboración entre agentes, patrones poéticos-musicales, dinámicas de circulación institucional y geográfica. Se emplean técnicas de análisis de redes

(Gephi) y cartografía digital (QGIS, Leaflet) para visualizar la distribución de repertorios, teatros y trayectorias autorales.

- **Visualización crítica**: los datos se integran en una interfaz digital interoperable, con funcionalidades de navegación textual, sonora y espacial. Se prioriza el diseño de una infraestructura crítica que combine consulta académica con exploración pedagógica y experiencial, mediante visualizaciones interactivas, mapas, líneas de tiempo y grafos de colaboración.
- **Transferencia performativa**: se desarrollan acciones de activación patrimonial en colaboración con intérpretes, compañías de teatro, espacios escénicos y colectivos culturales. Estas acciones incluyen conciertos comentados, microproducciones escénicas, cápsulas audiovisuales, narrativas transmedia y propuestas de reapropiación comunitaria. Se plantea asimismo la integración curricular del archivo digital en asignaturas de grado y máster, como *Humanidades Digitales y Música*, *Dramaturgia del Teatro Musical* y *Documentación Musical*.
- **Evaluación y sostenibilidad**: se prevé una evaluación continua de la plataforma digital y de sus usos pedagógicos y performativos, así como la construcción de una red de colaboración con otros proyectos nacionales e internacionales en HD y patrimonio musical (EMOTHE, RISM, Music Encoding Initiative, etc.).

Resultados obtenidos

Durante la primera fase del proyecto (2024–2025) se ha recopilado y descrito un corpus inicial de más de 120 libretos de zarzuela barroca y comedia armónica conservados en archivos públicos, procedentes de instituciones como la BNE, Biblioteca de Catalunya, Real Academia de la Historia o la British Library. A partir de este corpus se han codificado ya, en versión piloto, tres obras completas en TEI-XML, vinculadas a ediciones parciales en MEI, con anotación de personajes, estructura escénica y metadatos críticos. La visualización mediante Verovio y GraphML ha permitido modelar redes de colaboración autoral (libretistas, compositores, escenógrafos), así como itinerarios de representación en teatros cortesanos y comerciales entre 1670 y 1750. Se han generado, además, tres visualizaciones experimentales que permiten explorar: (a) la evolución de la forma escénico-musical en diferentes periodos, (b) la distribución geográfica del repertorio en el territorio hispánico y virreinal, y (c) las relaciones intertextuales entre escenas, motivos poéticos y estructuras musicales. Estas visualizaciones se han integrado en un prototipo de entorno digital accesible, en desarrollo, que se presenta como *archivo performativo*, es decir, como interfaz viva para la experimentación historiográfica, la docencia crítica y la reapropiación cultural.

Paralelamente, el proyecto ha articulado una propuesta de formación universitaria vinculada al archivo, mediante seminarios, talleres de codificación y módulos docentes en asignaturas de grado y máster.

Conclusiones

El proyecto demuestra que es posible articular una metodología crítica e innovadora para el estudio y difusión del teatro musical hispánico, mediante la convergencia de codificación estructurada, análisis computacional y activación cultural. Más allá de la digitalización como objetivo técnico, se propone un modelo de archivo expandido que permita reescribir la historia del teatro musical desde sus márgenes, visibilizando prácticas olvidadas, géneros excluidos y formas híbridas de creación.

La codificación no se concibe como un fin en sí misma, sino como una tecnología epistemológica que habilita nuevas formas de acceso, lectura y representación. Las decisiones sobre jerarquías textuales, estructuras dramáticas o vínculos entre texto y música son también decisiones políticas sobre qué repertorios se legitiman, cómo se representan y para quiénes se hacen accesibles. La interfaz digital deviene así un espacio de mediación crítica, donde la historia del repertorio no se contempla como un pasado cerrado, sino como una constelación de futuros posibles.

Finalmente, este enfoque permite no solo enriquecer el campo de las Humanidades Digitales hispánicas, sino también proponer un modelo de musicología crítica y situada, capaz de integrar saberes técnicos, sensibilidades patrimoniales, lenguajes audiovisuales y compromisos sociales. El teatro musical en español se revela, así, como un objeto privilegiado para pensar la memoria, la representación y la comunidad en el siglo XXI.

Bibliografía

COTARELO Y MORI, E. (1917), *Orígenes y establecimiento de la ópera en España hasta 1800*, Madrid, Tipografía de la Revista de Archivos.

DABROWSKA, M., & MARTÍNEZ NIETO, J. (2022), «Ediciones digitales de teatro áureo: balance crítico y desafíos metodológicos», *Rilce*, 38(2), pp. 453–480.

DRUCKER, J. (2013), *Graphesis: Visual Forms of Knowledge Production*, Harvard University Press.

FLANDERS, J. (2009), «The productive unease of 21st-century digital scholarship», *Digital Humanities Quarterly*, 3(3).

LAMBEA, M., JOSA, L., & VALDIVIA, F. (2024), *Nuevo Íncipit de Poesía Española Musicada (NIPEM)*, Digital CSIC. https://hdl.handle.net/10261/358755

LAMBEA, M., & JOSA, L. (2023), *Tono barroco. Colección de partituras de Digital CSIC*, Biblioteca Virtual Miguel de Cervantes. https://www.cervantesvirtual.com/portales/literatura_y_musica/tono_barroco/

PEDRELL, F. (1910), *La ópera nacional*, Barcelona.

STEIN, L. K. (1993), *Songs of Mortals, Dialogues of the Gods*, Oxford University Press.

SUBIRÁ, J. (1953), *Historia de la música teatral en España*, CSIC.

UNSWORTH, J. (2000), «Scholarly primitives: What methods do Humanities researchers have in common?», recurso en línea.

VAN ZUNDERT, J. (2012), «If you build it, will we come? Large-scale digital infrastructures as a dead end for Digital Humanities», *Historical Social Research*, 37(3), pp. 165–186.

IA, prensa histórica y literatura hispanofilipina: las HD y el estudio de narrativas postcoloniales en español

Diana Villanueva-Romero
Universidad de Extremadura
Emilio Vivó Capdevila
Universiteit Antwerpen

Palabras clave: Humanidades digitales; IA; prensa histórica; literatura hispanofilipina; narrativas postcoloniales
Keywords: Digital humanities; AI; historical journalism; Hispanic-Philippine literature; postcolonial narratives

Introducción

Rocío Ortuño (2017) ha señalado la naturaleza paradójica de la literatura filipina en español: los autores filipinos que escriben en esta lengua forman parte del canon nacional pese a no contar con lectores actuales y estar marginados del canon occidental y global. Esta paradoja se sostiene sobre una doble articulación: el difícil acceso a los textos originales (por la falta de antologías y estudios críticos) y el propio desprestigio de la lengua española en el archipiélago. Según Ortuño (2017, 74-75), la ausencia de un relevo generacional cristalizó en el archipiélago en un repertorio canónico fosilizado, sin apenas contendientes dentro del polisistema multilingüe filipino y muy marcado por las jerarquías socioculturales de la colonización española.

En este contexto, la prensa en español desempeñó un papel clave: no solo fue el espacio donde publicaron muchos de los principales intelectuales nacionalistas filipinos, sino que siguió siendo un medio vivo hasta finales de los años sesenta, cuando el relevo generacional completó el proceso de cambio lingüístico. En el caso filipino, lejos de ser esferas separadas, la escritura literaria y el trabajo periodístico estuvieron profundamente entrelazados (Sinardet y Ortuño Casanova 2024). Para muchos autores, la prensa fue a la vez un oficio, un primer espacio de publicación y una plataforma de legitimación. Así, el dinamismo de la prensa en español alimentó el auge de la literatura hispanofilipina, marcando sus ejes temáticos y formales, y la decadencia de la primera supuso también el debilitamiento de la segunda.

En esta comunicación, proponemos demostrar cómo el uso de herramientas de inteligencia artificial y humanidades digitales ha permitido reconstruir y ampliar el canon de la literatura filipina en español y revalorizarlo, al tiempo que conocer mejor el contexto sociopolítico en el que se desarrolló.

Hacerlo permite también explicitar y deconstruir las políticas lingüísticas coloniales y la imposición cultural que atravesaron esta literatura en detrimento de las formas de expresión nativas.

Metodología y propósito de la comunicación

Los desarrollos teóricos y el acceso a textos de esta investigación han sido posibles, en parte, gracias al proyecto DigiPhilLit, financiado por el programa Erasmus+, que permitió poner en contacto a distintas universidades europeas activas en el estudio de la literatura filipina en español. Con el apoyo de las humanidades digitales, el proyecto ha contribuido a reconstruir un corpus, prestando especial atención a las fuentes hemerográficas. La investigación que vamos a presentar se enmarca, sin embargo, en un nuevo proyecto, GRESEL (Enriquecimiento de respuestas generadas por aplicaciones IA generativas con explicaciones basadas en características lingüísticas) (PID2023-151280OB-C22) financiado por el Ministerio de Ciencia e Innovación (https://gresel-uned.hypotheses.org). Su objetivo principal es desarrollar un entorno computacional y una infraestructura tecnológica que facilite el uso de herramientas de PLN en dominios específicos como los textos financieros, literarios y periodísticos, con especial atención a la lengua española.

Entre sus objetivos específicos se encuentran:

1. La creación de un marco de software que integre herramientas y algoritmos de PLN.
2. La experimentación con modelos generativos de QA (pregunta-respuesta) aplicados a corpus cerrados.
3. La exploración de nuevas preguntas de investigación en el ámbito del poscolonialismo español, especialmente en relación con el Caribe y Filipinas.
4. El diseño de recursos educativos abiertos (OER) para la formación en HD y PLN.

En nuestra comunicación nos centraremos en ofrecer un ejemplo del trabajo que se ha llevado a cabo desde la fecha de inicio del proyecto—1 septiembre 2024—en lo referido a los objetivos 2 y 3 señalados previamente. Para ello presentaremos dos estudios de caso que nos ayudarán a ejemplificar los retos y complejidades a los que nos enfrentamos al estudiar narrativas y fenómenos sociales coloniales y poscoloniales en el contexto filipino a través de su reflejo en prensa histórica y de su análisis mediante herramientas de HD, IA y PLN.

Estudio de caso 1. La resignificación simbólica de José Rizal a través de las HD

El poeta José Rizal es considerado el eje simbólico de la nación filipina (Rafael 2000, 211). Su obra forma indudablemente parte del canon filipino —en español y traducido al inglés—, y tanto esta como su vida fueron objeto durante

décadas de una lectura asimilacionista e hispanista que minimizaba sus contenidos revolucionarios. Sin embargo, sabemos poco sobre las motivaciones sociopolíticas y estéticas que condujeron a este cambio narrativo en la vida y obra del poeta dentro de la esfera pública hispanoparlante del archipiélago.

Para analizar esto, se ha construido un corpus de referencias literarias y políticas al poeta en la prensa filipina en español. Este corpus ha sido elaborado mediante la consulta de hemerotecas digitales como *Southeast Asian Newspapers* (East View Global Press Archive®), la Universidad de Santo Tomás de Manila, la Universidad de Filipinas Dilimán y diversos fondos físicos en España. Gracias al uso de tecnologías de OCR se ha podido identificar materiales relacionados con palabras clave como “Rizal”, pero no sin algunos problemas metodológicos que se compartirán en el marco de la conferencia: primero, las transcripciones hechas mediante modelos no basados en el español o que solo permiten búsquedas simples; y, segundo, las limitaciones materiales que supone la digitalización y digitación “manuales” de algunas fuentes.

Aun así, esta exploración de la prensa filipina en español entre 1896 y 1944 ha generado la hipótesis, que se presentará en esta comunicación, de que la poesía en español publicada en la prensa filipina fue un espacio privilegiado para resignificar la figura del héroe nacional y sirvió activamente para promover una idea de nación filipina católica (aunque no necesariamente hispanohablante), independiente pero enmarcada dentro de una “*colonialidad del saber*” de raíz eurocéntrica, que negaba las formas culturales nativas (Walsh 2007, 103). Sobre todo, a partir de los años 20, Rizal pasó de simbolizar la intransigencia y el fin del colonialismo español a articular una visión idealizada de la mision civilizadora de España en el archipiélago que fue celebrada primero por las órdenes misioneras españolas y por las autoridades españoles después, pero también por las élites filipinas.

Gracias a la creación de un modelo de Transkribus adaptado a un corpus de prensa y de un lenguaje de etiquetado que permita identificar palabras clave y géneros como la poesía o las novelas por entregas, se pretende generar herramientas de inteligencia artificial adaptadas a este corpus que permitan, primero, poner a prueba las hipótesis anteriores; segundo, ampliar el corpus identificando de forma automática textos vinculados con otras figuras heroicas, especialmente aquellas relacionadas con la revolución de 1896; y, tercero, encontrar textos inéditos en formato libro, ampliando así el canon de la literatura filipina en español. Combinando lecturas distantes y atentas, mi investigación dentro del proyecto GRESEL permitirá seguir, de forma contextualizada, la evolución del significante “Rizal” en la cultura letrada hispanófona.

Estudio de caso 2: La prensa histórica como recurso para estudiar el asociacionismo femenino en Filipinas y la prensa de mujeres

Durante el proyecto DigiPhiLit, las HD facilitaron la comprensión de las circunstancias en las que la escritora Paz Zamora Mascuñana escribió sus

cuentos entre 1919 y 1924. En la investigación publicada al respecto (Villanueva-Romero 2023) se expuso cómo dicha escritora estuvo ligada a algunas de las asociaciones de mujeres de su tiempo interesadas en el bienestar público y en la representación política femenina. En esta comunicación se expondrá la manera en la que el proyecto GRESEL permite avanzar en el estudio del asociacionismo femenino en Filipinas de 1900 a 1930. Esto resulta de interés no solo porque dichas asociaciones lograron importantes mejoras en materia de salud y de educación, sino también porque impulsaron la creación de una prensa femenina y el derecho al voto para la mujer conseguido en 1937.

En el último año, a partir de la hemeroteca digital *Southeast Asian Newspapers* (East View Global Press Archive®), se ha elaborado un corpus acotado al periodo que va de 1900 a 1930, años durante los cuales se produce la americanización del archipiélago, y restringido a publicaciones periódicas en español aún abundantes en las Filipinas de ese tiempo (Godoy 2015). Los hallazgos principales se han obtenido a partir de los periódicos: *El Renacimiento* y *La Vanguardia*. Se han usado los significantes “Paz Zamora”, “Maria Paz Zamora” y “Paz Zamora Mascuñana” ampliándose más tarde al de “Rosa Sevilla de Alvero” por el interés de esta figura en el afianzamiento de un modelo de feminismo filipino en oposición al anglosajón.

A partir de estas búsquedas se ha logrado trazar un mapa de redes de sororidad protagonizadas por asociaciones como la Liga Nacional de Damas y la Federación Nacional de Clubes Femeninos que en general fomentaban la creación de una prensa hecha por y para mujeres. Esta prensa se convirtió en repositorio de lo que activistas como Sevilla de Alvero consideraban ser una “lectura útil” que se expresaba en textos de ficción—cuentos, novelas por entregas, obras teatrales—y de no ficción—ensayos, recetas—que pretendían formar a las filipinas en un ideario de feminismo nacionalista y cristiano.

Teniendo este contexto en cuenta en la comunicación se explicará de qué manera el proyecto GRESEL mediante la creación de un RAG que trabajará con un corpus de prensa histórica puede ayudar a responder preguntas cuantitativas y cualitativas que ayuden a entender tanto las redes de sociabilidad de las activistas filipinas como su expresión periodística y literaria.

Inspirada por los resultados del proyecto “BIESES: Bibliografía de escritoras españolas” se espera también poder llegar a visualizar los resultados relativos a estas redes con herramientas de software libre como Gephi y conocer mejor la producción literaria femenina en las primeras décadas del siglo XX.

Bibliografía

GODOY, A. C. (2015), «La prensa filipina en español entre dos guerras (1899–1941)», *Revista Internacional de Historia de la Comunicación*, 1(4), pp. 22–51.

ORTUÑO CASANOVA, R. (2017), «Philippine literature in Spanish: Canon away from canon», *Iberoromania*, 85, pp. 58–77. https://doi.org/10.1515/iber-2017-0003

RAFAEL, V. L. (2000), *White Love and Other Events in Filipino History*, Durham, Duke University Press.

SINARDET, E., & ORTUÑO CASANOVA, R. (2024), «Prensa y literatura: espacios de diálogo y de intersección», en *Introducción a la literatura hispanofilipina*, London & New York, Routledge, pp. 220–230.

VILLANUEVA-ROMERO, D. (2023), «Nationalism, feminism, and modernity in Maria Paz Zamora-Mascuñana's *Mi óbolo*», *Kritika Kultura*, 42, pp. 114–141. https://doi.org/10.13185/KK2023.004206

WALSH, C. (2007), «¿Son posibles unas ciencias sociales/culturales otras? Reflexiones en torno a las epistemologías decoloniales», *Nómadas*, n.º 26, pp. 102–113.

Los modelos 3D como herramienta de análisis y preservación patrimonial: el caso del Obrador Botifoll (Martorell)

Isidre Ot Padilla
Virtual Past
Esther Travé Allepuz
Universitat de Barcelona – Institut d'Arqueologia

Palabras clave: Modelado 3D; arqueología; alfarería; ciencia de los materiales; digitalización.
Keywords: 3D modeling; archaeology; pottery; materials science; digitization.

La digitalización tridimensional se ha consolidado como una herramienta clave en el ámbito de las Humanidades Digitales, no solo por sus posibilidades de visualización y difusión, sino también por su valor como instrumento de investigación (Hernández-Muñoz 2023). En el marco de un proyecto de investigación interdisciplinario centrado en la producción tradicional de cerámica en Cataluña, presentamos aquí una colección de modelos 3D de piezas elaboradas por el último alfarero del municipio de Martorell, cuya familia –de apellido Botifoll– había regentado el último alfar conocido en el municipio, del que algunos habitantes todavía tienen recuerdo. La colección incluye enseres de fabricación propia y algunos otros adquiridos al por mayor para su venta en el propio taller del alfarero (Figura 1). Nuestro objetivo es poner en valor el patrimonio material e inmaterial asociado a esta actividad, a menudo en manos privadas, y explorar el potencial de una documentación exhaustiva en soportes digitales para la investigación arqueológica de base.

Figura 1. Representación de algunos de los modelos contenidos en la colección del Obrador Botifoll.

La iniciativa forma parte de un proyecto de investigación más amplio **(Figura 2)** que combina arqueología, historia, antropología y ciencias de la información para estudiar la producción de cerámica en Cataluña desde una perspectiva diacrónica, con especial atención a los procesos de resignificación de los materiales, en un contexto como el del siglo XX, donde buena parte de ellos han perdido su función inicial para convertirse en elementos con valor sentimental para las familias que los poseen o como elementos patrimoniales por sí mismos.

En el caso de Martorell, este municipio cuenta con una tradición alfarera documentada desde el siglo XIV – probablemente con algunas interrupciones– hasta su extinción definitiva en los años sesenta del siglo XX (Romero & Rosal 2014). El proyecto se propone reconstruir esta trayectoria productiva a partir de una metodología integrada que incluye vaciado documental, análisis de redes, estudio tipológico y documentación digital.

En este contexto, los modelos 3D se convierten en una herramienta clave tanto para la investigación como para la difusión. Por un lado, facilitan el análisis tipológico y morfométrico de las piezas, permitiendo comparaciones precisas entre formas y funcionalidades. Por otro, permiten la preservación digital de bienes patrimoniales frágiles y, en casos como el nuestro, de acceso restringido, así como su consulta abierta por parte de investigadores, museos y público

general, características que se vuelven fundamentales cuando se trabaja con colecciones temporales.

Figura 2. Imagen corporativa del proyecto ARTIFACTA utilizada con fines de difusión

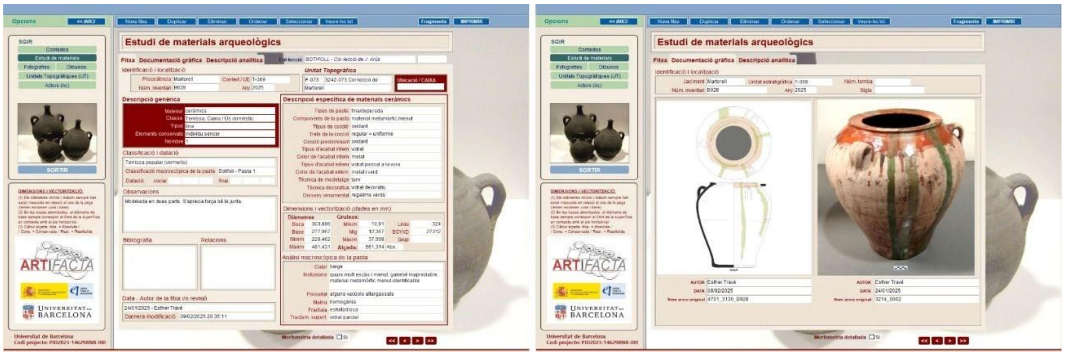


Figura 3. Interfaz de visualización de la base de datos del proyecto.

La colección digital del Obrador Botifoll se ha generado mediante fotogrametría de alta resolución, y cada modelo ha sido integrado en un sistema de información patrimonial interoperable con otras herramientas de gestión (Travé et al. 2021). Este enfoque no solo garantiza la estandarización de los registros, sino que permite su vinculación con otras fuentes documentales, arqueológicas y museográficas. En este sentido, la digitalización tridimensional no se presenta como un mero recurso gráfico, sino como un elemento estructural del flujo de trabajo científico y patrimonial. Los modelos, hoy disponibles en un repositorio público en Sketchfab, van acompañados de fichas técnicas detalladas que describen la morfología, funcionalidad, cronología y procedencia de cada artefacto (Figura 3). Hemos seleccionado, en este caso, muestras

pertenecientes a colecciones privadas, cedidas temporalmente para su estudio. Esta condición hace especialmente relevante su documentación mediante técnicas digitales puesto que garantiza el acceso futuro y público a la documentación de estos materiales, que son devueltos a sus legítimos propietarios tras el pertinente estudio de los mismos (Figura 4).

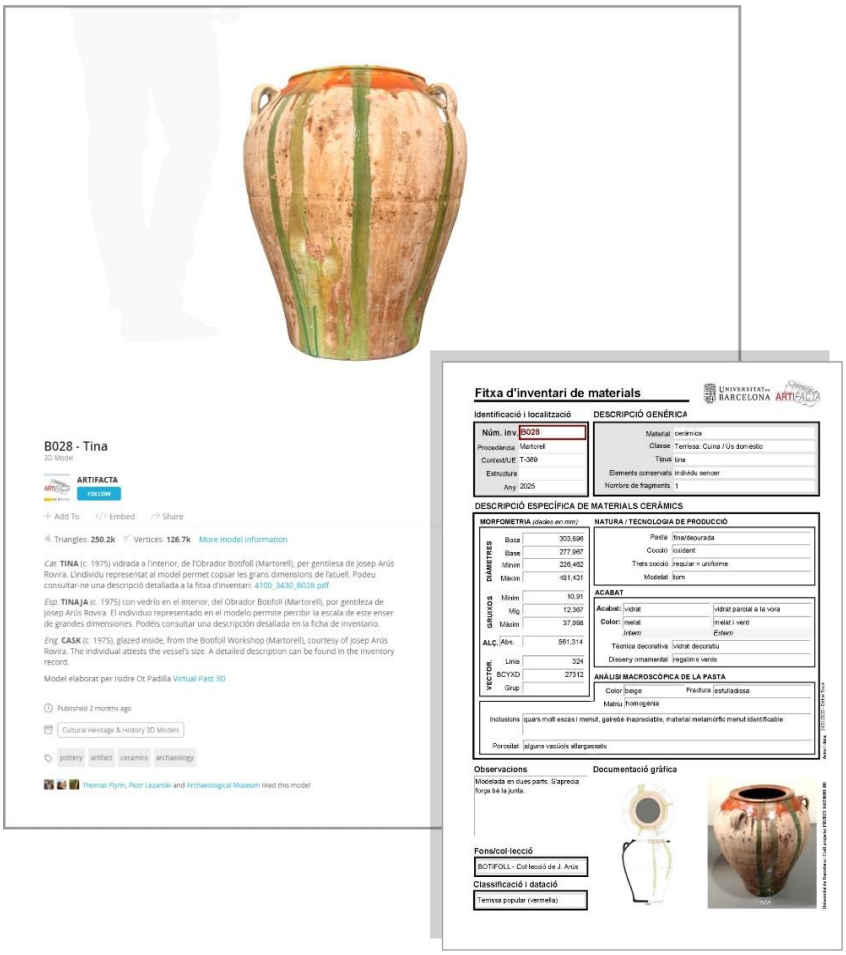


Figura 4. Información que acompaña a cada uno de los modelos publicados, entre la que se incluye una breve descripción en tres idiomas, una selección de descriptores de localización y una ficha detallada del elemento.

Uno de los aspectos más relevantes de esta experiencia es el potencial de los modelos 3D para aportar nuevos elementos al estudio tipológico. La precisión geométrica y la facilidad para el análisis morfométrico (Travé 2022) que permiten estas representaciones digitales favorecen una clasificación y comparación más objetiva y reproducible, posibilitando la identificación de patrones productivos, transformaciones formales y relaciones entre funcionalidad y diseño. Además, la integración de datos cuantitativos (dimensiones, volúmenes, proporciones) y cualitativos (forma, uso, decoración) en un entorno manipulable y visual refuerza el valor interdisciplinario de estas herramientas (Figura 5).

Este trabajo adquiere aún mayor relevancia en un contexto como el de Martorell, donde la producción alfarera tradicional ha desaparecido hace relativamente poco tiempo, y donde todavía es posible rastrear memorias,

relatos y prácticas vinculadas a esta actividad. La colaboración con coleccionistas y propietarios particulares ha sido esencial para la localización, cesión y descripción de las piezas, lo que pone de manifiesto la importancia del trabajo de campo y de la participación comunitaria en la preservación del patrimonio.

Más allá de su valor científico, este repositorio digital cumple también una función esencial en la socialización del conocimiento. La publicación abierta de los modelos, junto con metadatos descriptivos y contextuales, permite una aproximación divulgativa al patrimonio, ofreciendo recursos para instituciones museísticas, centros educativos y público interesado. Esta accesibilidad refuerza una visión democrática del conocimiento y contribuye a visibilizar un patrimonio a menudo olvidado o subestimado.

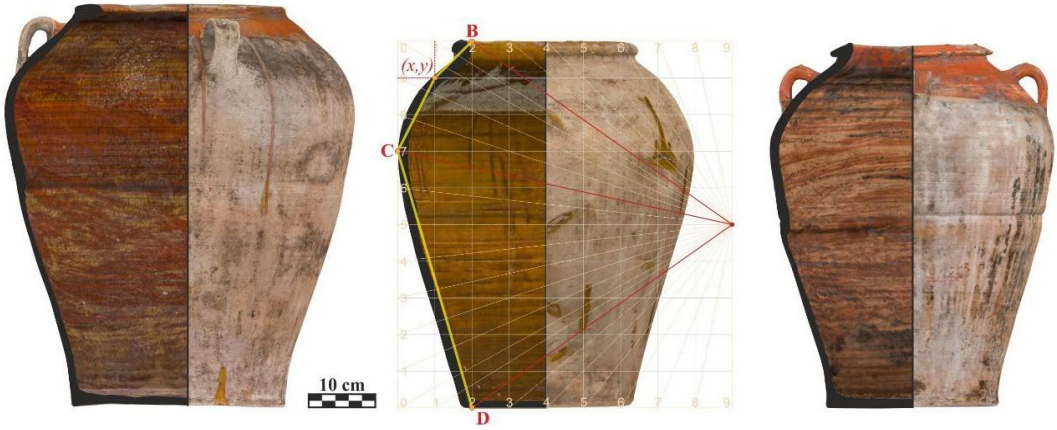


Figura 5. Proceso de estudio tipológico a partir de volúmenes digitalizados.

Sin embargo, en relación a los protocolos y el flujo de trabajo que hemos generado, cabe señalar que no son exclusivos de este trabajo y que, por lo tanto, invitan a la reflexión acerca del propio proceso de digitalización, y de los mecanismos de preservación de la documentación y los expedientes generados. El proceso de documentación de las piezas –tanto la creación de modelos tridimensionales como los dibujos y fichas descriptivas según el método arqueológico tradicional–, la selección de repositorios, o del material que finalmente será publicado, requieren de un proceso de reflexión previo, que es necesario abordar, y que tiene que ver con los volúmenes de información generada, su procesamiento y gestión, así como los criterios para conservar o destruir determinados elementos del expediente.

Solamente para generar un modelo tridimensional de las características como los que aquí presentamos, se generan una cantidad ingente de imágenes que oscilan entre las 300 y las 400 fotografías de alta resolución en RAW (a las que cabe sumar la cantidad equivalente de imágenes reveladas en .jpg), lo que supone un gran volumen de información que requiere de una capacidad de almacenamiento cada vez mayor y de protocolos detallados para asegurar su trazabilidad y reutilización. Por ejemplo, los trabajos relativos a las 39 piezas

digitalizadas de la colección, han generado un total de 47794 archivos (incluyendo imágenes en RAW, jpg, archivos de procesado, revelado y resultados de trabajo), lo que nos da un promedio de 1225 archivos por pieza. Asimismo, la realización de dibujos, la toma de medidas y la descripción alfanumérica de las piezas también genera expedientes documentales que es necesario procesar y almacenar, en entornos interoperables que garanticen la preservación de todo el proceso de trabajo, y que permitan el acceso y recuperación fácil de la documentación original. Garantizar la normalización y trazabilidad de estos procesos de documentación también exige de un proceso de reflexión teórica que conviene no eludir (Del Fresno et al. 2024).

Nuestra contribución pretende aportar nuestra experiencia en este campo, a partir de la realización de un proyecto concreto y ejemplificar cómo las herramientas digitales no solo transforman la forma en que documentamos y conservamos el patrimonio material, sino también cómo lo conocemos, lo analizamos y lo compartimos. El repositorio del Obrador Botifoll es una muestra del enorme potencial que estas técnicas tienen para preservar colecciones efímeras, a menudo invisibilizadas, que solo pueden darse a conocer, especialmente a un público amplio, gracias a su digitalización.

La documentación arqueológica en este sentido, especialmente en lo referido al estudio y catalogación de materiales, acaba teniendo un valor patrimonial intrínseco que nos lleva a concebir este tipo de colecciones no solo como archivos visuales, sino como espacios dinámicos de investigación y diálogo interdisciplinario. Su construcción debe atender a criterios de rigor científico, pero también a una sensibilidad comunitaria que garantice que el patrimonio que preservamos sea también patrimonio que interpela, que conecta y que transforma. Reflexionar sobre este tipo de procesos y marcar algunas líneas de actuación puede contribuir a generar procesos de normalización (Dallas 2009, Medina 2024).

El proyecto contempla también la realización de una exposición monográfica dedicada al Obrador Botifoll, que permita contextualizar históricamente la actividad alfarera en Martorell y presentar los resultados de la investigación. En este marco, uno de los principales retos reside en definir cómo integrar los modelos 3D dentro del espacio expositivo de manera significativa. No se trata únicamente de mostrar los contenidos digitales como un añadido tecnológico, sino de explorar fórmulas museográficas que potencien su valor interpretativo y su impacto en la experiencia del visitante. Esta decisión implica una reflexión crítica sobre la relevancia de los modelos en relación con los discursos curatoriales, así como una evaluación realista de los costes asociados a su producción, implementación técnica y mantenimiento. En este sentido, será necesario valorar si su inclusión aporta un valor añadido suficiente —en términos de accesibilidad, interactividad o comprensión del objeto— como para justificar los recursos invertidos, especialmente en un contexto en el que las instituciones culturales deben optimizar tanto sus presupuestos como sus estrategias de mediación.

Bibliografía

DALLAS, C. (2009), «From artefact typologies to cultural heritage ontologies: Or, an account of the lasting impact of archaeological computing», *Archeologia e Calcolatori*, 20, pp. 205–221.

DEL FRESNO, P., MEDINA, S., & TRAVÉ, E. (2024), «HORAI: An integrated management platform for historical information», en *CAA2023 Conference Proceedings (Computing Applications in Archaeology, Ámsterdam, 3–6 abril 2023)*, PCI Archaeology, Zenodo. https://doi.org/10.5281/zenodo.10822889

HERNÁNDEZ-MUÑOZ (2023), «Analysis of digitized 3D models published by archaeological museums», *Heritage*, 6, pp. 3885–3902.

MEDINA, S. (2024), *Paisajes de montaña y periferias en la Alta Edad Media. Procesos de gestión de la información para un relato histórico integrado* [Tesis doctoral], Universidad de Barcelona. https://www.tdx.cat/handle/10803/694214#page=1

ROMERO, A., & ROSAL, J. (2014), *La Terrissa a Catalunya*, Figueres, Brau.

TRAVÉ, E., MEDINA, S., DEL FRESNO, P., VICENS, J., & MAURI, A. (2021), «Towards an ontology-driven information system for archaeological pottery studies: The GREYWARE experience», *Applied Sciences*, 11, art. 7989. https://doi.org/10.3390/app11177989

TRAVÉ, E. (2022), «Statistical analysis of morphometric data for pottery formal classification: Variables, procedures, and digital experiences of medieval and postmedieval greyware clustering in Catalonia (Twelfth–Nineteenth centuries AD)», *Open Archaeology*, 8, pp. 1269–1285.

ChemiNova: Nuevas Herramientas para la Inspección y Documentación del Patrimonio Cultural en Riesgo

Carmen Cano
Belén Palma
Esperanza Villuendas
Laya Targa
Álvaro Solbes
Sergio Casas
Ester Alba
Cristina Portalés
Universitat de València

Palabras clave: Conservación; Realidad Aumentada; Realidad Virtual; interdisciplinariedad.
Keywords: Conservation; Augmented Reality; Virtual Reality; Interdisciplinarity.

Introducción

A pesar de los avances logrados por la comunidad científica, las estrategias para la protección del patrimonio cultural (PC) frente a amenazas naturales y antrópicas no están integradas de forma exhaustiva en los Planes Nacionales de adaptación al cambio climático y reducción de riesgos (Comisión Europea, 2022). Una de las razones de esta brecha persistente es la falta de una metodología estandarizada y comúnmente aceptada para el monitoreo y la cuantificación del daño y la pérdida del patrimonio cultural.

El proyecto ChemiNova, liderado por un equipo interdisciplinar de la Universitat de València, tiene el propósito de abordar algunos de estos retos. ChemiNova es un proyecto de investigación financiado bajo el programa Horizonte Europa (ChemiNova, 2024), que está centrado en el diseño y la implementación de un conjunto de aplicaciones informáticas para ayudar a mejorar la conservación, el análisis y el monitoreo del patrimonio cultural. Mediante la adquisición de diversos tipos de datos, el proyecto ayuda a abordar los daños estructurales y químicos a los bienes culturales, con especial atención al cambio climático. Los daños a los materiales de construcción patrimoniales causados por variaciones lentas y continuas de los parámetros climáticos y de contaminación son una consideración importante (Bonazza y Sardella, 2023).

Metodología

ChemiNova implica la estrecha cooperación de un equipo interdisciplinario que abarca áreas como conservación, historia del arte, arquitectura, restauración, desarrollo de capacidades, interpretación del patrimonio, química, biología, informática, ingeniería multimedia y fotogrametría. Además, el consorcio es diverso tanto a nivel geográfico (España, Italia, Francia, Alemania, Austria, Rumanía, Ucrania y Grecia) como a nivel institucional (cinco universidades, un instituto de investigación, una ONG, dos museos, dos pymes y una institución intergubernamental), siendo un total de doce socios. El trabajo en equipo de expertos de diferentes campos del conocimiento tiene como objetivo llevar nuevas tecnologías —incluyendo la Realidad Aumentada (RA) y la Realidad Virtual (RV)— al ámbito de la conservación de nuestro patrimonio cultural, de tal forma que puedan integrarse en los flujos normales de trabajo.

En el proyecto se elaboran modelos digitales tridimensionales enriquecidos, a los que hemos llamado e3D, y que consisten en una serie de capas de información mapeadas sobre una geometría base. En el proyecto, consideramos las siguientes dimensiones en los e3D:

- Espacial/geométrica: nube densa de puntos (por ejemplo, obtenido a través de un escáner 3D).
- Espectral: imágenes en diferentes espectros (visible, infrarrojo cercano, térmico, etc.).
- Fotométrica: Información generara mediante RTI (*Reflectance Transformation Imaging*).
- Semántica: estimaciones de daños en bienes del patrimonio cultural (derivados mediante un módulo de inteligencia artificial).
- Temporal: Registrar los bienes en distintos momentos.
- Otras dimensiones: por ejemplo, ED-XRF, espectroscopías FT-NIR.

El proyecto cuenta con cuatro pilotos (en España, Austria, Italia y Ucrania), los cuales proporcionan datos para la confección de los e3D. Estos modelos digitales, serán integrados en diversas aplicaciones informáticas para distintos usos. En este artículo, centramos nuestra atención en dos de estas herramientas, las cuales están siendo desarrolladas por la Universitat de València:

- ChemiInspection: está diseñado para su uso con dispositivos portátiles y se centra en la inspección in situ de artefactos. La aplicación emplea la tecnología de RA para facilitar una inspección guiada siguiendo procedimientos específicos —como en García-Pereira, 2020)—. La aplicación permite la digitalización de informes de estado y la determinación de la relación espacial entre la información virtual y los elementos del mundo real durante la inspección in situ.
- ChemiSensing: Entorno colaborativo multiplataforma —como en (García-Pereira, 2018)— en el que usuarios locales y remotos podrán trabajar conjuntamente en la inspección y documentación de daños. En

un entorno virtual, el área bajo investigación será recreada utilizando modelos e3D.

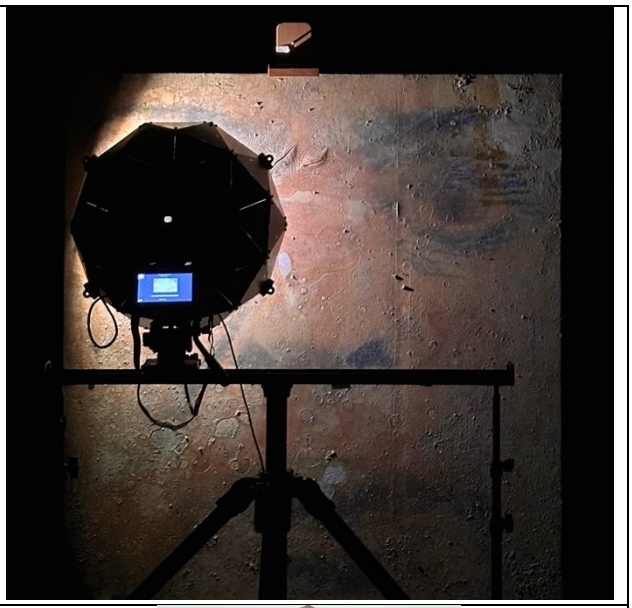
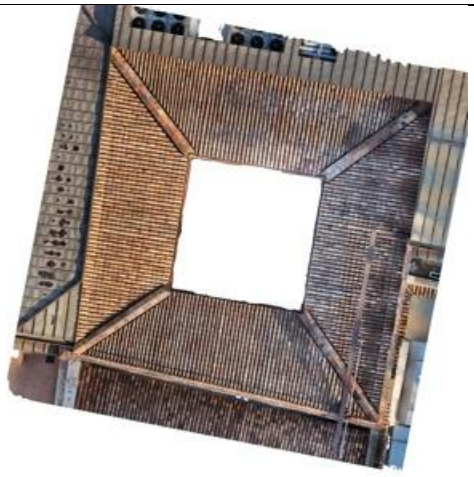
Resultados

En este apartado mostramos los primeros resultados del piloto de España. Este piloto en un principio se centraba en el edificio histórico La Nau (de la Universitat de València) y alguna de sus colecciones, como la Colección contemporánea de Martínez Guerricabeitia. Sin embargo, tras la riada del 29 de octubre del año 2024 causada por una DANA, se decidió ampliar el piloto a obras dañadas por esta causa. Centraremos nuestra atención no solo en la adquisición de los datos, sino también en el resto de las actividades llevadas a cabo o previstas, haciendo también hincapié en los aspectos interdisciplinares, los retos y el valor añadido.

Adquisición de datos:

La primera adquisición de datos para el proyecto coincidió con el comienzo del mismo (durante el llamado *kick-off meeting*), aprovechando la ocasión de que el consorcio entero se reunió en Valencia. Tras esta primera campaña, desde la Universitat de València hemos tomado más datos según se ha requerido. En estos momentos estamos planificando otra recolección de datos por parte de todos los socios, que probablemente tendrá lugar en octubre del 2025. En la Tabla 1 puede verse el resumen de la toma de datos para algunas de las obras que serán posteriormente integradas en ChemiSensing y ChemiInspection.

Objeto patrimonial y características principales	**Datos adquiridos (hasta mayo 2025) y tecnologías relacionadas**	**Imagen de ejemplo y/o imagen tomada durante la captura de datos**
Peana de la estatua de Lluís Vives.	Geometría, imagen visible, imagen hiperespectral, RTI, imagen térmica IR, Fluorescencia de rayos X	

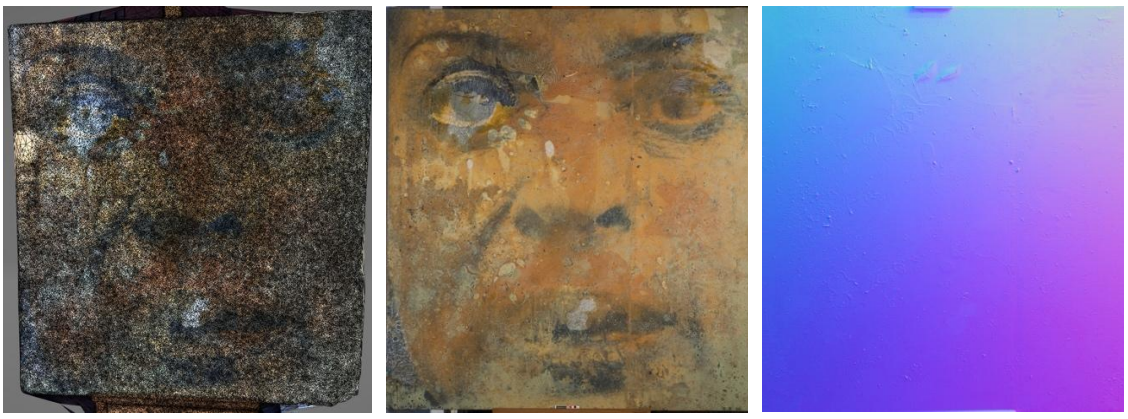
Retrato de Lola Flores. Dimensiones: 128 x 118 cm Autor: Darío Villalba	Imagen visible, imagen hiperespectral, RTI, imagen multibanda, reflectografía infrarroja, nube de puntos a partir de LiDAR	
Fuente del Claustro	Imagen visible, imagen hiperespectral, nube de puntos a partir de LiDAR	
Tejado de La Nau	Imagen visible y nube de puntos a partir de LiDAR	

Tendur, Maria, esposa Dimensiones: 81 x 60 cm Autor: Francesc Sempere Ferández de Messa (Obra afectada por la DANA)	Imagen visible, imagen hiperespectral, imagen multibanda, reflectografía infrarroja, fotogrametría	

Tabla 1. Relación de objetos patrimoniales documentados en el piloto de España (desde el *kick-off meeting* del proyecto hasta mayo del 2025).

Generación de modelos e3D:

En estos momentos, se prevé tener una primera versión de los modelos e3D de los objetos relacionados en la Tabla 1 para finales de julio 2025 (integrando al menos la geometría y diferentes capas de imagen), y una versión mejorada (con todos los datos integrados) para noviembre del 2025. Por otro lado, para alguno de los objetos ya tenemos una primera versión, como es el caso del cuadro de Lola Flores. En la Figura 1 podemos ver un detalle del cuadro atendiendo a su geometría, imagen visible, y el mapa de normales realizado con la técnica del RTI.



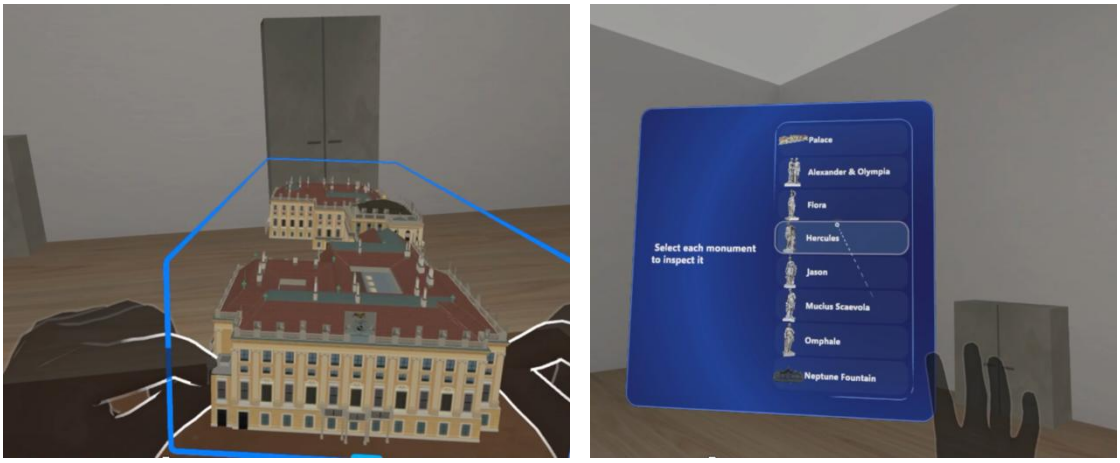
(a) (b) (c)

Figura 1. Detalle del cuadro de Lola Flores, donde: (a) geometría; (b) imagen visible; (c) mapa de normales (RTI).

Demos de RA y RV y primer contacto con expertos:

Ya en la primera reunión de proyecto se implementaron dos entornos interactivos, uno basado en la RA y otro en la RV. El objetivo es que los socios del proyecto pudieran interactuar con estas tecnologías, para poder obtener sus primeras impresiones de cara al diseño e implementación de las aplicaciones ChemiInspection y ChemiSensing. Es por ello por lo que nos referimos a ellas como “demos” (demostradores).

En cuanto a la demo de RV, se preparó un entorno virtual para interactuar con objetos patrimonio, en esta ocasión, del palacio de Schönbrunn (que constituye el piloto de Austria). Los usuarios llevaban puestas unas gafas de RV y la interacción con los objetos se realizaba de forma natural, mediante movimientos de las manos (Figura 2).



(a) (b)

Figura 2. Detalle de dos momentos durante la interacción de uno de los usuarios con el entorno virtual desarrollado para el kick-off meeting del proyecto, donde: (a) rotación y escalado del modelo; (b) selección de un objeto a partir de un panel que muestra la lista de estatuas disponibles.

En cuanto a la demo de RA se prepararon dos entornos, uno para la estatua de Lluís Vives (Figura 3), y el otro para la fuente del pórtico. En este caso, el dispositivo empleado es una tableta (iPad Pro).

Figura 3. Entorno de RA desarrollado para el kick-off meeting del proyecto, en el entorno de la estatua de Lluís Vives. En la parte derecha podemos ver los dos escenarios disponibles (fuente del pórtico y estatua de Lluís Vives). En la parte inferior podemos ver otros parámetros para seleccionar que permiten cambiar la textura de la estatua

Conclusiones

En este artículo se ha presentado el piloto de España del proyecto ChemiNova, que está centrado en el edificio de La Nau de la Universitat de Valencia junto con alguna de sus colecciones. Se ha presentado el trabajo realizado hasta mayo del 2025, desde la adquisición de datos, modelos e3D (ejemplo básico) y aplicaciones demo de RA y RV.

Como ya se ha comentado anteriormente, la siguiente campaña de adquisición de datos de todos los socios en La Nau está prevista para octubre de este mismo año. Los datos adquiridos serán integrados en los e3D para dotarlos de más capas de información. Posteriormente (finales de noviembre), está previsto hacer la primera evaluación de ChemiInspeciton y ChemiSensing en La Nau, puesto que está previsto que dispongamos ya de las versiones funcionales de ambas herramientas.

Financiación

ChemiNova ha recibido financiación del Programa Marco Horizonte Europa de la Unión Europea en virtud del acuerdo de subvención 101132442.

Bibliografía

BONAZZA, A., & SARDELLA, A. (2023), «Climate change and cultural heritage: Methods and approaches for damage and risk assessment addressed to a practical application», *Heritage*, 6(4), pp. 3578–3589. https://doi.org/10.3390/heritage6040190

CHEMINOVA (s. f.), recurso en línea. https://cheminova.eu/ [Acceso: 29/05/2005].

COMISIÓN EUROPEA, Directorate-General for Education, Youth, Sport and Culture (2022), *Strengthening cultural heritage resilience for climate change – Where the European Green Deal meets cultural heritage*, Publications Office of the European Union. https://data.europa.eu/doi/10.2766/44688

GARCÍA-PEREIRA, I., GIMENO, J., PÉREZ, M., PORTALÉS, C., & CASAS, S. (2018), «MIME: A mixed-space collaborative system with three immersion levels and multiple users», *International Symposium on Mixed and Augmented Reality (ISMAR)*.

GARCÍA-PEREIRA, I., PORTALÉS, C., GIMENO, J., & CASAS, S. (2020), «A collaborative augmented reality annotation tool for the inspection of prefabricated buildings», *Multimedia Tools and Applications*, 79, pp. 6483–6501.

Recreando un final del Camino que no pudo ser. El modelo virtual de la *Rúa Galicia* ideada por Antonio Palacios para Santiago de Compostela (1932)

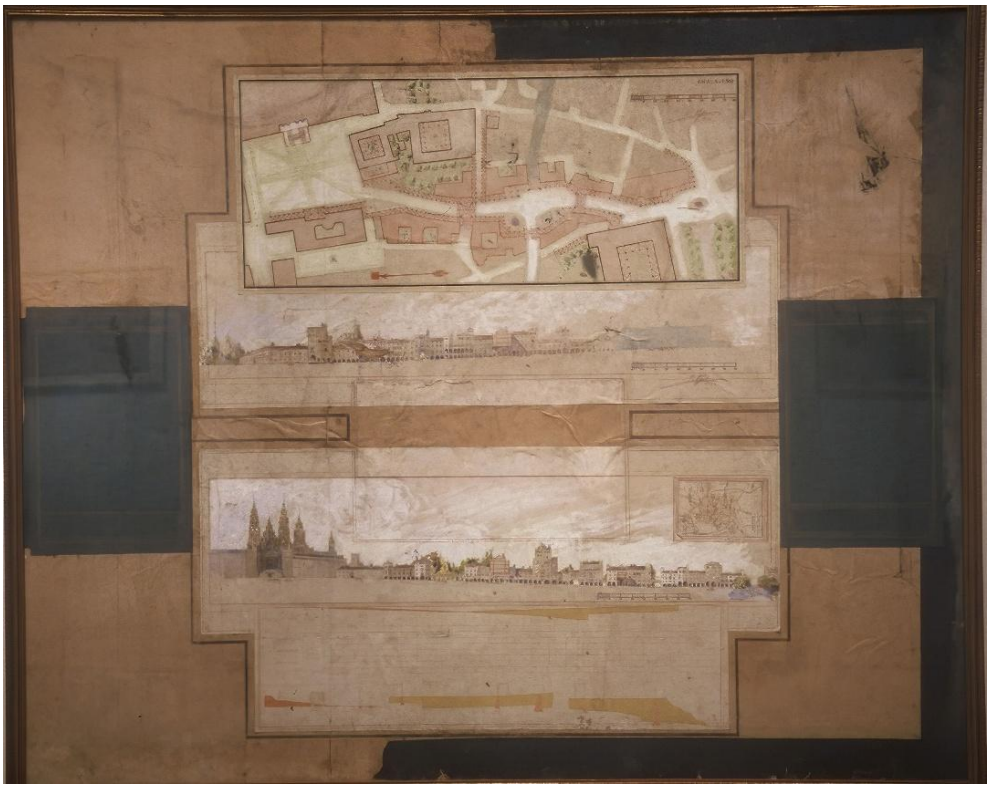
Jesús Ángel Sánchez García
Universidade de Santiago de Compostela

Palabras clave: Recreación virtual; propuesta urbana; Santiago de Compostela; Antonio Palacios; modelos 3D
Keywords: Virtual recreation; urban proposal; Santiago de Compostela; Antonio Palacios; 3D models

Problema de investigación

El arquitecto gallego Antonio Palacios Ramilo presentó en el año 1932 una ambiciosa propuesta de nueva avenida para canalizar el creciente flujo de turistas que arribaban a la plaza del Obradoiro en Santiago de Compostela. Con el objetivo de evitar que coches y autobuses cruzaran las viejas rúas asoportaladas, esta avenida de 300 m de longitud, bautizada como *Rúa Galicia*, ofrecería un acceso alternativo combinando distintos tramos y plazas desde el colegio de San Clemente hasta el lateral del palacio de Raxoi (Costa Buján 2016, 138-144; Sánchez García 2022). Entregada al ayuntamiento, la idea se sometió a varios reajustes durante los dos años siguientes con el fin de completar las viviendas y tiendas previstas inicialmente con las sedes de varios edificios públicos que necesitaba la ciudad. Las adhesiones de distintos sectores locales hacia la que se estimaba como una mejora para la "monumentalidad artística" de Compostela, sin olvidar la oportunidad de empleo en tiempos de crisis económica, se tradujeron en un apoyo general y la aprobación final del anteproyecto en la sesión municipal del 13 de marzo de 1936.

Aun contando con ese prometedor contexto, las circunstancias de la insurrección militar y guerra civil desencadenada a partir de julio de 1936 terminaron por aparcar la idea, de modo similar a lo ocurrido con otros proyectos de Antonio Palacios como la plaza a crear ante la catedral de Ourense, o el utópico plan de reforma y ensanche de Vigo (Sánchez García 2021 y 2022). Los dibujos originales fueron almacenados, conservándose hoy en día en una composición enmarcada de gran tamaño en el Museo do Pobo Galego (Fig. 1).



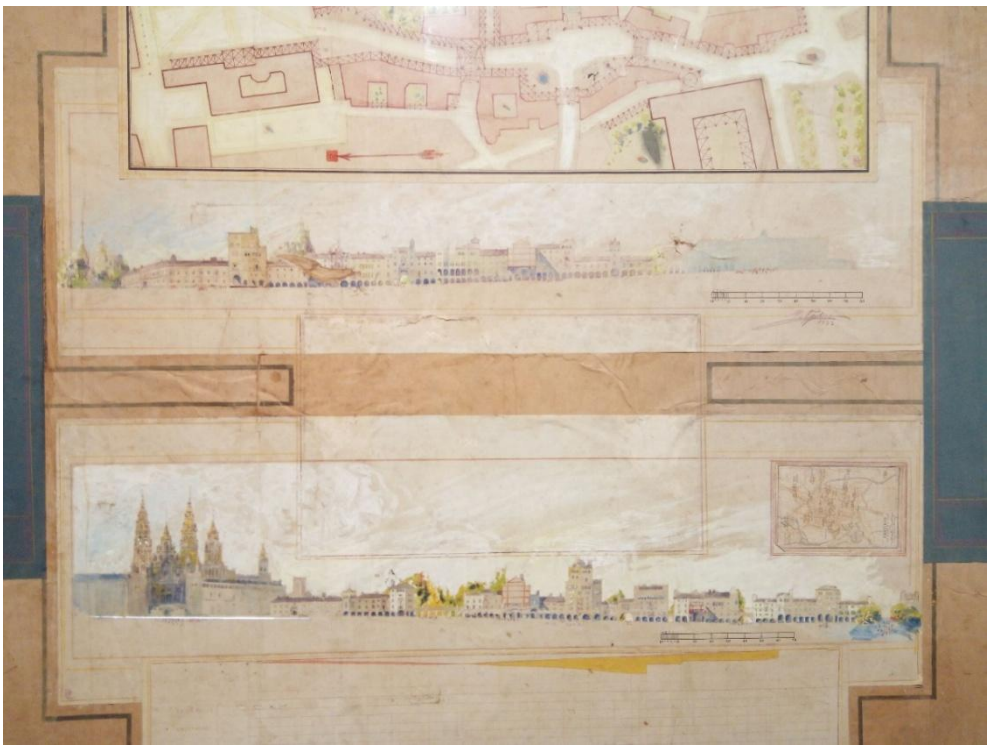
La radical transformación que se hubiera operado en el casco histórico de Santiago de Compostela mediante la intervención soñada por Antonio Palacios aconsejó incluir este caso entre las muestras más destacadas de arquitecturas utópicas elaboradas en Galicia dentro del proyecto ARSOGAL - *Arquitecturas soñadas. Proyectos fallidos y pensamiento utópico en la historia de la arquitectura en Galicia*, Código PID2022-137098NB-I00 -, financiado para el período comprendido entre los años 2023-2027 por el Ministerio de Ciencia e Innovación. Continuador de trabajos previos dedicados a las arquitecturas desaparecidas de Galicia, ARSOGAL es un proyecto de investigación en Historia del Arte en el que se integran las herramientas de diseño y modelado 3D como parte de las metodologías de análisis, investigación y difusión, de modo que se adscribe al nuevo paradigma de conocimiento de las Humanidades Digitales. En concreto, el trabajo desarrollado sobre los planos de Antonio Palacios ha permitido acometer la recreación virtual de todo el trazado de la *Rúa Galicia*, facilitando así una mejor comprensión de la ambiciosa propuesta, con un impacto previsto a través de diferentes soportes e iniciativas dirigidas tanto a especialistas en historia de la arquitectura como al gran público.

Metodología

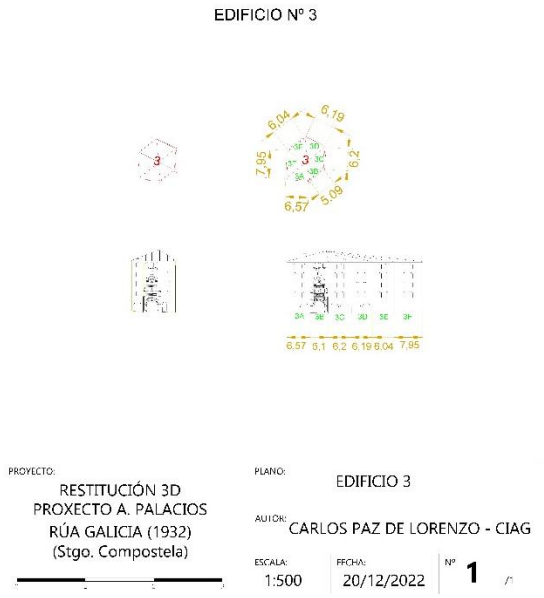
Asumiendo como marco metodológico los principios de la *Digital Art History*, el trabajo sobre la propuesta de *Rúa Galicia* ha buscado la intersección entre las fuentes tradicionales de conocimiento histórico y las tecnologías digitales para todo el arco de acciones comprendidas desde la organización, tratamiento y análisis de información a la difusión de resultados (Rodríguez Ortega 2019;

Milligan 2022, 7-9). La preparación de los materiales documentales generados para la presentación y tramitación del anteproyecto de la *Rúa Galicia* incluyó la localización y transcripción de la memoria escrita por Antonio Palacios, así como la ampliación y mejora digital de todos los planos elaborados entre 1932 a 1936. Junto a los planos que contienen el desarrollo en superficie de la nueva avenida, los dibujos enmarcados para su exhibición en el Museo do Pobo Galego han constituido la fuente gráfica principal a escanear, tratar y analizar (González Gracia y Angulo Fornos 2020). Gracias a un escaneado de alta resolución de dichos dibujos realizado con motivo de la exposición *Antonio Palacios. Soños de Modernidade. Compromiso con Galicia* (Sánchez García 2020), los alzados completos de los edificios que bordearían la nueva avenida fueron tratados para eliminar manchas e incrementar su definición. A partir de esta fase entró en juego la aportación del diseñador gráfico Carlos Paz de Lorenzo, profesional integrado en el Centro Infográfico Avanzado de Galicia -CIAG-, que ya cuenta con dilatada experiencia de colaboraciones en los proyectos dirigidos por el autor de la comunicación, destacando el proyecto Portomarín Virtual, en coautoría con Julio Vázquez Castro, profesor en la Universidade de Santiago de Compostela y especialista en arquitectura medieval, que mereció en 2024 el premio a la mejor herramienta o recurso concedido por la sociedad HDH.

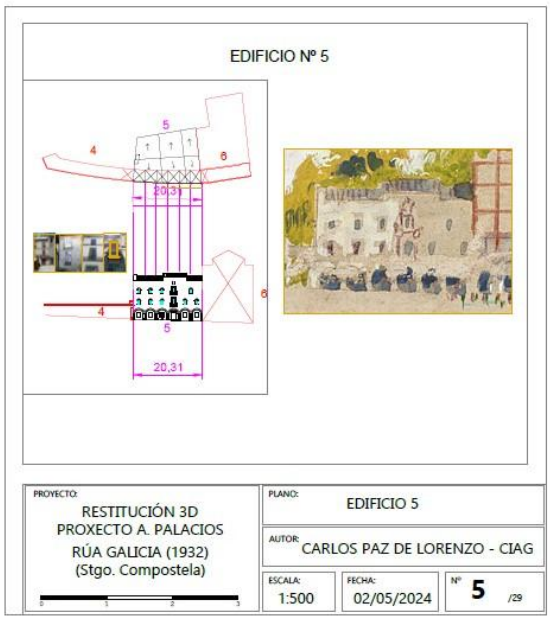
En una primera fase se procedió a aislar cada edificio previsto por Palacios, redibujando planta y alzado principal en Autocad 2D, ajustando con total precisión su ubicación dentro de los quiebros y cambios de dirección de todo el recorrido de la nueva vía por la topografía y entorno construido en el sector occidental de la ciudad histórica de Santiago de Compostela, tal como se encontraba en el año 1932 (Fig. 2).

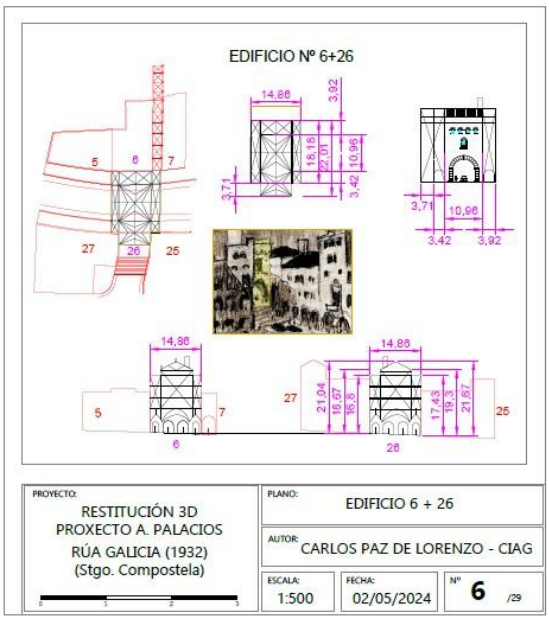


Uno a uno, cada edificio dibujado por Palacios fue interpretado gráficamente en 2D para obtener las líneas compositivas y de contorno de su fachada, con sus vanos, adornos y cubiertas (Fig. 3). En estas fases iniciales del trabajo digital sobre las fuentes gráficas radicó una primera dificultad, puesto que los diseños de Palacios se conservan en el nivel de anteproyecto, es decir, con una definición que aporta los volúmenes y formas esenciales, pero que deja abocetados muchos detalles decorativos. Además, en algunos casos las roturas y pérdidas de papel en los planos originales obligaban a completar elementos desaparecidos.

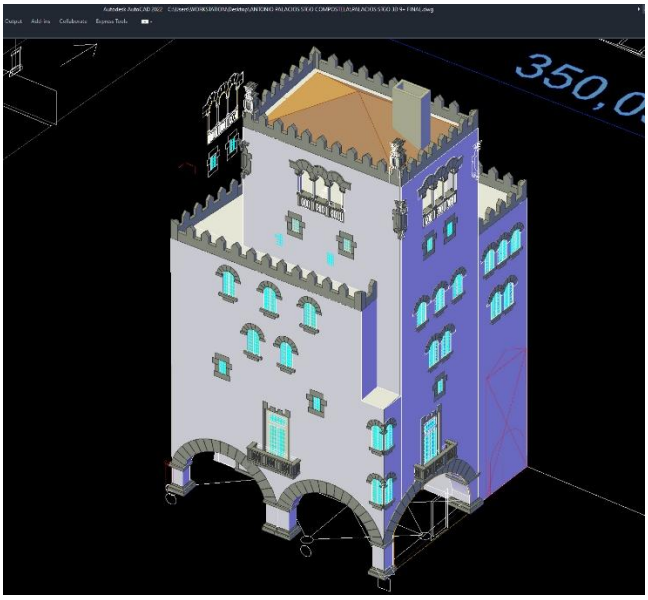


Con el fin de lograr la restitución más fiel a las ideas de Antonio Palacios, la metodología de interpretación recurrió a buscar inspiración en elementos de tipo historicista presentes en otros proyectos regionalistas de este arquitecto -Ayuntamiento de O Porriño, Central del Tambre, Iglesia de Panxón…-, combinados con modelos reales existentes en la propia ciudad de Santiago de Compostela. La intención de lograr una armonía y perfecta integración estética de su *Rúa Galicia* con el resto del tejido de la vieja Compostela, explicitada por Antonio Palacios en la memoria de acompañamiento, justificó esta selección de formatos de arcos, puertas, balcones, ventanas, escudos y almenas replicados a partir de modelos identificados en todo el conjunto histórico. En alguna ocasión, casos más minoritarios, los elementos que sirvieron de inspiración se localizaron en muestras de la arquitectura gallega y española del Renacimiento al Barroco en tipologías como pazos -pazo de Fefiñáns, pazo de Sistallo-, o bien en la arquitectura civil de ciudades históricas como Betanzos, Toledo y Salamanca que Antonio Palacios apreciaba y conocía bien (Fig. 4).





La revisión y corrección de los alzados 2D dio paso a la creación de los modelos 3D, preparados en un primer momento como edificios aislados para definir sus volúmenes básicos (Fig. 5).



El ajuste de los distintos elementos compositivos y ornamentales implicó otra fase de trabajo de gran complejidad para afinar el estilo y rasgos formales presentes en cada fachada. En un tercer momento se procedió al montaje de los distintos grupos de edificios que integrarían las manzanas, y finalmente el montaje completo de las dos hileras de edificaciones, hasta definir toda la avenida y sus plazas interiores (Fig. 6). El proceso de producción, combinando hasta aquí Autocad, 3Ds Max, Sketchup y Blender, dio paso a la generación de texturas específicas de arquitectura tradicional con la ayuda de Photoshop, Materialize y CrazyBump. Se utilizó también Reality Capture para incorporar algunos elementos fotogramétricos de gran complejidad. Para completar la

ambientación hiperrealista, se modelaron y añadieron elementos accesorios como farolas de iluminación, fuentes, una columna ostentando la placa con el nombre de la vía, un monumento situado al inicio y unos pavimentos pétreos con tratamientos diferenciados para peatones y vehículos. Finalmente, dado el elevado nivel de geometría requerida para este proyecto, se empleó Twinmotion, software que usa el motor de videojuegos e infoarquitectura UnrealEngine, especializado en el procesamiento por GPU, acelerando la generación de imágenes infográficas y la integración del modelo urbano de la calle en sus distintas perspectivas.

Logradas las texturas definitivas para cada material y los detalles más realistas en superficies e iluminación, las revisiones de los sucesivos renderizados 3D se realizaron en archivos jpg 1600x900, como soporte más manejable para la fase de correcciones de detalles.

Resultados obtenidos

La interpretación y restitución 3D de todo el recorrido de la *Rúa Galicia* se desarrolló entre los meses de diciembre de 2023 a julio de 2024, siempre a cargo del diseñador Carlos Paz y con la supervisión del autor de esta comunicación e Investigador Principal del proyecto ARSOGAL. El resultado del trabajo de modelado virtual se ha mantenido inédito hasta la fecha, esperando las oportunidades de difusión en foros especializados, incluida la publicación final del proyecto financiado. La restitución completa y en alta definición del conjunto de la *Rúa Galicia* permite visualizar la utópica propuesta de nueva entrada de los peregrinos tramo a tramo, con recorrido y perspectivas que podrán enfocarse a ras de suelo o a vista de pájaro (Fig. 7).

Rematando los trabajos de posproducción, dos fotomontajes completos de la *Rúa Galicia* posibilitan valorar los efectos que su construcción hubieran provocado en la estructura urbana actual de Santiago de Compostela. Estas fotocomposiciones han conllevado un trabajo específico de posproducción dada la complejidad de encajar el modelo 3D en la trama del casco histórico actual en toda su complejidad. La producción gráfica de los renderizados finales y fotocomposiciones, ambas en formato 4K, en PNG y JPG, en 2 tamaños, hasta una resolución máxima de 2500x1406, ha dado como resultado un total de 45 archivos, incluyendo las dos vistas en perspectiva con fotomontajes para las dos orientaciones de inicio y final de la calle.

Ahora bien, la posibilidad de difusión más interesante para todo este proyecto de restitución virtual de una arquitectura de papel reside en la opción de aplicar un motor de animación gráfica tipo UnrealEngine para conseguir una experiencia de recorrido inmersivo y en primera persona. De hecho, esta propuesta se ha hecho llegar al Consorcio de Santiago como producción audiovisual a integrar en una exposición sobre el contexto de la ciudad en la etapa de desarrollo del regionalismo arquitectónico, exposición que provisionalmente se ha programado para inaugurar en la Casa del Cabildo de Santiago de Compostela en la primavera de 2026.

Bibliografía

COSTA BUJÁN, P. (2016), *Periferias y (des)bordes. Evolución urbana y cambios morfológicos, Santiago de Compostela (1778–1950)*, Santiago de Compostela, Consorcio de Santiago / Teófilo Edicións.

GONZÁLEZ GRACIA, E., & ÁNGULO FORNOS, R. (2020), «Revisión del concepto de modelo digital de información. El papel del dibujo y la imagen gráfica», en F. PINTO PUERTO (ed.), *Modelos digitales de información en la tutela sostenible del patrimonio cultural. Contribución al conocimiento e innovación social*, Sevilla, Universidad de Sevilla, pp. 85–97.

MILLIGAN, I. (2022), *The Transformation of Historical Research in the Digital Age*, Cambridge, Cambridge University Press. https://doi.org/10.1017/9781009026055

RODRÍGUEZ ORTEGA, N. (2019), «Digital Art History: The questions that need to be asked», *Visual Resources: An International Journal on Images and Their Uses*, 35(1–2), pp. 6–20. https://doi.org/10.1080/01973762.2019.1553832

SÁNCHEZ GARCÍA, J. A. (ed.) (2020), *Antonio Palacios. Soños de modernidade. Compromiso con Galicia*, Vigo, Museo do Mar de Galicia / Xunta de Galicia.

SÁNCHEZ GARCÍA, J. A. (2021), «Reformas en las "Ciudades de Arte" de España a comienzos del siglo XX. Antonio Palacios y la conservación de una integridad estética en las propuestas para Ourense y Santiago de Compostela», *Storia Urbana*, 168, pp. 77–113. https://doi.org/10.3280/SU2021-168004

SÁNCHEZ GARCÍA, J. A. (2022), «Antonio Palacios y el proyecto para la nueva Rúa Galicia en Santiago de Compostela. Imaginar una ciudad histórica viva y armónica en el contexto del regionalismo», *Quintana*, 21, pp. 1–13. https://doi.org/10.15304/quintana.21.8733

SÁNCHEZ GARCÍA, J. A., VÁZQUEZ CASTRO, J., & VIGO TRASANCOS, A. (eds.) (2019), *Arquitecturas desvanecidas. Memoria gráfica del patrimonio desaparecido en Galicia*, Madrid, Abada Editores.

SÁNCHEZ GARCÍA, J. A., VÁZQUEZ CASTRO, J., & VIGO TRASANCOS, A. (eds.) (2023), *Arquitecturas añoradas: Memoria gráfica del patrimonio destruido en Galicia en el siglo XX*, Gijón, Trea.

Herramientas digitales para la conservación, análisis y activación del patrimonio del diseño en el Arxiu Valencià del Disseny

Arabella León Muñoz
Universitat de València

En los últimos años, las humanidades digitales han proporcionado marcos conceptuales y herramientas tecnológicas capaces de transformar las prácticas tradicionales de conservación, documentación e interpretación del patrimonio cultural. En el contexto de esta transformación, marcada por la digitalización y la expansión de los datos culturales abiertos, la gestión del patrimonio requiere enfoques innovadores que integren accesibilidad, contextualización y análisis relacional. En este marco se sitúa el presente trabajo, vinculado al proyecto de excelencia *PROMETEO* titulado *Arxiu Valencià del Disseny: Shaping the future, rescuing the past*, financiado por la Generalitat Valenciana.

Este proyecto aborda el desarrollo e implementación de un ecosistema digital avanzado para la conservación, catalogación y difusión del *Arxiu Valencià del Disseny (AVD)*, una iniciativa pionera centrada en el patrimonio del diseño gráfico, industrial y visual producido en la Comunitat Valenciana desde mediados del siglo XX hasta nuestros días. El AVD propone un modelo de archivo digital integral, interoperable y semánticamente enriquecido, que responde a los desafíos de un patrimonio caracterizado por su diversidad material, sus vínculos con la cultura visual contemporánea y su escasa formalización archivística.

El proyecto ha sido desarrollado por un equipo multidisciplinar formado por investigadores del Departamento de Historia del Arte y del Instituto de Robótica y Tecnologías de la Información y la Comunicación (IRTIC) de la Universitat de València. Esta colaboración entre especialistas en humanidades y en ingeniería del conocimiento ha permitido diseñar soluciones tecnológicas adaptadas a las necesidades específicas del archivo de diseño. Asimismo, se ha contado con la participación de profesionales del ámbito del diseño, que han asesorado a lo largo de todo el proceso, garantizando la adecuación del modelo a las prácticas y discursos del sector. Esta sinergia entre disciplinas ha sido clave para generar un proyecto sólido, innovador y con proyección internacional.

El origen del propio proyecto es una doble constatación. Por un lado, la creciente visibilización del diseño como disciplina fundamental para comprender los procesos de modernización estética, económica y social del entorno urbano e industrial en España. Por otro, la urgente necesidad de preservar el AVD y estructurar este patrimonio documental extremadamente frágil y heterogéneo. Frente a esta situación, PROMETEO plantea una respuesta desde el campo de

las humanidades digitales, articulando una serie de herramientas tecnológicas que permiten no sólo conservar digitalmente los fondos, sino también enriquecerlos semánticamente, visualizarlos en red y convertirlos en un recurso activo para la investigación, la docencia y la mediación cultural.

El núcleo del proyecto es el diseño y desarrollo de un ecosistema digital semántico que articula cuatro componentes fundamentales. La primera fase del proyecto se centró en la construcción de una ontología propia que permitiera modelar conceptualmente el conocimiento sobre el diseño en el contexto valenciano. Para hacerla lo más interoperable posible se basó en estándares internacionales como CIDOC-CRM. La ontología resultante permite representar entidades como diseñadores/as, empresas, productos, técnicas, soportes, eventos, encargos y relaciones interdisciplinares que conectan los datos. El objetivo principal no es otro que estructurar un grafo de conocimiento que permita identificar no sólo los objetos físicos del archivo, sino también las redes sociales, contextos históricos y dinámicas creativas en que fueron producidos. Además, esta ontología ha sido validada mediante pruebas de modelado sobre casos reales —por ejemplo, el estudio de los trabajos de Paco Bascuñán o la colaboración entre diseñadores gráficos y movimientos sociales en los años ochenta—, lo que ha permitido afinar las clases, propiedades y relaciones necesarias para capturar la especificidad de este patrimonio.

La ontología se apoya en un tesauro multilingüe en tres idiomas (castellano, valenciano e inglés) que se ha desarrollado paralelamente a la catalogación con el objetivo de estandarizar los términos de diseño. Se ha abarcado el máximo de áreas de diseño posibles para cubrir cualquiera de los documentos y productos que se encuentra en el AVD. El desarrollo de la herramienta, que además de servir como vertebración es un recurso más accesible al público, se ha hecho con perspectiva de género y teniendo presente la inclusión respetuosa de los términos. Todo su desarrollo se hizo con Vocbench y está disponible en SKOSMOS.

En una segunda fase se organizó un sistema de gestión documental avanzado fundamentado en un software libre y tecnologías abiertas (Omeka S, Fedora Commons, Wikibase). Así pues, se ha implementado un sistema de gestión documental capaz de integrar metadatos normalizados (Dublin Core, MODS, EAD) y enriquecerlos con vocabularios controlados y descripciones multilingües. El sistema está diseñado para facilitar tanto la catalogación archivística hecha a través de WordPress en una planilla predetermina y consensuada, como la reutilización investigadora, permitiendo a los usuarios navegar por colecciones, acceder a documentación escaneada en alta resolución, y consultar fichas descriptivas interrelacionadas.

Un rasgo distintivo del sistema es su capacidad para registrar procesos y no solo productos: es decir, se documentan los flujos de trabajo, los encargos, los bocetos, las versiones intermedias y los contextos productivos, permitiendo reconstruir la historia material y social del diseño.

Una de las aportaciones clave del proyecto ha sido la creación de un buscador semántico que permite explorar el archivo mediante visualizaciones dinámicas y navegaciones relacionales. A través de grafos de conocimiento interactivos, cronologías contextualizadas y mapas de relaciones, el usuario puede acceder a la dimensión conectiva del diseño. Este visor no solo amplía las formas de acceso al patrimonio, sino que propone una nueva epistemología visual para su interpretación. De este modo, se rompen las jerarquías lineales tradicionales y se favorece una lectura plural, rizomática y situada del archivo. También se conectan documentos con productos y con una herramienta etnográfica que complementa y enriquece las relaciones y la investigación.

Además del trabajo archivístico, PROMETEO ha desarrollado herramientas orientadas al análisis de datos culturales y su reutilización. Entre ellas se incluyen sistemas de georreferenciación de proyectos y diseñadores, y visualizaciones interactivas que permiten detectar patrones históricos o estilísticos a lo largo del tiempo. La referenciación en mapas espaciotemporales resulta muy interesante para trazar las relaciones entre diseño y producción, por ejemplo. Esta herramienta ha sido aplicada en proyectos de investigación y docencia en colaboración con universidades, escuelas de diseño y museos, generando líneas de trabajo que abordan el archivo como laboratorio vivo, susceptible de activar nuevas preguntas historiográficas y nuevos relatos visuales.

A lo largo del desarrollo del proyecto se han integrado más de 30 colecciones documentales, entre las que destacan los fondos de diseñadores como Daniel Nebot, Pepe Gimeno, Marisa Gallén, Juan Nava o Carmen Baselga. Cada una de estas colecciones ha permitido poner a prueba las herramientas desarrolladas, validando la estructura semántica, el sistema documental y el visor relacional. Por ejemplo, el tratamiento digital del fondo Paco Bascuñán ha permitido reconstruir las redes de colaboración con el mundo teatral y político, visualizar la evolución de su lenguaje visual en relación con los movimientos sociales de la transición democrática, y trazar vínculos con diseñadores contemporáneos mediante análisis de coautoría y estilos compartidos.

Además, se han desarrollado experiencias de mediación digital en colaboración con otros proyectos como DHECA o Cheminova. Estas experiencias han demostrado el potencial del archivo como recurso pedagógico e instrumento para la democratización del acceso al patrimonio.

Más allá del desarrollo tecnológico, nuestro proyecto PROMETEO plantea una reflexión crítica sobre el papel de los archivos digitales en el siglo XXI. Frente al riesgo de digitalizar sin transformar, el proyecto propone una redefinición del archivo como infraestructura abierta, relacional e inteligente. En este contexto, “inteligencia” no remite a procesos automatizados o algorítmicos, sino a la capacidad del archivo para generar conocimiento situado, contextualizado, múltiple y reutilizable. Las búsquedas de nuestra web semántica utilizan un lenguaje natural para intentar garantizar la eficiencia de los datos obtenidos y la eficacia de los resultados.

Además, todo el proyecto se ha desarrollado con un enfoque de ciencia abierta, la reutilización de datos culturales y los objetivos de desarrollo sostenible (ODS), especialmente en lo relativo al acceso inclusivo a la cultura, la preservación de la memoria colectiva y la innovación educativa están presentes en todo su desarrollo. Sin duda, PROMETEO contribuye así a redefinir la noción de archivo desde una perspectiva contemporánea, situada en la intersección entre tecnología, arte y sociedad.

En conclusión, el proyecto PROMETEO del Arxiu Valencià del Disseny constituye una contribución significativa a la línea de conservación y digitalización del patrimonio cultural. A través de una aproximación transdisciplinar, el proyecto ha desarrollado herramientas que permiten no sólo conservar y estructurar un patrimonio cultural específico, sino activarlo como campo de investigación, espacio de creación, así como dispositivo de investigación y pedagógico. En un momento en que los archivos culturales se enfrentan al desafío de su integración en ecosistemas digitales sostenibles, transparentes y críticos, PROMETEO ofrece un modelo replicable y adaptable que combina excelencia técnica, rigor académico y compromiso social. Este trabajo se presenta como una invitación a repensar el archivo no como un lugar de almacenamiento, sino como un espacio relacional, abierto y transformador.

Bibliografía

BACA, M. (ed.) (2016), *Introduction to Metadata*, Los Angeles, Getty Research Institute. http://www.getty.edu/publications/intrometadata/

BERMAN, J., & BRADY, M. (2021), *Humanities Data: A Guide to Its Logical Structure and Interpretation*, MIT Press.

DOERR, M. (2003), «The CIDOC Conceptual Reference Model: An ontological approach to semantic interoperability of metadata», *AI Magazine*, 24(3), pp. 75–92. https://doi.org/10.1609/aimag.v24i3.1720

FLANDERS, J., & JANNIDIS, F. (2019), «The shape of data in Digital Humanities: Modeling texts and text-based resources», *Digital Scholarship in the Humanities*, 34(suppl_1), pp. i1–i9. https://doi.org/10.1093/llc/fqz060

GALLEGO, M. T., & HERNÁNDEZ, L. (2021), «Archivos de diseño: prácticas, políticas y retos contemporáneos», en M. T. GALLEGO (ed.), *El diseño como patrimonio. Prácticas archivísticas y narrativas visuales*, Valencia, Universitat de València, pp. 23–48.

MANOVICH, L. (2020), *Cultural Analytics*, Cambridge (MA), MIT Press.

MARTÍNEZ, S. (2018), «Representar el conocimiento en las Humanidades Digitales: Ontologías, modelos y taxonomías», *Revista de Humanidades Digitales*, 3, pp. 59–86. https://doi.org/10.5944/rhd.vol.3.2018.21562

MOREIRO, J. A. (2017), *Gestión de la información y del conocimiento en archivos y museos digitales*, Madrid, Síntesis.

PRESNER, T., SHEPARD, D., & KAWANO, Y. (2014), *HyperCities: Thick Mapping in the Digital Humanities*, Cambridge (MA), Harvard University Press.

ZINN, K. (2022), «Open access and design archives: A roadmap for sustainable digital infrastructures», *Design Issues*, 38(2), pp. 80–92. https://doi.org/10.1162/desi_a_00698

Cartografías entrelazadas: modelar la transnacionalidad en el ecosistema expositivo español (2000-2023) mediante análisis de redes

Nuria Rodríguez-Ortega
iArtHis_Lab – Universidad de Málaga
Bárbara Romero-Ferrón
Provenance Lab – Leuphana University;
iArtHisLab – Universidad de Málaga

Palabras clave: Transnacionalidad cultural; exposiciones artísticas; análisis de redes; sistemas culturales complejos
Keywords: Cultural transnationality; art exhibitions; network analysis; complex cultural systems

Este trabajo presenta una aproximación metodológica para estudiar la transnacionalidad en los centros de arte contemporáneo españoles entre 2000 y 2023. Partiendo de un corpus extraído de Expofinder y del grafo de conocimiento Complexhibit, se aplican técnicas de teoría de grafos y análisis de redes para cartografiar los flujos de artistas e instituciones y, con ello, contribuir al cuestionamiento de las fronteras geopolíticas como principio de clasificación cultural, de acuerdo con la propuesta que planteara Thomas da Costa Kaufmann en su obra seminal Toward a Geography of Art (2004). El estudio revela (1) la existencia de un núcleo densamente interconectado de museos e instituciones españolas que actúan como hubs de internacionalización, (2) patrones asimétricos de circulación que privilegian determinadas nacionalidades y (3) una dialéctica global-local que produce geografías culturales diferenciadas dentro del propio territorio. Estos hallazgos contribuyen a una comprensión más compleja de las lógicas expositivas y proporcionan un marco replicable para otros contextos.

Introducción: contexto, objetivos y preguntas de investigación

En las dos últimas décadas la noción de transnationality se ha consolidado como categoría analítica para describir flujos culturales que atraviesan, pero no necesariamente suprimen, las fronteras nacionales. Frente a discursos celebratorios de la «globalización», el giro transnacional propone interrogar

simultáneamente las persistencias del Estado-nación y las formas de circulación que lo desbordan (Goyal, 2017). El dominio de las exposiciones artísticas, con sus circuitos y desplazamientos, constituye un campo privilegiado para observar esta tensión. Así lo han mostrado estudios previos en el ámbito de los exhibition studies, subrayando el papel de las exposiciones, bienales y ferias como infraestructuras de movilidad simbólica (O'Neill, 2012; Jones, 2017; Atkinson et al., 2022). Complementariamente, proyectos como DoME para Europa (1905-1915)[1] y Artl@s a escala global[2] —entre otros— han demostrado el potencial de la analítica de datos para producir conocimiento significativo en este campo.

En este marco de investigación, los proyectos Exhibitium (2015-2016)[3] y Andalex (2017-2021)[4], evolucionados hoy en el proyecto Complexhibit (2021-)[5], se han centrado en el caso español. La primigenia base de datos relacional Expofinder[6] se ha transformado en un grafo de conocimiento basado en la ontología OntoExhibit[7] —alineada parcialmente con CIDOC-CRM— interrogable mediante un endpoint SPARQL que facilita la recuperación de información en formato RDF (Rodríguez Ortega, et al. 2024, Rodríguez Ortega, 2024). Esta infraestructura, que cuenta con un volumen de 27.699 exposiciones celebradas en España, constituye el fundamento empírico del presente estudio.

La singular configuración del panorama expositivo español —caracterizado, por un lado, por una red densa de museos y centros de arte públicos de escala estatal y autonómica, y, por otro, por la irrupción paulatina de fundaciones corporativas que diversifican la gestión cultural— constituye un laboratorio idóneo para interrogar las dinámicas de transnacionalidad. El reto central radica en comprender cómo se erige un ecosistema transnacional dentro de unas fronteras geopolíticas claramente delimitadas —el Estado español—, que, a su vez, se fragmenta en contextos autonómicos con identidades culturales y políticas propias. Sobre esta premisa, las preguntas de investigación se orientan a desentrañar la lógica con la que actores, flujos y territorios articulan, modulan o tensionan la construcción de una trama transnacional en el sistema expositivo del arte.

Por su parte, adoptar la transnacionalidad como objeto de estudios implica ir más allá del concepto de «internacionalización», orientado a describir y cuantificar la presencia de lo foráneo dentro de un contexto nacional o a la

[1] DoME: Database of Modern Exhibitions. Accesible en: https://exhibitions.univie.ac.at
[2] Artl@s Project. Accesible en: https://artlas.huma-num.fr/map/#/
[3] Proyecto Exhibitium. Generación de conocimiento sobre exposiciones artísticas temporales para su reutilización y aprovechamiento multivalente (Fundación BBVA). https://exhibitium.es/
[4] Proyecto Andalex. Ecosistema de datos abiertos y enlazados del subsector cultural de las exposiciones artísticas: formalización ontológica, soluciones tecnológicas y modelos de explotación para la generación de conocimiento y valor en el ámbito de las ICC (Junta de Andalucía). https://andalexproject.iarthislab.eu/
[5] Proyecto Complexhibit. Estudio e interpretación del dominio de las exposiciones artísticas como sistema cultural complejo mediante analítica de datos y procesamiento del lenguaje natural. PID2021-125037NB-I00 (Ministerio de Ciencia e Innovación). https://github.com/orgs/Complexhibit-Project/repositories
[6] https://expofinder.iarthislab.eu/web/
[7] https://complexhibit-project.github.io/OntoExhibit/index-en.html

inversa, la presencia de artistas e instituciones españoles en contextos extranjeros. La transnacionalidad, por el contrario, nos permite poner el foco en la estructura relacional que posibilita y reconfigura esa presencia: no basta saber que un museo «importa» un número determinado de artistas; queremos mostrar cómo se tejen flujos multilaterales, cómo operan capas superpuestas de identidad y de qué manera las exposiciones redibujan la geografía cultural tradicional (Rodríguez Ortega, 2018).

Las preguntas de investigación de nuestro estudio se articulan en torno a seis ejes: (1) evaluar la internacionalización efectiva del sistema expositivo español (2000-2023) a través de la presencia y centralidad de agentes extranjeros; (2) definir indicadores de transnacionalidad basados en grafos (E-I, intermediación, modularidad); (3) caracterizar los roles de museos estatales, centros autonómicos y fundaciones corporativas como mediadores de flujos transnacionales; (4) analizar cómo se superponen dinámicas globales y redes autonómicas y detectar «micro-plataformas» locales; (5) identificar morfologías alternativas (clústeres, puentes, periferias conectadas) que desafían la lógica centro-periferia; y (6) mapear las conexiones y asimetrías inter-CCAA y su implicación en cooperación, competencia y dependencia.

Metodología

Fases de trabajo

La investigación combina tres capas de trabajo:

- Extracción y normalización de datos. Para este estudio se han filtrado eventos celebrados entre enero de 2000 y diciembre de 2023 en 570 museos o centros de arte contemporáneo de titularidad estatal, autonómica o fundacional (privada o público-privada). Se han considerado las siguientes dimensiones: actores (rol artista / rol comisario), instituciones (rol organizador), lugar de celebración y fechas.
- Análisis de estadística descriptiva: Se han realizado análisis de estadística descriptiva para conocer la estructura general del ecosistema: volumen de artistas extranjeros participantes en términos globales y desagregados por sexo y edad; nacionalidades más representadas en términos globales y desagregadas por institución, comunidad autónoma y sexo; y análisis dinámico de las distintas variables para conocer le evolución temporal.
- Modelado de la red y construcción del grafo: Se ha trabajado con un grafo bipartito (A agentes: artistas, comisarios) y B (instituciones) con proyecciones derivadas basadas en redes de coaparición en exposiciones. Se descartaron vínculos con frecuencia < 2 para reducir ruido y se ponderaron los restantes por número de colaboraciones.

Métricas fundamentales utilizadas en el análisis

Indicador	Descripción
Centralidad de grado	Identifica los hubs de visibilidad dentro de la red.
Grado de intermediación	Señala los nodos que actúan como puentes entre comunidades, facilitando la conectividad entre grupos diversos.
Modularidad multinacional (Louvain γ = 1)	Detecta comunidades mixtas de agentes e instituciones en función de sus vínculos, revelando la estructura modular del sistema transnacional.
Índice E-I (External-Internal)	Mide la apertura transnacional mediante el balance entre enlaces heteronacionales (E) y homonacionales (I). El índice varía de -1 (bloques mononacionales) a +1 (alta heterofilia).
Diversidad de Simpson + Entropía de Shannon	Cuantifican la riqueza y el equilibrio de nacionalidades en la red local (egonetwork) de cada institución. Permiten evaluar si las instituciones o exposiciones actúan como vitrinas multiculturalmente diversas o si concentran la atención en un único foco geopolítico.
Motif analysis	Estudia configuraciones microestructurales, como triángulos heteronacionales o cuadrángulos puente, que revelan patrones de integración transnacional y la formación de micro-configuraciones que desafían la lógica de bloques geopolíticos clásicos.

En conjunto, esta batería de métricas —centradas en visibilidad, mediación, cohesión modular, apertura, riqueza diversificada y estructuras micro-relacionales— proporciona un retrato multidimensional de cómo las instituciones españolas integran la presencia y circulación de artistas extranjeros en el ecosistema expositivo.

Todo el procesamiento se ha efectuado con Python/NetworkX y las visualizaciones finales se han realizado con Gephi (layout ForceAtlas2). Una vez se muestren y publiquen los resultados, los códigos y datasets estarán disponibles en el GitHub del proyecto Complexhibit (doi en trámite) bajo licencia CC-BY-SA.

Corpus y sesgos

La muestra contiene 15.633 exposiciones, 13.720 artistas procedentes de 141 nacionalidades y 570 instituciones. Es importante considerar que, debido a la naturaleza de nuestro proyecto y a la forma en que se han ido recopilando los datos —a través de distintos paquetes de trabajo, con objetivos y fuentes de financiación diversas— no disponemos de la misma información para todas las instituciones analizadas. Esta heterogeneidad puede generar cierto desbalance en el corpus. Asimismo, es fundamental tener en cuenta la idiosincrasia de las entidades incluidas: instituciones como el Museo del Prado o los museos de

Bellas Artes, con una orientación más decimonónica, tienden a exhibir artistas italianos, franceses, ingleses, flamencos y españoles, mayoritariamente de los periodos renacentista, barroco o del siglo XIX. Su inclusión, sin embargo, nos permite entender cómo conviven lógicas curatoriales históricas y contemporáneas y, así, analizar las tensiones entre tradición y modernidad en el panorama expositivo.

Resultados

Con el fin de ajustarnos a la limitación de extensión, presentamos a continuación un breve resumen de los resultados obtenidos, los cuales serán desarrollados y debatidos en profundidad durante la presentación en el congreso.

Desde un punto de vista estadístico, los datos revelan que el ecosistema expositivo español se configura como un entorno abierto y plural, en el que la presencia de artistas extranjeros supera a la de nacionales. Solo el 46 % de los artistas representados son españoles, mientras que el resto proviene de una amplia gama de países, sin que ninguna nacionalidad extranjera resulte hegemónica. Estados Unidos (7 %), Francia (6 %) e Italia (4 %) encabezan la lista, pero con porcentajes relativamente bajos y equilibrados entre sí, lo que sugiere una red cultural que, lejos de estar dominada por una lógica «doméstica» u orientada hacia contextos geopolíticos dominantes, opera bajo dinámicas plurales. Esta distribución descentralizada apunta hacia una estructura policéntrica del sistema, en la que múltiples focos de origen coexisten sin que uno se imponga sobre los demás.

Desde una perspectiva interseccional, el cruce entre género y nacionalidad añade una capa significativa al análisis (fig. 1). Las artistas extranjeras más presentes son estadounidenses, seguidas de alemanas y portuguesas, lo que indica una mayor apertura hacia la internacionalización femenina en ciertos contextos culturales. Sin embargo, casos como el de Italia, donde la proporción de mujeres es particularmente baja respecto al conjunto de artistas italianos, evidencian la persistencia de desigualdades de género vinculadas a marcos nacionales específicos.

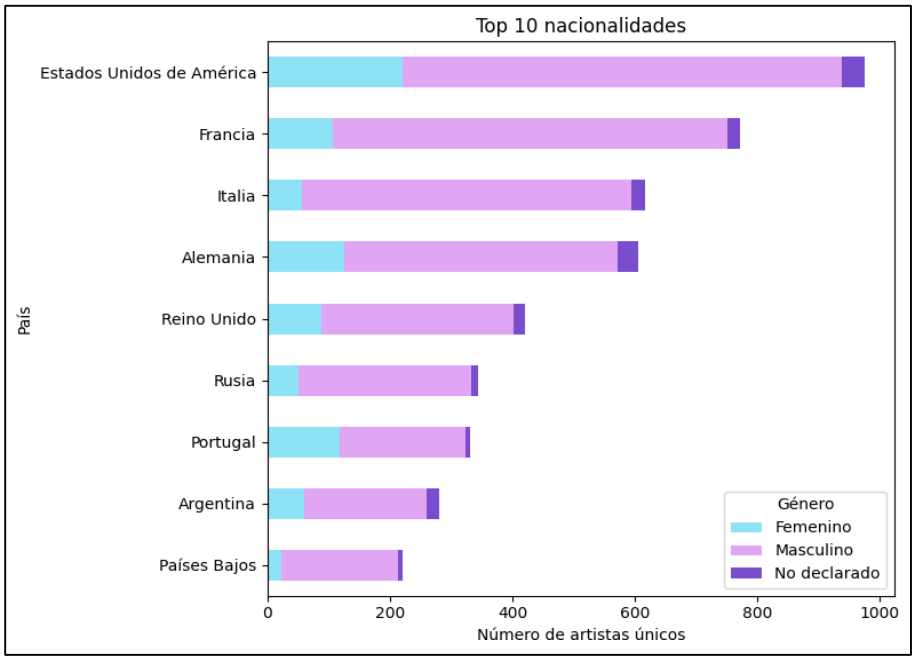


Figura 1. Distribución de nacionalidades por género. Fuente: Expofinder-OntoExhibit

A nivel temporal, la evolución de estas dinámicas refleja tanto continuidades como rupturas (fig. 2): la estabilidad británica contrasta con el ascenso portugués a partir de 2015, interrumpido abruptamente por la pandemia de la COVID19. Asimismo, se aprecia que la presencia de artistas estadounidenses ha sido significativa desde los inicios del siglo XXI, aunque con fluctuaciones notables.

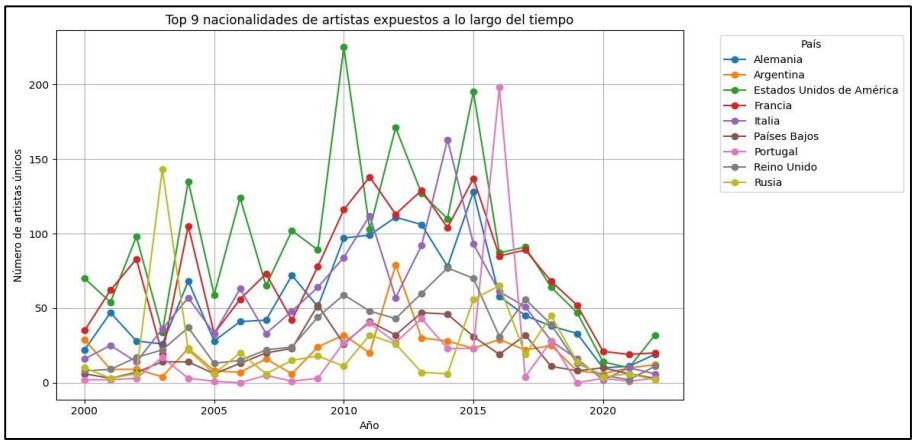


Figura 2. Análisis dinámico de la presencia de nacionalidades en el ecosistema expositivo español. Fuente: Expofinder- OntoExhibit

El análisis de la estructura global del ecosistema expositivo español, modelada como un grafo compuesto por 15.757 nodos y 693.169 aristas, revela una distribución de grado que se aproxima a una ley de potencias con un exponente $\alpha \approx 4.5$. No obstante, esta tendencia se verifica únicamente en un número reducido de nodos altamente conectados, mientras que la mayoría se desvía de dicho patrón. La comparación con una distribución exponencial no ofrece resultados estadísticamente significativos ($p = 0.16$), lo que debilita la

hipótesis de una red propiamente de tipo escala libre. En su lugar, nos encontramos ante una estructura híbrida, situada entre una red aleatoria y una red jerárquica, que refleja una cierta resistencia a la centralización total del sistema. Esta configuración puede estar relacionada con el peso específico que ejercen las comunidades autónomas como polos de visibilidad relativamente independientes. Todo ello sugiere la existencia de un sistema semiegalitario con focos de concentración parcial, capaz de sostener dinámicas de circulación transnacional sin depender exclusivamente de unos pocos actores dominantes. Se favorecen, así, configuraciones multicéntricas que facilitan la inclusión de agentes periféricos o emergentes (Fig.3).

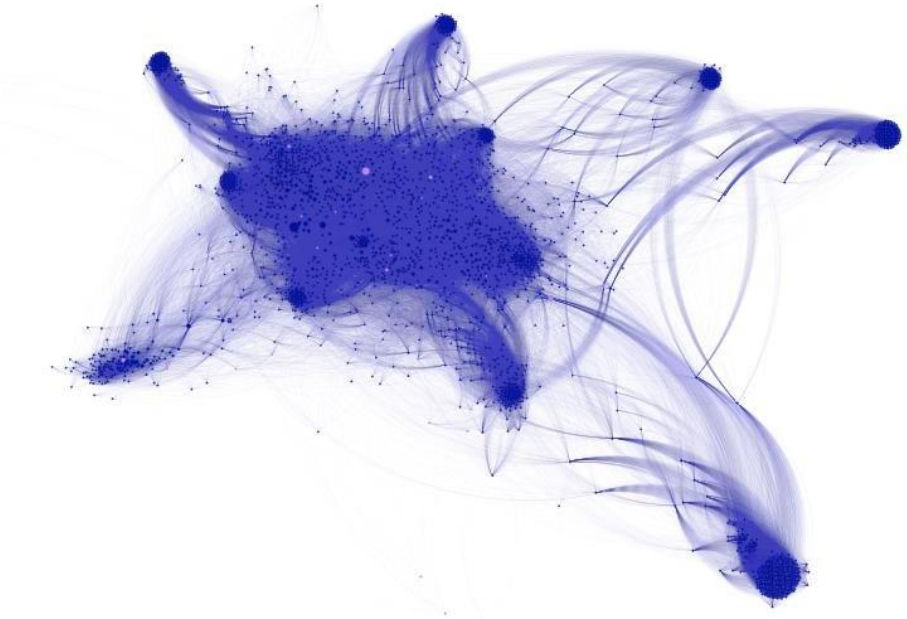


Figura 3. Red de coaparición en exposiciones. El tamaño de los nodos corresponde al grado ponderado. La red está coloreada en azul oscuro los actores (aristas y comisarios) y en lila las instituciones culturales. Fuente: Expofinder- OntoExhibit

De hecho, el análisis de motivos (Milo et al., 2002) aplicado a las capas territoriales revela una arquitectura en red de tipo multinivel. En este contexto, el índice β[8] (Boccaletti et al., 2006) —que mide la densidad relativa de conexiones entre instituciones locales y artistas internacionales— sugiere que, junto a los grandes centros que concentran y articulan los flujos globales (Madrid, Cataluña o el País Vasco), están emergiendo micro-plataformas periféricas en ciudades de tamaño medio, como León a través del MUSAC (fig. 4). Aunque más pequeñas en volumen, estas plataformas desempeñan un papel estratégico al acoger artistas de diversas procedencias; en particular, de contextos específicos apenas representados en el resto del sistema, ampliando, así, la geografía transnacional y la especialización temática del ecosistema expositivo español.

[8] El índice β se basa en métricas de conectividad interescalares utilizadas en análisis de redes complejas (cf. Boccaletti et al., 2006).

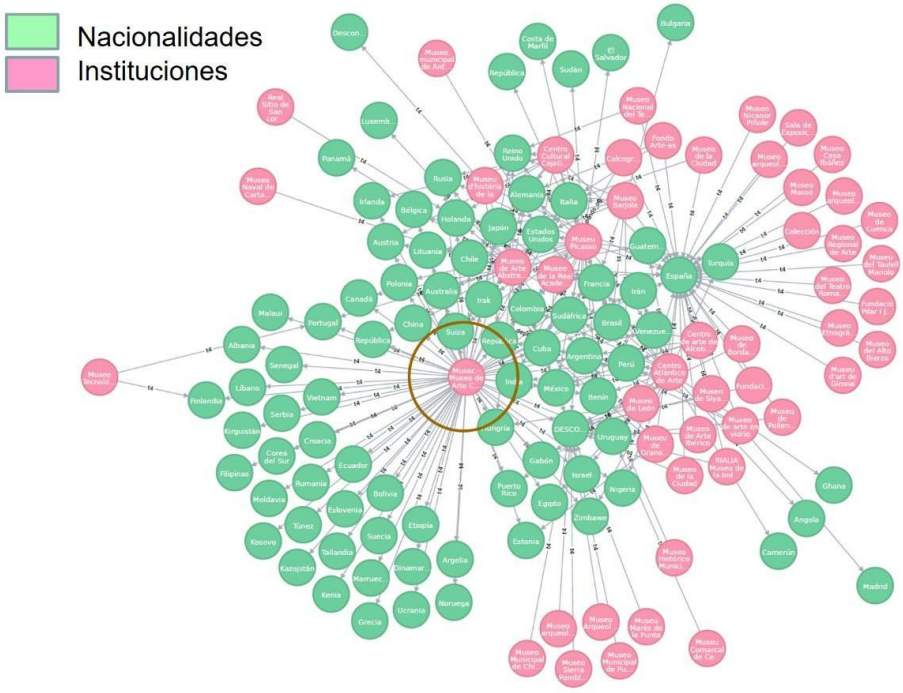


Figura 4. Red instituciones-nacionalidades. Fuente: Expofinder-OntoExhibit

En esta línea, los resultados dialogan con estudios sobre redes artísticas globales que han señalado la persistencia de desigualdades estructurales en el acceso a la visibilidad internacional (Quemin, 2013; Velthuis & Baia Curioni, 2015), pero también con investigaciones que destacan la emergencia de nodos intermedios y escenas periféricas capaces de alterar parcialmente dicha lógica (Perczel & Vedres 2025).

El análisis de modularidad, basado en la coaparición de nacionalidades en exposiciones, revela que la transnacionalidad en el ecosistema expositivo español no se manifiesta de forma unificada, sino que se estructura en al menos tres comunidades diferenciadas. La primera, compuesta por países de Europa occidental, del norte y Estados Unidos, presenta baja diversidad nacional y altos niveles de coexposición, lo que refuerza un núcleo hegemónico y relativamente cerrado en torno al canon artístico dominante. La segunda comunidad agrupa principalmente países del Sur Global—junto con algunas naciones de Europa oriental— y se caracteriza por una mayor apertura y heterogeneidad en sus patrones de coaparición, lo que sugiere una lógica curatorial más inclusiva hacia perspectivas periféricas. La tercera comunidad conecta países europeos no centrales, especialmente del Este y del Norte, donde las coexposiciones se dan de forma más localizada y con escasa vinculación al Sur Global. Estos patrones indican que la transnacionalidad no conforma una red integrada, sino un sistema segmentado en bloques con grados variables de visibilidad, centralidad y reconocimiento, atravesados por jerarquías geopolíticas.

A nivel institucional, los datos sugieren tres tipologías:

- Global brokers (E-I > 0.5, alta intermediación) —principalmente museos estatales—.

- Nodos locales reforzados (E-I ≈ 0, alto grado interno) —centros autonómicos—.
- Plataformas periféricas (E-I negativo, pero conectadas a brokers mediante exposiciones itinerantes).

Como era de prever, los museos estatales tienden a aparecer como global brokers con betweenness elevada y alta diversidad de Simpson; los centros autonómicos, en cambio, exhiben una densidad interna fuerte pero menor capacidad de puente, actuando a modo de anclas regionales. Las fundaciones corporativas ocupan posiciones intermedias: menos centralidad de grado que los grandes museos, pero mayor E-I que muchos centros autonómicos, evidenciando su rol de pasarela hacia circuitos de patrocinio internacional.

En resumen, el estudio demuestra que en el ecosistema expositivo español la transnacionalidad no es homogénea sino estratificada. Se observa un mosaico de transnacionalidades donde la tradicional posición geográfica centro-periferia coincide solo parcialmente con la función cultural que cada institución desempeña en la red como dinamizador de dinámicas transnacionales. Este patrón confirma la tesis de que la transnacionalidad genera nuevas fronteras simbólicas dentro del territorio nacional, redefiniendo el «adentro» y el «afuera» en función del capital relacional, no de líneas políticas (Goyal, 2017).

Limitaciones y líneas futuras

A pesar de su enfoque relacional sólido, este estudio presenta ciertas limitaciones derivadas de la cobertura desigual del corpus, centrado principalmente en instituciones de gran escala. La ausencia de indicadores cualitativos —como los presupuestos de producción, el perfil del público o los criterios curatoriales— limita la comprensión de las dinámicas internas de las exposiciones. Asimismo, el uso de snapshots anuales impide captar variaciones a corto plazo o responder con precisión a eventos disruptivos como la pandemia de la COVID-19. Con el objetivo de afinar este diagnóstico, futuras investigaciones ampliarán el universo analizado para incluir galerías medianas y pequeñas, así como espacios independientes con menor visibilidad institucional. Se incorporarán, además, datos económicos y de audiencia que permitan caracterizar con mayor precisión las lógicas de funcionamiento del sistema expositivo. También se mejorará la resolución temporal del análisis, lo que facilitará la detección de choques exógenos y la evaluación de su impacto en la programación. Finalmente, se desarrollará un análisis de contenido basado en técnicas de topic modeling y clasificación manual orientado a identificar los ejes discursivos más presentes en las exposiciones —como género, poscolonialismo o ecología— y así contrastar la transnacionalidad de origen de los artistas con la transdiscursividad que promueven las instituciones.

Bibliografía

ATKINSON, H., CLARKSON, V., & LICHTMAN, A. (eds.) (2022), *Exhibitions Beyond Boundaries: Transnational Exchanges through Art, Architecture, and Design 1945–1985*, New York, Bloomsbury Visual Arts.

BOCCALETTI, S., LATORA, V., MORENO, Y., CHAVEZ, M., & HWANG, D.-U. (2006), «Complex networks: Structure and dynamics», *Physics Reports*, 424(4–5), pp. 175–308. https://doi.org/10.1016/j.physrep.2005.10.009

GOYAL, Y. (2017), «Introduction: The transnational turn», en *The Cambridge Companion to Transnational American Literature*, Cambridge, Cambridge University Press, pp. 1–16.

JONES, C. A. (2017), *The Global Work of Art: World's Fairs, Biennials, and the Aesthetics of Experience*, Chicago, University of Chicago Press.

KAUFMANN, T. D. (2004), *Toward a Geography of Art*, Chicago, University of Chicago Press.

MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D., & ALON, U. (2002), «Network motifs: Simple building blocks of complex networks», *Science*, 298(5594), pp. 824–827. https://doi.org/10.1126/science.298.5594.824

O'NEILL, P. (2012), «Biennial culture and the emergence of a globalized curatorial discourse», en *The Culture of Curating and the Curating of Culture(s)*, Cambridge (MA), MIT Press, pp. 29–48.

PERCZEL, J., & VEDRES, B. (2025), «Careers in the global art field: Geo-capital and globalizer venues in the consecration of Central-Eastern European artists», *Poetics*, 118. https://doi.org/10.1016/j.poetic.2024.101961

QUEMIN, A. (2013), «International contemporary art fairs in a 'globalized' art market», *European Societies*, 15(2), pp. 162–177. https://doi.org/10.1080/14616696.2013.767927

RODRÍGUEZ ORTEGA, N. (2018), «Nuevas geografías culturales: cartografías construidas mediante algoritmos», en E. ORTEGA ARJONILLA (coord.), *El paisaje. Percepciones interdisciplinares desde las Humanidades*, Granada, Editorial Comares, pp. 29–42.

RODRÍGUEZ-ORTEGA, N. (2024), «Contours of Knowledge: Epistemological Implications of Semantic Models in the Representation of the Art Exhibition Domain through the Lens of the OntoExhibit Ontology», *Život umjetnosti*, 114(1), pp. 122–147. https://doi.org/10.31664/zu.2024.114.06

RODRÍGUEZ ORTEGA, N., ROLDÁN GARCÍA, M.ª M., PLATAS DÍEZ, M.ª L., & SALVACHÚA, M. (2024), «OntoExhibit: An ontology for modeling the field of art exhibitions and its semantic-discursive dimension», en J. KARAJGIKAR, A. JANCO, & J. OTIS (eds.), *Digital Humanities 2024: Book of Abstracts*, VA 2024. https://doi.org/10.5281/zenodo.13761079

VELTHUIS, O., & BAIA CURIONI, S. (eds.) (2015), *Cosmopolitan Canvases: The Globalization of Markets for Contemporary Art*, Oxford, Oxford University Press.

Desafíos y oportunidades en la preservación digital de las colecciones textiles de seda en museos europeos

Mar Gaitán
Universitat de València

Palabras clave: Museos; tejidos; tesauro; digitalización; web semántica
Keywords: Museums; textiles; thesaurus; digitization; semantic web

Introducción

Los museos europeos desempeñan un papel crucial en la conservación del patrimonio cultural, tanto tangible como intangible. Dentro de sus colecciones, los textiles —y en particular los tejidos de seda— representan una categoría especialmente delicada y valiosa, no sólo por su fragilidad material, sino también por la riqueza histórica y técnica que encapsulan. A pesar de ello, la mayoría de las instituciones que albergan este tipo de patrimonio son de tamaño pequeño o mediano, y enfrentan importantes limitaciones estructurales, económicas y tecnológicas que obstaculizan su adaptación al entorno digital contemporáneo.

Actualmente, existe una creciente producción de datos digitales relacionados con estas colecciones, pero esta se caracteriza por su heterogeneidad, dispersión y falta de estandarización. Desde repositorios locales elaborados en Excel o Word hasta sistemas más complejos con APIs, la gestión de la información digital varía ampliamente entre instituciones. Esta diversidad no sólo refleja la disparidad de recursos humanos y técnicos, sino también una brecha significativa en competencias digitales dentro del sector patrimonial.

La documentación, la catalogación y el acceso a las colecciones textiles en formato digital siguen siendo procesos desiguales, con pocos esfuerzos coordinados o políticas comunes. El riesgo es doble: por un lado, perder información relevante sobre objetos únicos y frágiles; por otro, desaprovechar la oportunidad de reutilizar estos datos para la investigación, la educación o la innovación cultural. Además, la conservación del patrimonio inmaterial asociado (como las técnicas de tejido o los saberes artesanales) requiere metodologías capaces de ir más allá de la simple digitalización de objetos físicos.

Este estudio se propone analizar de manera crítica y descriptiva cómo los museos europeos con colecciones de seda están afrontando estos desafíos digitales. En concreto, busca comprender:

- ¿Qué tecnologías están utilizando?
- ¿Qué barreras encuentran en términos de documentación y gestión digital?

- ¿Cómo se accede y reutiliza la información disponible?
- ¿Qué estrategias se aplican —si existen— para conectar el patrimonio tangible con el intangible?

El objetivo final es aportar recomendaciones basadas en la evidencia que contribuyan al fortalecimiento digital de estas instituciones, permitiéndoles conservar mejor su acervo y hacerlo más accesible y significativo para la sociedad.

Metodología

Para comprender cómo los museos europeos que gestionan colecciones de seda están utilizando la tecnología en sus procesos de conservación, documentación y acceso, se adoptó una metodología cualitativa mediante el uso del método de encuesta. El instrumento de recolección de datos consistió en un cuestionario electrónico estructurado en tres bloques temáticos y distribuido a través de una plataforma en línea. Se empleó una muestra de conveniencia no probabilística, enfocada en profesionales que trabajan directamente con colecciones textiles: conservadores, curadores y documentalistas.

La encuesta fue enviada por correo electrónico a museos europeos con colecciones relevantes de seda, y se recibieron un total de 42 respuestas completas. El cuestionario incluyó un total de 30 preguntas, abiertas y cerradas, lo que permitió triangular técnicas cuantitativas y cualitativas de análisis. Esta combinación metodológica buscó garantizar tanto la estandarización de ciertos ítems como la posibilidad de recoger perspectivas más amplias sobre las prácticas institucionales.

La encuesta se organizó en tres bloques principales:

a) Datos institucionales: tipo de museo, tamaño, personal involucrado en la digitalización, disponibilidad de presupuesto, etc.
b) Procesos de documentación y catalogación: sistemas utilizados, formatos de archivo, nivel de estandarización, políticas internas, etc.
c) Acceso abierto y gestión de datos digitales: plataformas públicas, interoperabilidad, condiciones de reutilización, uso de tecnologías emergentes (como modelos 3D, realidad aumentada, etc.).

Una vez recopilados los datos, se procedió a la codificación de las respuestas y a la elaboración de un análisis descriptivo de frecuencias relativas. Este análisis se realizó utilizando el software Microsoft Excel, permitiendo identificar tendencias generales, casos particulares y puntos críticos recurrentes entre las instituciones encuestadas. Además, se incorporó una lectura cualitativa de las respuestas abiertas para identificar discursos, expectativas y desafíos no visibles en los datos cuantificados.

Resultados obtenidos

Los resultados del estudio confirman que los museos europeos con colecciones de seda presentan grandes disparidades en cuanto a recursos digitales, capacidades técnicas y estrategias de conservación digital.

- De los museos analizados, el mayor número de textiles de seda se encuentra en los museos de artes decorativas, que representan el 21 % del total, seguidos de los museos de tejidos, con un 15 %. Las terceras categorías más representadas son los museos de moda y trajes y los museos diocesanos, con un 13 % respectivamente. Sin embargo, de los museos que participaron en la encuesta, solo el 35,7 % procedían de museos textiles.
- Todas las instituciones encuestadas afirmaron que es importante estandarizar el vocabulario utilizado en la terminología textil; el 71,4 % utiliza tesauros, el 66 % de los cuales se crearon específicamente para cada una de las colecciones, lo que dificulta la interoperabilidad y la estandarización de los datos entre las distintas instituciones. Es cierto que hay una minoría de museos dedicados específicamente a los textiles, de los cuales la gran mayoría se basa al menos en CIETA, pero la falta de un vocabulario estándar abierto en formato SKOS dificulta las tareas de catalogación y, por lo tanto, de conservación y difusión.
- La mayoría de los encuestados digitalizó su información; de hecho, 41 de las 42 instituciones creen que es importante compartir su colección; sin embargo, solo el 35,7 % de los encuestados estaba familiarizado con la web semántica y las ontologías. En este sentido, el 80 % cree que la interoperabilidad es importante, lo que requiere una terminología estandarizada.
- Todos los encuestados consideraron que la tecnología es útil para la conservación, difusión y protección del patrimonio de la seda. En este sentido, 29 instituciones estarían interesadas en recibir una mejor formación en gestión de datos digitales y en el intercambio de sus colecciones, lo que mejoraría su aprendizaje y también su difusión. Esto demuestra el interés de los museos por utilizar diversos aspectos tecnológicos en los servicios de sus colecciones. En este sentido, las humanidades digitales y la historia del arte pueden ser de especial utilidad.

Conclusiones y perspectivas

Los museos europeos que conservan colecciones de seda enfrentan un panorama complejo y fragmentado respecto a su digitalización. Aunque son conscientes del potencial de las nuevas tecnologías para enriquecer la conservación, el acceso y la valorización de su patrimonio, sus posibilidades reales se ven limitadas por falta de formación técnica, carencia de personal especializado y escasos recursos económicos. La digitalización del patrimonio

de la seda no debe entenderse sólo como una cuestión técnica, sino como un proceso profundamente cultural y social. Requiere diálogo entre disciplinas, colaboración entre instituciones y una visión integradora que conecte el pasado con los desafíos del futuro. Esta propuesta busca ser una contribución significativa a ese proceso.

Se demostrará las diferentes intenciones que tienen las instituciones culturales a la hora de aplicar la tecnología, ya sea con fines de conservación o para gestionar la información de los usuarios.

Este estudio ofrece una visión de los retos digitales a los que se enfrentan los museos en el siglo XXI. Hacer accesibles las colecciones permite descubrir nuevas conexiones, mejorar y llevar a cabo nuevas investigaciones en Historia del Arte, y también establecer conexiones con los usuarios y el territorio, redescubriendo el estrecho vínculo entre el objeto textil y el territorio para su mejora mutua.

Este trabajo se enmarca en el cruce entre las ciencias sociales, las humanidades digitales y los estudios del patrimonio, y pretende contribuir al diálogo interdisciplinar en torno a la conservación digital, proponiendo soluciones concretas y sostenibles para preservar las colecciones textiles europeas en el siglo XXI.

Bibliografía

KLEPACKI, L., & SCHRÖDER, M. K. (2023), «The transformation of museum exhibitions in the era of digital objects», en [obra no especificada], pp. 179–188. https://doi.org/10.1007/978-981-19-3915-0_15

MELCHIOR, M. R. (2019), «Digital fashion heritage: Understanding EuropeanaFashion.eu and the Google Cultural Institute's "We Wear Culture"», *Critical Studies in Fashion & Beauty*, 10(1), pp. 49–68.

STAMENOV, S., & STAMENOVA, V. (2021), «Geoinformation technologies for conservation of cultural heritage», en M. IOANNIDES, E. FINK, L. CANTONI, & E. CHAMPION (eds.), *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, Cham, Springer, pp. 535–542.

TIMANS, R., WOUTERS, P., & HEILBRON, J. (2019), «Mixed methods research: What it is and what it could be», *Theory and Society*, 48(2), pp. 193–216.

TOMOVIC, T. (2021), «Silk Memory: Targets beyond the web portal», en *Weaving Europe: Silk Heritage and Digital Technologies*, Valencia, Tirant lo Blanch, pp. 279–288.

WINTER, T. (2021), «The geocultural heritage of the Silk Roads», *International Journal of Heritage Studies*, 27(7), pp. 700–719.

Persuasión y emoción: la prosodia como herramienta argumentativa en el discurso político oral

Carlos Castelló Vercher
Noelia Ruano Piqueras
Universitat de València

Palabras clave: Prosodia; emoción; discurso político oral; oralidad prefabricada; persuasión
Keywords: Prosody; emotion; oral political discourse; prefabricated orality; persuasion

Introducción. Descripción del problema de investigación

El estudio de la lengua oral a lo largo de las últimas décadas ha permitido comprender que la comunicación humana trasciende la mera transmisión de información lógica o datos objetivos. Al hablar, no solo comunicamos ideas, sino que también expresamos emociones, matices afectivos y estados internos. A pesar de ello, tradicionalmente, el estudio de las emociones ha sido considerado como un tema marginal dentro del análisis lingüístico, ya que tienden a ser percibidas como fenómenos idiosincrásicos, subjetivos o irracionales. Sin embargo, estudios recientes (Barret, 2017) evidencian que emoción y pensamiento están profundamente interrelacionados, lo que influye tanto en nuestra forma de hablar como en la manera en que interpretamos lo que escuchamos.

En este marco de análisis, la participación de la prosodia resulta determinante; efectivamente, elementos como la variación tonal, el volumen y el ritmo revelan información emocional de manera automática e inconsciente. En la misma línea, emociones intensas como la ira o la sorpresa tienden a elevar la frecuencia fundamental, aumentar la intensidad vocal y acelerar la velocidad del habla, mientras que otras como la tristeza producen un tono más bajo, menor energía y una ralentización en la velocidad elocutiva (Padilla, 2022).

Por lo demás, estudios actuales subrayan que los parámetros prosódicos son fundamentales para delimitar con mayor exactitud el análisis de las emociones, pero además ponen el acento en la necesidad de estudiarlas en contextos comunicativos genuinos y no artificiales, recurriendo a material oral espontáneo (Padilla, 2022). Desde esta perspectiva, el discurso coloquial representa un ámbito privilegiado para observar la expresión natural de las emociones, pues se desarrolla en situaciones de interacción cotidiana entre

interlocutores que comparten códigos lingüísticos, expectativas comunicativas y marcos interaccionales comunes.

En todo caso, dada la complejidad del fenómeno emocional y la diversidad de registros en que se manifiesta, resulta igualmente pertinente explorar contextos comunicativos donde la entonación emocional adquiere una dimensión performativa y convencionalizada. Por ello, en la presente investigación se ha optado por trabajar con materiales que, si bien no son estrictamente coloquiales o espontáneos, permiten analizar cómo se construye y modula la emoción a través de recursos prosódicos en situaciones de comunicación pública. En este sentido, se ha seleccionado un corpus compuesto por textos audiovisuales procedentes del discurso político o parlamentario de cuatro partidos políticos. Esta elección responde a lo señalado por Cantero (2014), quien sostiene que los actores y oradores recurren a un repertorio entonativo basado en convenciones compartidas con su audiencia, lo que evidencia la existencia de un código expresivo relativamente estable capaz de suscitar respuestas emocionales eficaces. Este tipo de material, denominado “oralidad prefabricada” (Baños Piñero, 2009), se sitúa así en un punto intermedio entre la oralidad espontánea y la actuación, ofreciendo una fuente especialmente útil para profundizar en el estudio de la prosodia emocional.

Entendemos, además, que el discurso político es cada vez más accesible y cercano al ser retransmitido de modo insistente desde las redes sociales, tanto por los propios partidos como por candidatos o afiliados a una determinada formación política. Este nuevo espacio se concibe como un escenario, donde los propios actores políticos hacen llegar directamente sus mensajes a la ciudadanía. Esta transformación del ámbito comunicativo contribuye a una creciente teatralización del discurso político, que desplaza el debate público racional y lo sustituye por una apelación preferente a la dimensión emocional. El objetivo principal es captar la atención de la audiencia y propiciar la adhesión de nuevos simpatizantes a una determinada opción política (Cala Siria, 2019). En palabras de Christian Salmon, “se produce la sustitución del debate público por la mera captación de emociones” (Salmon, 2010: 222).

Este nuevo escenario ha sido abordado desde diferentes enfoques. Neuman, Marcus, MacKuen y Cigler (2007) aluden así al denominado “efecto del afecto”, según el cual son las emociones las que impulsan la acción ciudadana. Por su parte, Perniola (2004, apud Dader, 2009) introduce el concepto de “sensiología” para referirse al estudio sistemático de las emociones, mientras que Dader (2015) adapta esta noción para acuñar el término “sensiocracia”, con el que designa un modelo de evaluación política dominado por criterios emocionales. En la misma línea, Drew Western (2007) sostiene que los ciudadanos tienden a votar por aquellos candidatos que logran despertar los sentimientos adecuados, más que por quienes ofrecen los argumentos más sólidos.

En definitiva, la emoción parece ocupar un lugar central en la construcción del discurso político oral contemporáneo: la persuasión ya no depende solo de

la solidez argumentativa, sino de la capacidad de generar respuestas afectivas en el público. La prosodia, como componente esencial del habla, permite no solo expresar dichas emociones de forma espontánea o estratégica, sino también delimitar y analizar su manifestación mediante parámetros observables como el tono, la intensidad o la velocidad del habla. El discurso político —especialmente aquel que se desarrolla en contextos mediáticos y parlamentarios— emplea, pues, recursos prosódicos codificados para activar emociones concretas, enmarcando el mensaje dentro de una lógica más emocional que racional. Así, conceptos como “sensiocracia” u “oralidad prefabricada” ayudan a ilustrar este “nuevo” modo en que la dimensión afectiva funciona como herramienta discursiva clave, empleando la prosodia como vehículo clave de conexión entre el orador político y su audiencia.

Metodología y resultados

Con el objetivo de explorar estas dinámicas, esta investigación ha trabajado con un corpus compuesto por discursos de los líderes de los cuatro principales partidos políticos españoles (PSOE, PP, VOX y SUMAR). Se han seleccionado intervenciones difundidas en redes sociales por los propios partidos que mostraban de forma clara la manifestación de alguna emoción. Para analizar este material se han llevado a cabo dos pruebas perceptivas, diseñadas con el software FOLERPA (https://ilg.usc.gal/folerpa/), herramienta que permite preparar y gestionar experimentos de percepción, facilitando la presentación controlada de estímulos audiovisuales a los participantes. La primera prueba ha consistido en identificar las emociones expresadas escuchando únicamente el audio de los discursos. La segunda ha incorporado además el soporte visual, permitiendo valorar el impacto de la imagen en la percepción emocional. Ambos experimentos se han aplicado a dos grupos distintos de informantes. El contraste entre ambas pruebas permitió observar hasta qué punto la gestualidad influye en la percepción emocional.

Además, dado que el foco del estudio se centra en la dimensión sonora del discurso, el trabajo perceptivo se ha complementado con un análisis acústico. Para ello se ha utilizado el software Praat (Boersma y Weenink, 2025), junto con el script “Analyse tier”, desarrollado por Daniel Hirst, que permite extraer de forma automatizada parámetros prosódicos (frecuencia fundamental, duración, intensidad) a partir de segmentos previamente anotados. Mediante estas herramientas se han obtenido los datos acústicos de las muestras, organizados en archivos TextGrid, que han tomado como unidad de análisis el “acto” conversacional (según la definición de Val.Es.Co., 2014).

Se han recopilado medidas medias, máximas y mínimas de frecuencia fundamental (F0), intensidad vocal y duración absoluta. Estos datos se han volcado luego en un documento Excel, lo que ha permitido realizar un análisis comparativo detallado. Así se ha podido identificar qué emociones predominaban en los discursos de cada partido y cómo se manifestaban prosódicamente. Como

hipótesis de trabajo, y tomando como base lo señalado por Cala Siria (2019), se plantea que los partidos con discursos más emocionales presentan una prosodia más marcada, mientras que aquellos que apuestan por un discurso más equilibrado entre razón y emoción tienden a una realización más neutra.

En conclusión, la prosodia no constituye un mero acompañamiento del discurso político, sino que desempeña un papel estructurador y potenciador que incide de manera directa en su eficacia comunicativa. En un escenario discursivo en el que la dimensión emocional se erige como un factor determinante en la construcción del vínculo con el electorado, resulta fundamental analizar los modos en que dicha emocionalidad se articula a través de los recursos sonoros. En definitiva, la configuración melódica de la voz, sus inflexiones y silencios estratégicos se revela como una vía principal para la persuasión política en el contexto contemporáneo, de tal manera que dota al mensaje de una carga expresiva que trasciende el contenido verbal y fortalece su impacto sobre la audiencia.

Bibliografía

BAÑOS PIÑERO, R. (2009), «Estudio descriptivo-contrastivo del discurso oral prefabricado en un corpus audiovisual comparable en español: oralidad prefabricada de producción propia y de producción ajena», en P. CANTOS GÓMEZ & A. SÁNCHEZ PÉREZ (eds.), *Panorama de investigaciones basadas en corpus*, Murcia, Asociación Española de Lingüística del Corpus, pp. 399–413.

BARRET, L. F. (2017), *How Emotions Are Made: The Secret Life of the Brain* (Paperback edition), London, Pan Books.

BOERSMA, P., & WEENINK, D. (2025), *Praat: Doing Phonetics by Computer* [programa informático], recurso en línea. http://www.praat.org/

CALA SIRIA, R. (2019), «Emoción y razón en la construcción del discurso político: el caso de la proclamación de la no reconocida República Catalana en octubre de 2017», *Estudios de Lingüística del Español*, 40, pp. 95–115.

CANTERO SERENA, F. J. (2014), «Códigos de la entonación y entonación emocional», en *Actas del Congreso Internacional de la Asociación Española de Lingüística Aplicada*.

DADER, J. L. (2015), «Fascinados por “Podemos”. Un fenómeno natural de la “Democracia sentimental”», *H-ermes: Journal of Communication*, 4, pp. 13–45. Disponible en: https://siba-ese.unisalento.it/index.php/h-ermes/issue/view/1306

GRUPO VAL.ES.CO. (2014), «Las unidades del discurso oral. La propuesta Val.Es.Co. de segmentación de la conversación (coloquial)», *Estudios de Lingüística del Español*, 35(1).

NEUMAN, R., MARCUS, G., MACKUEN, M., & CRIGLER, A. (2007), *The Affect Effect: Dynamics of Emotion in Political Thinking and Behavior*, Chicago, The University of Chicago Press.

PUJANTE, D. (2010), *El discurso político como discurso retórico. Estado de la cuestión*, Valladolid, Biblioteca Virtual Universal, Universidad de Valladolid.

PADILLA, X. A. (2022), «La voz como reacción emocional. De qué nos informa la prosodia», *Spanish in Context*, 19(1), pp. 72–98.

SALMON, C. (2010), *Storytelling: la máquina de fabricar historias y formatear las mentes*, Barcelona, Ediciones Península.

WESTERN, D. (2007), *The Political Brain: The Role of Emotions in Deciding the Fate of the Nation*, Nueva York, Public Affairs.

Corpus ADUMAT: claves y retos para la creación de un corpus de habla de adultos mayores

Cristina Villalba Ibáñez
Universitat de València

Palabras clave: Corpus; discurso oral; adultos mayores; envejecimiento; lingüística de corpus
Keywords: Corpus; oral discourse; older adults; aging; corpus linguistics

El envejecimiento de la población es un fenómeno consolidado en los países desarrollados, donde la esperanza de vida ha aumentado y los nacimientos son cada vez más escasos. Consecuentemente, es preciso que desde la comunidad científica surjan trabajos centrados en este sector de la población, de modo que puedan ayudar a anticipar respuestas a futuros retos derivados de esta tendencia.

Cabe señalar que, en algunas áreas del conocimiento como la medicina o la psicología, llevan décadas realizando investigaciones centradas en adultos mayores; sin embargo, en el área de las humanidades y, en concreto, en la lingüística hispánica, el desarrollo de estudios monográficos sobre la tercera edad es todavía escaso, especialmente si se compara con el ámbito de la salud. Por ello, la creación de bases de datos y la compilación de muestras lingüísticas que documenten el discurso de los mayores puede entenderse como una manera de apoyar la investigación sobre este sector poblacional a la vez que abre la puerta a estudios con especialistas de distintas disciplinas.

En relación con lo dicho hasta ahora, el objetivo de esta comunicación es presentar el corpus ADUMAT, un corpus de reciente creación dedicado al habla de los adultos mayores y reflexionar desde un punto de vista metodológico sobre los principales retos que debe afrontar el/la analista a la hora de crear un corpus compuesto por informantes de la tercera edad. La propuesta se inscribe en el marco de las humanidades digitales, al integrar tecnologías de grabación, transcripción y anotación lingüística, así como principios de accesibilidad y sostenibilidad de los datos. A través de este enfoque, se busca no solo crear un recurso útil para la investigación lingüística, sino también contribuir a la representación digital de voces tradicionalmente invisibilizadas.

El discurso en personas mayores se ha estudiado principalmente desde el ámbito de la psicolingüística y la sociolingüística. Por una parte, la psicolingüística adopta una perspectiva eminentemente clínica. Su interés radica en explicar o prevenir las manifestaciones patológicas o el deterioro del lenguaje a nivel cognitivo, acústico o articulatorio. Sin embargo, esta disciplina tiende a

descuidar la dimensión social del lenguaje. Por otra parte, la sociolingüística aborda el estudio del discurso considerando diversos factores, entre los cuales figura la edad. Con todo, los estratos etarios más avanzados apenas han captado el interés individualizado por parte de la comunidad investigadora, a diferencia de lo que ocurre con otros grupos como el habla adolescente o de adultos jóvenes, que han sido objeto de mayor estudio (Coupland, 2004; Pichler *et al*., 2018).

Una de las consecuencias de esta situación es que puede derivar en un sesgo edadista, tanto a nivel científico como a nivel general de la población, pues se puede caer en el peligro de concebir la población mayor como un grupo patologizado o, en el mejor de los casos, entenderlo como "grupo de control" de lo que se considera el estándar adulto. En esta línea, el corpus ADUMAT pretende ser una herramienta que permita a los investigadores acercarse al discurso de este grupo etario de manera monográfica y, al mismo tiempo, ofrecer al investigador claves metodológicas que puedan ayudar a aliviar los posibles escollos que surjan durante la creación de corpus centrados en la población mayor.

Por lo que respecta a la metodología, se han compilado conversaciones espontáneas de entre 2 y 4 hablantes. Para favorecer la espontaneidad y autonomía de los participantes, las grabaciones se realizaron en espacios de socialización habitual (cafeterías y espacios del club social) y se ha procurado que los hablantes tengan un conocimiento común compartido (la relación entre los participantes va desde amigos, matrimonios y compañeros de clase hasta miembros del mismo club social). Pese al cuidado en la creación de espacios amables para que fluyera la conversación, en aquellos intercambios en los que los participantes se encontraban cohibidos, fue necesaria la intervención de la persona al cargo de la grabación como participante.

En lo que respecta a la selección de informantes, los criterios de selección de informantes parten de ocho categorías. Estas aparecen descritas de forma resumida en la tabla siguiente.

Criterios de selección de participantes	
Edad	A partir de 60 años
Nacionalidad	Española
Zona geográfica	Residentes en Valencia y su área metropolitana durante más de 10 años
Lengua de comunicación	Castellano (o grado alto de competencia lingüística en castellano)
Nivel educativo	1 (sin estudios y estudios primarios), 2 (estudios secundarios), 3 (educación superior)
Género	Hombre y mujer
Profesión	Respuesta abierta
Convivencia	Solo/a, en pareja, con familiares

Tabla 1. Criterios de selección de participantes

El proceso de grabación, transcripción y almacenamiento de los datos de este corpus cumple con las directrices éticas de la Universitat de València. En primer lugar, se informó verbalmente y por escrito a los informantes interesados de que iban a ser grabados y del tratamiento que iban a recibir sus datos de acuerdo con la política de privacidad de la Universitat de València. Una vez informados y firmados los consentimientos pertinentes, se inició el proceso de registro de la conversación mediante un dispositivo de grabación y se completó una ficha técnica de los informantes con los datos mencionados en la Tabla 1. Seguidamente, se inició el proceso de transcripción mediante la herramienta ELAN con el objetivo de alinear el audio con la transcripción, además de anonimizar los audios y las transcripciones durante este proceso. Para enriquecer el texto transcrito y permitir búsquedas especializadas, se siguió un sistema de etiquetado que combina símbolos del sistema de transcripción de Val.Es.Co. (Pons Bordería, en línea) y etiquetas basadas en el sistema desarrollado por el proyecto PRESEEA (Moreno Fernández, 2021), tal y como se propone en el corpus AMERESCO (Albelda & Estellés, en línea). Estas etiquetas son compatibles con XML, de modo que se garantiza la recuperación posterior de la información mediante motores de búsqueda. Finalmente, el acceso al corpus transcrito y las fichas técnicas puede realizarse a través de www.uv.es/adumat. En el caso de las grabaciones, en estos momentos es necesario completar un formulario indicando los datos del investigador y el propósito de la investigación para darle acceso a este contenido.

La elaboración del corpus hasta este momento nos ha permitido experimentar y reflexionar sobre algunos de los principales retos vinculados a la elaboración de un corpus de estas características. Uno de los primeros desafíos metodológicos ha sido establecer el umbral etario para definir al colectivo de “adultos mayores”. No existe consenso entre disciplinas respecto a cuándo comienza esta etapa vital. En lingüística, esta indefinición se traduce en distintas formas de categorización etaria en los corpus sociolingüísticos, donde la forma más extendida es situar el tercer grupo de edad a partir de los 55. Sin embargo, en lugar de fijar únicamente criterios cronológicos, la experiencia de campo sugiere que pueden resultar más relevantes otros factores de carácter social y biográfico, como el nivel de socialización, la trayectoria laboral o el contexto de jubilación, lo que plantea la necesidad de incorporar nuevas categorías para caracterizar lingüísticamente a esta población y nuevos retos para acceder a la población con un grado de socialización más bajo.

Esta conclusión enlaza con el hecho de que las personas mayores constituyen un grupo altamente heterogéneo, cuyas vivencias acumuladas inciden en su identidad discursiva. Dicha variabilidad, sin embargo, contrasta con la tendencia de los estudios variacionistas clásicos a excluir a quienes presentan deterioros físicos o cognitivos. Este enfoque, como señalan Pichler *et al.* (2018), ofrece una visión parcial de la tercera edad, obviando que muchas de estas condiciones forman parte natural del envejecimiento. En nuestro corpus, no se han establecido filtros diagnósticos, aunque el contacto inicial con informantes

se ha canalizado a través de espacios de socialización activa, lo que tiende a incluir a personas con buena autonomía funcional. En estadios futuros, se plantea incluir en el corpus a informantes cuyo grado de dependencia les haya llevado a ingresar en residencias para evaluar en qué medida este factor también puede afectar al discurso.

Por último, es preciso destacar que, a nivel logístico, el acceso a informantes ha sido complejo. Las restricciones institucionales para grabar en centros públicos y la reticencia de los propios participantes —motivada por cierta inseguridad lingüística o miedo al uso indebido de su voz, especialmente en los sectores con menor grado de instrucción— han exigido estrategias empáticas y flexibles. Resulta llamativo en nuestros datos, además, la observación de un mayor interés por participar en este estudio por parte de la población femenina.

Bibliografía

ALBELDA, M., & ESTELLÉS, M. (s. f.), *Corpus Ameresco* [software], Universitat de València, recurso en línea. http://esvaratenuacion.es/consulta

COUPLAND, N. (2004), «Age in social and sociolinguistic theory», en J. F. NUSSBAUM & J. COUPLAND (eds.), *Handbook of Communication and Aging Research*, Routledge, pp. 69–90. https://doi.org/10.4324/9781410610171-11

MORENO FERNÁNDEZ, F. (ed.) (2021), *Marcas y etiquetas mínimas obligatorias para materiales de PRESEEA*, Universidad de Alcalá. https://preseea.uah.es/sites/default/files/2022-02/Marcas%20y%20etiquetas%20m%C3%ADnimas%20obligatorias%20para%20materiales%20de%20PRESEEA_Moreno%20Fern%C3%A1ndez%20%282021%29.pdf

PICHLER, H., WAGNER, S. E., & HESSON, A. (2018), «Old-age language variation and change: Confronting variationist ageism», *Language and Linguistics Compass*, 12(6), art. e12281. https://doi.org/10.1111/lnc3.12281

PONS BORDERÍA, S. (s. f.), *Corpus Val.Es.Co. 3.0* [software], Universitat de València, recurso en línea. http://www.valesco.es

Memoria colectiva y reconstrucción virtual orientados a la terapia del deterioro cognitivo: estudio piloto del caso de la ciudad de Burgos en 1970

Roberto Martín González
Montserrat Santamaría Vázquez
Olalla Sáiz Vázquez
Universidad de Burgos

Palabras clave: DCL; RV; memoria; reconstrucción virtual; terapia de reminiscencia
Keywords: MCI; VR; memory; virtual reconstruction; reminiscence therapy

Introducción

El envejecimiento global ha incrementado el interés por abordar el deterioro cognitivo leve (DCL), donde la terapia de reminiscencia se presenta como una estrategia eficaz para estimular la memoria y mejorar el bienestar emocional. Aunque muchas investigaciones han explorado el uso de realidad virtual (RV) con este fin, la mayoría se centran en terapias grupales, ya que personalizarlas requiere más tiempo y recursos.

La RV permite recrear lugares inaccesibles o desaparecidos, como entornos patrimoniales (Lau et al., 2024), y resulta especialmente útil para diseñar terapias de reminiscencia individualizadas. En este contexto, la reconstrucción virtual de espacios significativos favorece la evocación de recuerdos y refuerza la conexión emocional con el pasado.

Siguiendo esta línea, se desarrolló una prueba piloto basada en la recreación virtual de la ciudad de Burgos en los años 70, incluyendo hitos urbanos reconocibles. La aplicación, con una interfaz sencilla y adaptada al entorno local, se creó mediante un enfoque multidisciplinar que integra Terapia Ocupacional, Historia, Comunicación y Nuevas Tecnologías. Además de validar su viabilidad técnica, el proyecto explora el valor terapéutico de la RV en personas con DCL, acercándolas al patrimonio cultural y reforzando sus capacidades cognitivas.

Estado de la cuestión

La enfermedad de Alzheimer (EA) es una forma común de demencia que suele comenzar con un estadio de deterioro cognitivo leve (DCL), caracterizado por síntomas que afectan las actividades de la vida diaria (Riaz et al., 2021).

Detectar estos signos tempranos es crucial, ya que algunos déficits podrían ser reversibles (Petersen, 2016).

La terapia de reminiscencia (TR) es una intervención no farmacológica que facilita el recuerdo de experiencias pasadas, promoviendo el bienestar emocional y la aceptación personal (Jahanbin et al., 2014). Según Jahanbin, Mohammadneja y Sharif (2014), la TR grupal puede prevenir problemas asociados al envejecimiento como la EA. En su modalidad individual y personalizada, ha demostrado mejorar la socialización, reducir la apatía y depresión (Chiang et al., 2010), y favorecer funciones cognitivas y calidad de vida (Justo-Henriques et al., 2021).

Vaillant-Ciszewicz et al. (2023) evaluaron una TR personalizada mediante realidad virtual basada en escenarios familiares, observando mejoras en estado de ánimo y memoria, así como una reducción de síntomas depresivos y de apatía.

Metodología

Tipo de estudio

La metodología trazada para esta prueba piloto se establece en tres fases (figura 1). En primer lugar, se plantea un estudio cualitativo basado en grupos focales para identificar hitos locales y seleccionar qué lugares de la ciudad de Burgos podrían resultar más reconocibles para los usuarios del estudio final. En segundo lugar, se diseña y desarrolla la aplicación en base a lo concluido en la fase 1. Y la tercera y última plantea una validación preliminar para evaluar la viabilidad de la propuesta y el diseño experimental de cara a futuros estudios.

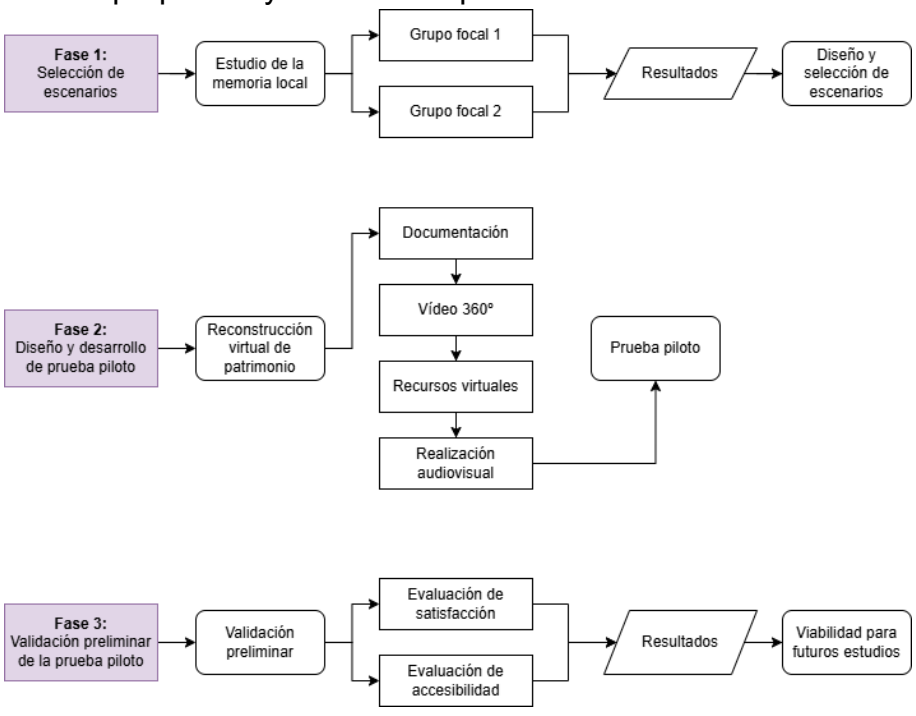


Figura 1. Diagrama de flujo de la metodología establecida

Muestra

El estudio contó con dos muestras: una en la fase 1 (grupos focales) y otra en la fase 3 (validación con usuarios; ver apartado 4.3). Para la fase 1, los criterios de inclusión fueron: personas nacidas entre 1950 y 1960, que vivieron en Burgos en los años 70 entre los 18 y 27 años (franja de mayor conservación de la memoria autobiográfica). Aunque compartían este perfil, existían diferencias en ocupación, nivel educativo y barrio de residencia. En la fase 3, desarrollada con la Asociación Alzheimer Burgos (Afabur), los criterios de inclusión fueron: personas nacidas entre 1950 y 1960, con experiencia de vida en Burgos durante los años 70, puntuación en el MEC entre 20 y 30 (para garantizar seguridad en el uso de la aplicación) y sin patologías visuales graves u otras condiciones asociadas. Fueron criterios de exclusión en esta fase: puntuación en el MEC inferior a 20, presencia de enfermedades neurológicas graves o problemas severos de visión u otras condiciones que impidieran el uso de gafas de realidad virtual.

Procedimiento

Fase 1: Selección de escenarios y eventos a reconstruir

En esta fase se llevó a cabo una investigación cualitativa mediante grupos focales, con el objetivo de identificar los escenarios más adecuados para la reconstrucción virtual. Se diseñó un guion de entrevista que abordaba desde recuerdos generales de la ciudad hasta aspectos concretos como lugares de encuentro, ocio, vestimenta y actitudes de la época. Esto permitió superar la barrera generacional, ya que ninguno de los participantes había vivido esa década como adulto. La muestra se dividió en dos grupos mixtos (n=6 y n=5), lo que permitió alcanzar saturación de la información, es decir, coincidencias suficientes como para no requerir un tercer grupo. Tras analizar las transcripciones, se seleccionaron dos escenarios clave para su recreación virtual.

Fase 2: Diseño y desarrollo de la prueba piloto

En esta segunda fase, una vez analizados los resultados de la fase 1, se abordó la reconstrucción del patrimonio en RV. En primer lugar, se planteó una fase de recopilación documental con el objetivo de vislumbrar el estado de los escenarios a reconstruir durante la década de los 70’. Una vez obtenida la documentación suficiente, se pudo elegir la ubicación del usuario para realizar la fotografía 360º que serviría de base para el posterior trabajo virtual, basado en las distintas metodologías de desarrollo digital como la fotogrametría, el modelado 3D y el matte painting. Diseñado el escenario virtual, se planteó el rodaje con actores caracterizados con la ropa y actitudes de la época para su posterior integración en la experiencia.

Fase 3: Validación preliminar de la prueba piloto

En esta fase se contactó con una muestra de personas con DCL de la Asociación Azlheimer Burgos. Se presentó el escenario virtual a cada persona por un tiempo de 4 minutos, bajo la asistencia y guía de un terapeuta ocupacional. Se llevaron a cabo entrevistas en profundidad a cada uno de los usuarios, apoyados en dos pruebas de validación: la MARS: Mobile App Rating Scale y la SSQ: Simulator Sickness Questionnaire.

Resultados

Fase 1: Selección de escenarios y eventos a reconstruir

A partir de los grupos focales, se identificaron los lugares más recordados y con mayores transformaciones desde los años 70, destacando la Plaza Mayor, el paseo del Espolón, el paseo de la Quinta y, especialmente, el Reloj "El Morito", punto de encuentro habitual en la época. También se obtuvieron datos sobre vestimenta, peinados y actitudes propias de una década marcada por el cambio social y político, así como impresiones comunes sobre la peatonalización creciente y el tráfico intenso en el centro. Para la prueba piloto se eligió la zona de “El Morito” por reunir elementos clave (arquitectura modificada, negocios desaparecidos, vehículos, cartelería) y ofrecer una mayor viabilidad técnica para su recreación virtual.

Fase 2: Diseño y desarrollo de la prueba piloto

Una vez seleccionado el escenario, se inició la reconstrucción virtual en varias fases. Primero se recopilaron datos visuales a través de archivos municipales, provinciales, RTVE y revistas de la época, lo que permitió documentar la arquitectura, el mobiliario urbano, los negocios desaparecidos y los estilos de vestimenta e interacción social característicos de los años 70. Estos datos confirmaron hallazgos previos (ver apartado 5.1), como el deterioro de fachadas por la contaminación o la presencia de comercios emblemáticos como la Óptica Nacional y Radio Castilla.

Posteriormente, se realizó una fotografía 360º en la ubicación seleccionada (frente a la antigua Dirección General de Turismo), que ofrecía una vista estratégica de puntos clave como la calle Vitoria, el paseo del Espolón y edificios significativos. A partir de esta imagen, se clasificaron las fachadas mediante un código alfanumérico para ubicar con precisión los elementos a reconstruir.

El entorno digital se desarrolló utilizando fotogrametría para crear una malla 3D precisa, y Blender para modelar el suelo y objetos como farolas o vehículos, algunos obtenidos de bancos libres. Las fachadas lejanas se recrearon con matte painting en Photoshop, agilizando el proceso sin comprometer el detalle. Durante todo el diseño se utilizaron gafas de realidad virtual para comprobar proporciones y ubicaciones.

Finalmente, se llevó a cabo un rodaje con voluntarios caracterizados como ciudadanos de los años 70. Las escenas se basaron en documentación previa y se diseñaron como micronarraciones en bucle, adaptadas a la impredecibilidad del usuario (Alaguero et al., 2021). Se incluyeron acciones como un hombre esperando bajo el reloj “El Morito” o mujeres comiendo castañas. Por razones logísticas, se emplearon pocos figurantes, aunque se prevé ampliar el número en futuras versiones para reforzar la inmersión.

Figura 2. Prueba piloto del escenario virtual resultante

Fase 3: Validación preliminar de la prueba piloto

Finalizada la prueba piloto, el equipo acudió al centro de día de la Asociación Alzheimer Burgos para evaluar la experiencia con una muestra reducida (n=5; n=3 mujeres y n=2 hombres), en colaboración con terapeutas del centro e investigadores de Terapia Ocupacional de la UBU. Los resultados fueron positivos, con buena percepción de inmersión y presencia, y un alto nivel de satisfacción reflejado en la prueba MARS.

Los usuarios aportaron sugerencias como incluir más movimiento en el paseo del Espolón, añadir vehículos, mejorar la legibilidad de carteles (ej. bar "Pinedo") y recuperar negocios emblemáticos como la cafetería “Viena”. La experiencia estimuló la conversación y la evocación de recuerdos. Además, se propuso como mejora técnica conectar las gafas de RV a un monitor para que el terapeuta pueda seguir la sesión en tiempo real.

Conclusiones

Se reconocen las limitaciones del estudio, como el pequeño tamaño de la muestra y la necesidad de un diseño experimental más completo. No obstante, los resultados de esta primera prueba piloto muestran que la herramienta diseñada es segura y con potencial terapéutico, dando así luz verde a futuras investigaciones. Se valora la eficacia de la aplicación para evocar recuerdos y estimular la socialización entre los usuarios, además de la alta accesibilidad y niveles de inmersión. Se plantea, de cara al futuro, la ampliación de los entornos virtuales, la realización de un ensayo controlado más amplio y la exploración de formas de hacer esta tecnología más accesible para su uso clínico. En general, este proyecto interdisciplinar abre las puertas al estudio en profundidad del

potencial de la reconstrucción virtual del patrimonio basado en la memoria colectiva como enfoque innovador aplicado a la terapia cognitiva, aprovechando las conexiones emocionales con la historia y la cultura locales.

Bibliografía

ALAGUERO, M., ANDRÉS LÓPEZ, G., CHECA, D., SAIZ-MANZANARES, M., & CARLOS HUGO, S. (2021), «Nuevas tecnologías en la enseñanza: explorando la narrativa audiovisual en 360º para la reconstrucción virtual de entornos urbanos desaparecidos», recurso en línea. http://dx.doi.org/10.2307/j.ctv2gz3wp2.22

CHIANG, K. J., CHU, H., CHANG, H. J., CHUNG, M. H., CHEN, C. H., CHIOU, H. Y., & CHOU, K. R. (2010), «The effects of reminiscence therapy on psychological well-being, depression, and loneliness among the institutionalized aged», *International Journal of Geriatric Psychiatry*, 25(4), pp. 380–388. https://doi.org/10.1002/gps.2350

JAHANBIN, I., MOHAMMADNEJAD, S., & SHARIF, F. (2014), «The effect of group reminiscence on the cognitive status of elderly people supported by Ilam Welfare Organization in 2013; a randomized controlled clinical trial», *International Journal of Community Based Nursing and Midwifery*, 2(4), pp. 231–239.

JUSTO-HENRIQUES, S. I., PÉREZ-SÁEZ, E., & ALVES APÓSTOLO, J. L. (2021), «Multicentre randomised controlled trial about the effect of individual reminiscence therapy in older adults with neurocognitive disorders», *International Journal of Geriatric Psychiatry*, 36(5), pp. 704–712. https://doi.org/10.1002/gps.5469

LAU, J. S., TANG, Y. M., GAO, G., et al. (2024), «Development and usability testing of virtual reality (VR)-based reminiscence therapy for people with dementia», *Information Systems Frontiers*. https://doi.org/10.1007/s10796-024-10479-w

PETERSEN, R. C. (2016), «Mild cognitive impairment», *Continuum (Minneapolis, Minn.)*, 22(2, Dementia), pp. 404–418. https://doi.org/10.1212/CON.0000000000000313

RIAZ, W., KHAN, Z. Y., JAWAID, A., & SHAHID, S. (2021), «Virtual reality (VR)-based environmental enrichment in older adults with mild cognitive impairment (MCI) and mild dementia», *Brain Sciences*, 11(8), art. 1103. https://doi.org/10.3390/brainsci11081103

VAILLANT-CISZEWICZ, A. J., QUIN, C., MICHEL, E., SACCO, G., & GUERIN, O. (2023), «La réalité virtuelle personnalisée sur les troubles de l'humeur en établissement d'hébergement pour personnes âgées dépendantes et unité de soin de longue durée: étude de cas chez un résident présentant des troubles cognitifs modérés», *Annales Médico-psychologiques, Revue Psychiatrique*, 181(10), pp. 848–854. https://doi.org/10.1016/j.amp.2022.10.018

Confluencias significantes: explorando las relaciones de producción y las conexiones iconográficas entre las ediciones ilustradas de las *Metamorfosis* de Ovidio y los *Emblemata* de Alciato de los siglos XVI y XVII

Fátima Díez Platas
Universidad de Santiago de Compostela
María Luisa Díez Platas
Universidad Internacional de La Rioja
Paloma Centenera Centenera
Universidad San Pablo-CEU

Palabras clave: Libro ilustrado; grabados; temas mitológicos; emblemas; impresores.
Keywords: Illustrated book; engravings; mythological themes; emblems; printers.

Las ediciones ilustradas del poema de las *Metamorfosis* de Ovidio se convierten en un *best-seller* editorial en la Europa del siglo XVI, hasta el punto de que el siglo se conoce como la *Aetas Metamorphoseos.* De manera paralela, en 1531 aparece en Ausburgo la primera edición impresa del libro de los *Emblemata* del erudito italiano Andrea Alciato que inaugura un género editorial que gozará de una difusión inusitada en la Europa de los siglos XVI y XVII. Concebidas como obras cultas para educar y transmitir el conocimiento de la Antigüedad, ambas producciones librarias comparten una serie de características que las ponen en un nivel similar desde varios puntos de vista. Las ediciones ilustradas del poema de las *Metamorfosis* constituyen un género en sí mismo al establecer tradiciones de ilustración que muestran el desarrollo de la imagen que pretende representar los episodios mitológicos que Ovidio narra en el poema. Las series de grabados que, a partir de la edición veneciana de 1497, acompañan de manera regular a un nutrido grupo de ediciones del poema que incorporan ilustraciones para representar los contenidos mitológicos, las historias y episodios narrados por el poeta romano, constituyen un reservorio figurado de motivos y soluciones figurativas que tienen una evidente influencia en las propuestas figuradas en relación con la representación de los contenidos mitológicos.

Por su parte, los *Emblemata* de Andrea Alciato constituyen un género editorial novedoso que apunta al conocimiento y a la ilustración utilizando la síntesis a través de una estructura tripartita que se convierte en el sistema de

presentación de los contenidos. Sirviéndose de una máxima que se tipifica como *motto,* del valor de la poesía que glosa la explicación del *motto* en el epigrama, a partir de la publicación de Steyner de la primera edición impresa en 1531, la imagen se incorpora a los poemas titulados, la *pictura,* en forma de grabados, que se conciben como un refuerzo de significado que, a través del medio figurado, transmite de modo sintético los contenidos que el *motto* y el epigrama pretenden conferir.

La producción de las ediciones ilustradas de las *Metamorfosis* y los *Emblemas* de Alciato durante los siglos XVI y XVII se realiza en el entorno culto de la Europa del momento y se difunden en los mismos entornos, pero sobre todo se producen en los mismos lugares, en las mismas casas de impresión y son ilustrados por los mismos grabadores sirviéndose de los mismos motivos iconográficos, y en muchas ocasiones reutilizando de forma evidente los mismos grabados.

El propósito de esta contribución es mostrar un desarrollo que se está empezando a realizar a partir de la *Biblioteca Digital Ovidiana* (BDO).

Esta propuesta, que está en sus primeras fases, consiste en la exploración de las relaciones entre los impresores, libreros y grabadores en Europa durante los siglos XVI y XVII atendiendo a las conexiones entre las ediciones ilustradas de las obras de Ovidio y las ediciones de los *Emblemas* de Alciato que se editan de forma contemporánea. El objetivo es utilizar la información estructurada que contiene la BDO (http://www.ovidiuspictus.es/bdo.php) en los datos sobre impresores, libreros y lugares de impresión de las ediciones ilustradas de las *Metamorfosis* de Ovidio en los siglos XVI y XVII e integrarla y cruzarla con la información del mismo tipo del proyecto sobre los libros de los *Emblemas* de Alciato de la Universidad de Glasgow (https://www.emblems.arts.gla.ac.uk/alciato/index.php) con el objetivo de explorar si se producen coincidencias. De este modo, empleando técnicas de análisis de grafos de conocimiento, hemos iniciado el mapeado de las redes de producción libraria en los siglos XVI y XVII en Europa.

Como segundo objetivo, hemos comenzado, además, la exploración de las relaciones entre los temas mitológicos utilizados en ambos tipos de ediciones, examinando las representaciones visuales de los episodios del mito que se encuentran tanto en las ediciones ilustradas de las *Metamorfosis* como en las ediciones de Alciato. En su relevante artículo “The Interrelationship between Emblem Books and Ovid’s *Metamorphoses*: A Survey of Research” (*Emblematica,* 17, 2009, 41-147), Mason Tung facilita a los investigadores en el texto y la imagen un campo de trabajo preparado para poner en relación los libros de *emblemas* y las ediciones ilustradas de las *Metamorfosis* en términos temáticos y de producción de imágenes. El formidable trabajo, que se presenta en forma de recopilación esquemática, sistemática y exhaustiva de las ediciones y los temas compartidos entre la ilustración del poema mitológico de Ovidio y la producción emblemática, hasta donde se nos alcanza no ha sido explotado para escrutar las fructíferas relaciones entre estos dos entornos productores de

imágenes mitológicas con valor indicativo y simbólico. De este modo con nuestro análisis de esta información pretendemos responder a la sugerencia de Tung y comenzar a abordar el análisis de estas conexiones temático-iconográficas a través de desarrollos técnicos para identificar motivos y elementos iconográficos comunes que ilustren cómo las narrativas mitológicas han impregnado la cultura visual en diversos géneros.

Nuestra investigación pretende demostrar cómo técnicas como la minería de textos, el reconocimiento de imágenes y el análisis de redes pueden revelar patrones y relaciones que pueden no ser evidentes a primera vista en los estudios tradicionales. En última instancia, aspiramos a mostrar los datos mediante recursos interactivos que permitan a los usuarios explorar las conexiones entre los textos, las imágenes, sus creadores y el contexto cultural más amplio.

Bibliografía

ALCIATUS, A. (1985), *The Latin Emblems, Indexes and Lists (Index Emblematicus)*, ed. P. M. Daly con V. W. Callahan; asist. S. Cuttler, Toronto, University of Toronto Press.

DÍEZ PLATAS, F. (2003), «Tres maneras de ilustrar a Ovidio: una aproximación al estudio iconográfico de las *Metamorfosis* figuradas del XVI», en M.ª C. FOLGAR DE LA CALLE, A. E. GOY DIZ, & J. M. LÓPEZ VÁZQUEZ (eds.), *Memoria Artis. Studia in memoriam M.ª Dolores Vila Jato*, tomo I, Santiago de Compostela, Xunta de Galicia, pp. 247–267.

DÍEZ PLATAS, F. (2017), «Desde la imagen: la Biblioteca Digital Ovidiana como instrumento de estudio e investigación iconográfica del libro ilustrado», *Studia Aurea: Revista de Literatura Española y Teoría Literaria del Renacimiento y Siglo de Oro*, 11, pp. 55–72. https://doi.org/10.5565/rev/studiaaurea.265

GREEN, H. (1872), *Andrea Alciati and His Books of Emblems: A Biographical and Bibliographical Study*, London, Trübner.

HENKEL, M. D. (1926–1927), «Illustrierte Ausgaben von Ovids *Metamorphosen* im XV, XVI und XVII Jahrhundert», *Vorträge der Bibliothek Warburg*, pp. 94–95.

HUBER-REBENICH, G., LÜTKEMEYER, S., & WALTER, H. (2014), *Ikonographisches Repertorium zu den "Metamorphosen" des Ovid*, Berlín.

MOSS, A. (1982), *Ovid in Renaissance France: A Survey of the Latin Editions of Ovid and Commentaries Printed in France before 1600*, London, The Warburg Institute, p. 70, nº 75.

SCHEVILL, R. (1971 [1913]), *Ovid and the Renascence in Spain*, New York.

TUNG, M. (1989), «Towards a new census of Alciati's editions», *Emblematica*, 4, pp. 135–176. (Actualización en *SES Newsletter*, 46, enero 2010, pp. 26–29).

TUNG, M. (2009), «The interrelationship between emblem books and Ovid's *Metamorphoses*: A survey of research», *Emblematica*, 17, pp. 41–147.

Hécuba: uso de las HD para el estudio comparado de un mito en la enseñanza preuniversitaria

Teresa Santa María
María Luisa Díez Platas
Universidad Internacional de La Rioja

Palabras clave: Propuesta didáctica; Humanidades Digitales; Primero de Bachillerato; Tragedia griega; Teatro español contemporáneo
Keywords: Teaching proposal; Digital Humanities; First year of Baccalaureate; Greek tragedy; Contemporary Spanish theater

Diversos estudios, entre otros el de Menéndez de la Cuesta (20219), han planteado la necesidad de incluir herramientas de las Humanidades Digitales como recurso didáctico en las aulas de secundaria y bachillerato También han incidido en diferentes tipos de actividades en que puede realizarse dicha inclusión. Así, Mendoza Gamiño (2016) recoge cinco posibles áreas de las HD que se pueden vincular con la enseñanza-aprendizaje, tales como: las colecciones digitales; la lectura y análisis de textos electrónicos; los mapeos geoespaciales y críticas discursivas; los grandes volúmenes de datos, la informática social, la externalización de tares y creación de redes; así como los entornos de visualización inmersiva tridimensional.

Con esta propuesta didáctica se pretende trabajar tres de estas áreas – las dos primeras y la cuarta-, así como aportar una línea que ayude, en la medida de sus posibilidades, a paliar dos problemas que afectan a la enseñanza de la asignatura de “Literatura Universal” en Primero de Bachillerato. Nos referimos a que los estudiantes no sientan los mitos clásicos como algo ajeno y que empiecen a familiarizarse con herramientas y recursos de las Humanidades Digitales, que les serán de gran utilidad en su futuro universitario y profesional. Para ello se propone el estudio comparativo del mito de Hécuba en tres versiones, la de Eurípides, del s. V a. C.; *Hécuba triste*, de Fernán Pérez de Oliva, del siglo XVI y *La hija de Dios*, de José Bergamín, de 1945.

Para poder realizar dicha propuesta didáctica se utilizará los recursos de corpus griego clásico que se encuentran en el Drama Corpora Project (DraCor), así como las versiones que pueden consultarse en línea de las dos versiones españoles del mito. También se recurrirá a datos extraídos de Shiny DraCor y de Voyant Tools. Se trata de una metodología innovadora para ellos, activa y que permite a los estudiantes adentrarse a las obras no solo a través de la lectura sino con herramientas de análisis que les permite cuestionar, contabilizar vocablos o reflexionar sobre la estructura y contenido de las tres piezas estudiadas.

Entre las actividades que proponemos, destacaríamos, en primer lugar, la comparación de grafos en dos obras de Eurípides que cuentan con una protagonista femenina: la mencionada *Hécuba* (Imagen 1) y *Las troyanas* (Imagen 2). Podrían plantearse diferentes preguntas – cómo se indican si los personajes son femeninos y masculinos, qué señalan las líneas que unen diferentes personajes -; para acabar observando la información que de tragedia que se compara ofrece la "Play info" en Shiny DraCor (Imagen 3). Las cuestiones en este último caso se ceñirán a *Hécuba,* para que analicen qué personajes interactúan más entre sí y versarán sobre qué personajes son los que comparten escenas más veces; qué personajes no se relacionan o no hablan con los otros; cuáles son los dos únicos personajes que dialogan con todos los restantes. Con este mismo recurso pueden visualizar diferentes obras de este periodo, comparando los grafos resultantes con las de Lorca o Valle-Inclán y, de esta forma, se familiarizan con este tipo de imágenes de las obras y con la información que reflejan, que no es otra que cómo los personajes se relacionan entre sí en las diferentes escenas o partes de las obras dramáticas. También resulta interesante que conozcan y comenten proyectos como DraCor, Gutenberg, Hispana o Bieses para que den su opinión sobre su relevancia y sobre cuáles les parece a ellos que podría ser relevante llevar a cabo en el futuro.

Pasaríamos ya a comparar la tragedia griega con la primera versión española, la de Fernán Pérez de Oliva, en una segunda actividad. Nos parece interesante que los estudiantes busquen, como primera parte de esta tarea, diversas ediciones que hay en la red, en abierto. Pueden luego analizar las diferentes grafías, ortografía y manera de editar las obras, por ejemplo, en 1521 y en 2019 (Pérez de Oliva y edición crítica de Araceli Hernández López, respectivamente). Respecto a la primera, se trataría de hacerlos reflexionar sobre aspectos dispares como: para qué sirven las letras que aparecen abajo y a la derecha de cada página; qué les parece que desde la BNE se vayan digitalizando las obras de la literatura que ya no están bajo los derechos de autor; qué obras y autores digitalizarían ellos si fueran bibliotecarios o directores de instituciones que se dedicarán a esta función, etc. Después se pueden leer algunos fragmentos de la obra para trabajarlos en dos sentidos.

En primer lugar, para introducir en *Voyant Tools* uno de forma aleatoria. Tendrían que comentar y localizar a qué pasaje de la *Hécuba* de Eurípides – en su versión en español – pertenece y qué similitudes y diferencias observan. También pueden ver si hay secuencias u oraciones que se repiten en una y otra, así como palabras que resultan más repetidas en ambas piezas. En segundo lugar, tomarán las palabras finales de Agamenón en ambas piezas y analizarán los resultados obtenidos, para compararlas en cuanto a su longitud y contenido. Pueden responder a preguntas como: ¿se diferencian o se asemejan más o menos que el otro pasaje? ¿Qué final te parece más justo para Hécuba y por qué? También se les puede pedir que reescriban el discurso final de *Hécuba triste*, adaptándolo al siglo XXI: ¿qué sentencia se le puede dar a una mujer ucraniana que mata al compatriota que ha entregado a su hijo a los rusos? Y

volviendo a la herramienta Voyant Tools, pueden comparar las veces que Agamenón cita a los hijos de Hécuba, Polimnéstor o a la propia Hécuba (Imagen 4) en sus palabras finales.

Por último, llegamos a la tercera de las obras comparadas. Se trata de *La hija de Dios*, de José Bergamín. Los estudiantes pueden buscar la portada de la primera edición e indagar sobre el autor del dibujo de esta: ¿a qué autor y cuadro les recuerda? ¿Tuvo alguna conexión José Bergamín con Pablo Picasso? Pueden usar buscadores de internet para procurarse toda la información, así como el contexto del 27 y exilio en que se incluye al escritor madrileño.

Una vez situado el autor, podemos centrarnos en la obra. De esta forma, comentarían el cambio del título y de los personajes principales – Teodora, Teodosia, Leoncio -, para constatar en qué medida actualiza la tragedia griega a la situación histórica de la guerra civil que Bergamín vivió en primera persona. Resulta interesante que, respecto a esta pieza, se analice también con Voyant Tools diferentes pasajes que se asemejan al texto de Pérez de Oliva. Se trata de responder a una cuestión que plantea el propio autor en el breve texto introductorio a *La hija de Dios*, donde señala que “He seguido, en estas escenas actualizan por un episodio español de nuestro tiempo la tragedia *Hécuba* de Eurípides, la antigua versión castellana del maestro Fernán Pérez de Oliva”. En grupos de tres integrantes se puede dividir la obra para que los estudiantes encuentren tres pasajes que les parecen guardan coincidencias, los analicen y presenten al resto de grupos. Después pueden compararla con la parte correspondiente de *Hécuba*, de Eurípides y volver a la pregunta de qué personajes les parece que llevan el peso de la obra; es decir, si siguen siendo Hécuba y el coro de las mujeres. También se les propondrán analizar algún pasaje de la obra por Voyant Tools, para comentar aquellos sustantivos que más se repiten en la pieza de Bergamín.

Por último, se puede plantear un ejercicio que actualice la situación de Hécuba al contexto del siglo XXI. Para ello, deben tener en cuenta los conflictos bélicos actuales y plantearse si la situación de las “vencidas” sigue siendo como en la Antigüedad. Pueden realizar una línea del tiempo que contemple diferentes versiones – tanto literarias, como artísticas o cinematográficas – del mito de Hécuba; o poner rostro a las nuevas “reinas de Troya”, es decir, a las actuales víctimas de los conflictos armados y recoger todo el material, recabado por grupos, en un Google Site o repositorio. De esta manera, ven la importancia de recoger y mantener la investigación realizada para la consulta digital de personas interesadas en el tema, con posterioridad.

Esperamos que las actividades didácticas y una metodología que ayude a introducir recursos y herramientas de las HD en cursos preuniversitarios puedan resultar muy atractivos a estas edades, tanto para adentrarse en textos cuya comprensión puede resultar compleja, como para familiarizarse con una serie de análisis que ayudan a aproximarse al texto para profundizar en su contenido y ofrecer dimensiones que lo complementan y facilitan su comprensión. Creemos que abre una vía muy interesante de iniciar, planteando actividades y recursos

que se complementen y con los que los estudiantes sientan, manipulen y crean, a su vez, materiales digitales, tras comprobar la importancia y desarrollo que las HD tienen en la cultura y enseñanza actual.

Asimismo, trabajar con corpus en abierto y ver una obra, cuyos derechos de autor aún permanecen vigentes les puede ayudar a comprender y respetar los derechos de reproducción y cita; y qué autores y obras pueden digitalizarse y cuáles no. En último lugar, este tipo de actividades puede trasladarse a otros mitos clásicos grecolatinos –Antígona o Electra–; así como de la literatura universal –Hamlet, el Quijote, don Juan–.

Figuras, tablas, mapas o gráficas

Grafo de *Hécuba*, Eurípides

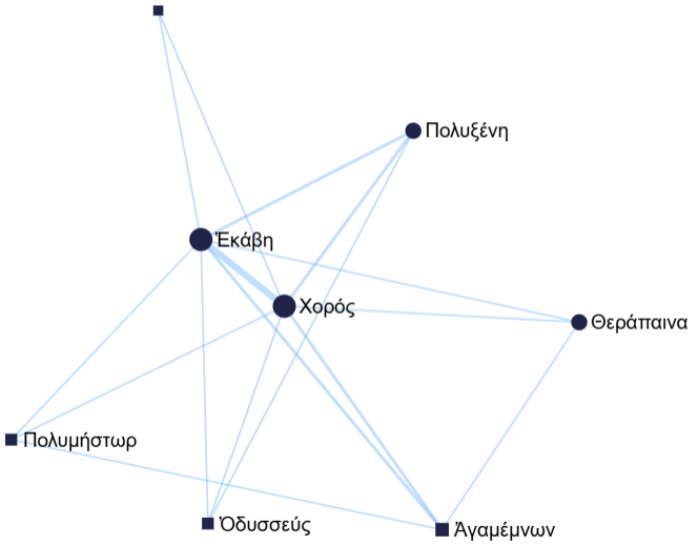


Imagen 1. Grafo extraído de DraCor: https://dracor.org/greek/euripides-hecuba

Grafo de *Las troyanas*, de Eurípides

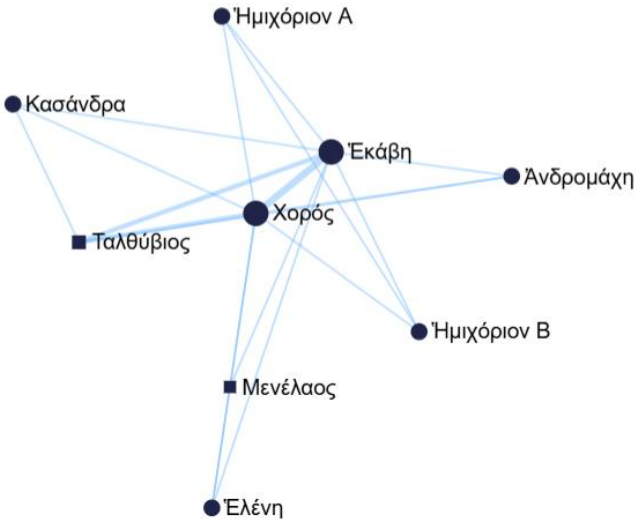


Imagen 2. Grafo extraído de DraCor: https://dracor.org/greek/euripides-the-trojan-women

Play info de *Hécuba*, de Eurípides

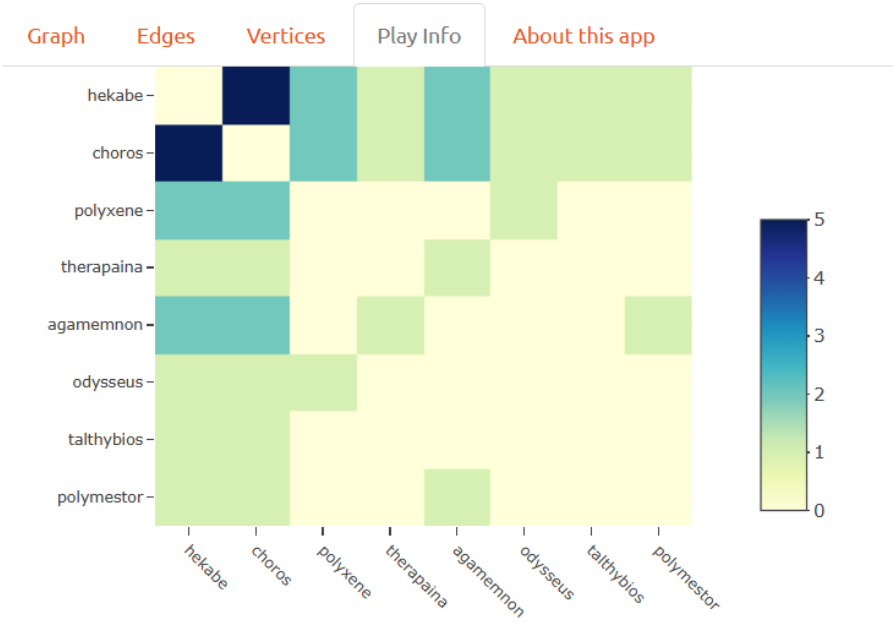


Imagen 3. Información extraída de Shiny DraCor: https://pozdniakov.shinyapps.io/Dracor/

Final de *Hécuba triste*, de Fernán Pérez de Oliva

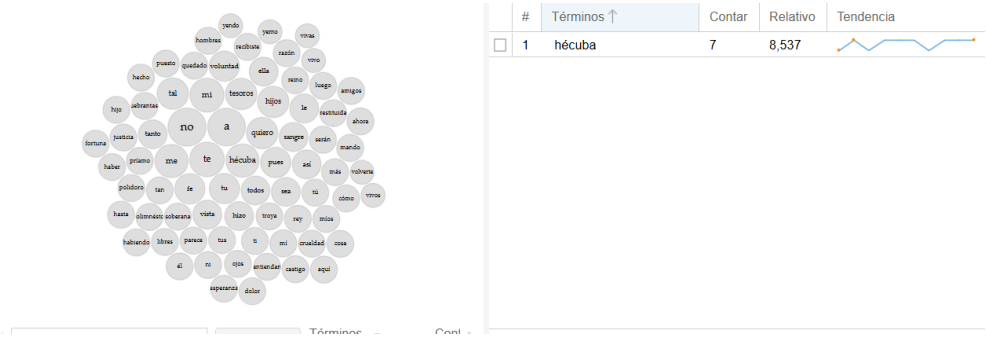


Imagen 4. Información extraída de Voyant Tools sobre el pasaje de *Hécuba triste:* https://voyant-tools.org/?corpus=097cceb248df04f4fb2bce0c826b70d7

Bibliografía

BERGAMÍN, J. (1945), *La hija de Dios y La niña guerrillera*, México, Manuel Altolaguirre Impresores. Cita en T. SANTA MARÍA, *El teatro en el exilio de José Bergamín* (tesis doctoral), tomo II, Barcelona, UAB, 2001, pp. 64–89. https://ddd.uab.cat/pub/tesis/2001/tdx-0327107-160203/mtsmf1de1.pdf

DRAMA CORPORA PROJECT (DRACOR) (s. f.), recurso en línea. https://www.dracor.org/

MENDOZA GAMIÑO, A. (2016), «El nuevo campo de investigación de las Humanidades Digitales y su vinculación con las metodologías didácticas», en R. ROIG VILA (coord.), *Tecnología, innovación e investigación en los*

procesos de enseñanza-aprendizaje, Universidad de Alicante, pp. 1741–1750.

MENÉNDEZ DE LA CUESTA GONZÁLEZ, A. (2019), «Humanidades Digitales: en español, en portugués», en *Encuentros digitales: escrituras, colecciones, aprendizajes en español*, Fundación BBVA.

PÉREZ DE OLIVA, F. (1586), *Las obras del Maestro Fernán Pérez de Oliva natural de Córdoba…: con otras cosas que van añadidas, como se dará razón luego al principio*, Córdoba, Imprenta de Gabriel Ramos Bejarano. Disponible en: http://bdh-rd.bne.es/viewer.vm?id=0000144496&page=1. Cita en A. HERNÁNDEZ LÓPEZ, *Edición crítica de la obra dramática de Hernán Pérez de Oliva: Sus tragedias La venganza de Agamenón y Hécuba triste* (tesis doctoral), Madrid, UCM, 2019, pp. 249–292. https://docta.ucm.es/rest/api/core/bitstreams/35f7b5a9-8ff7-491a-a756-dd6348080997/content

RODRÍGUEZ ADRADOS, F. (1999), *Del teatro griego al teatro de hoy*, Madrid, Alianza Editorial.

SANTA MARÍA, T. (2018), «El ciclo troyano en el teatro del exilio español», *Forma Breve*, 15, pp. 307–320.

SHINY DRACOR (s. f.), recurso en línea. https://pozdniakov.shinyapps.io/Dracor/

VOYANT TOOLS (2025), recurso en línea. https://voyant-tools.org/

Compartimos experiencias con AMIS (Advanced Metadata Intelligent System)

Fatiha Idmhand
Université de Poitiers
Amelia Sanz
Universidad Complutense de Madrid

AMIS (Advanced Metadata Intelligent System) es un proyecto financiado por la Unión Europea a través del llamado OSCAR (01-327) y liderado por la Universidad de Poitiers (Francia). Reúne a investigadores de Francia, Italia, Rumanía y España en un intento de resolver un problema concreto que toca de una u otra manera a todos los investigadores en el ámbito de las Ciencias Humanas y Sociales a la hora de manejar bibliotecas digitales y repertorios, corpus y ediciones para el análisis: los metadatos. Y es que en los últimos años ha crecido la necesidad de contar con documentos en formatos legibles por las máquinas y enmarcados por metadatos de calidad que los identifiquen y aporten potencialidad de explotación (Martin- Rodilla C González-Pérez, 2017).

Sin embargo, cualquier investigador se ha enfrentado en algún momento con la dificultad de encontrar datos (textos, ediciones, colecciones…), porque no están correctamente descritos por los metadatos, incluso son erróneos. Más aún, constatamos que los metadatos requeridos para trabajar sobre un manuscrito no son los mismos que el investigador requiere cuando comienza a trabajar con un manuscrito digitalizado. Por otro lado, sabemos que la curación de la calidad de los metadatos es tarea delicada que consume mucho tiempo, empezando por los desafíos del multilingüismo: contamos con tipologías textuales solo en francés (productos de las actividades de CAHIER), pero las ontologías suelen estar concebidas desde y para el inglés, de manera que se impone la necesidad de desambiguar términos partiendo de las lenguas románicas de los integrantes del proyecto.

Efectivamente, los metadatos suelen ser creados de forma manual, idiosincrática y extremadamente lenta sobre miles de ficheros (Galleron C Idmhand, 2020a, 2020b), lo que desemboca en un alto grado de incoherencia e inconsistencia a la hora de extraer información de un gran número de recursos, incluso dentro de un mismo campo y utilizando los mismos estándares de codificación (desde el muy flexible XML-TEI al XML-EAD, y hasta el exhaustivo Dublin Core).

AMIS propone la construcción de una herramienta accesible en forma de servicio web en abierto, destinada a producir y/o enriquecer metadatos coherentes con información adicional, de forma que se facilite al usuario la

creación rápida de metadatos de alta calidad, así como el acceso a los datos y, por ende, la descripción y análisis de contenido textual. Su objetivo es servir como asistente en la creación de metadatos, mejorar la visibilidad de los trabajos académicos y personalizar las recomendaciones de datos. Esta iniciativa responde a la "F" de "fácil de encontrar" y a la "R" de "reutilizable" de los principios FAIR.

(Semi)Automatizar ese etiquetado a partir de una aproximación contextual basada en vocabularios compartidos debe ser posible gracias a técnicas de aprendizaje automático, redes neuronales recurrentes (RNN) y memoria larga a corto plazo (LSTM) entrenados con datos almacenados en repositorios internacionales: necesitamos herramientas de Inteligencia artificial para recomendar datos también en nuestras ciencias.

La herramienta operará de la siguiente manera: en primer lugar, contará como *input* con archivos de metadatos dotados de una mínima estructura o archivos que contengan un documento textual con unos metadatos mínimamente estructurados, incluso un documento sin metadatos; en segundo lugar, en la fase de análisis, la herramienta "leerá" el archivo para identificar los rasgos mediante técnicas de PLN y bases de conocimiento con el fin de buscar y encontrar textos/documentos/metadatos similares en bases o repositorios de datos similares; en tercer lugar, como *output*, la herramienta producirá recomendaciones de metadatos a partir de vocabularios controlados existentes y permitirá al usuario elegir entre aceptar o rechazar las sugerencias; el resultado final entregado será un archivo en formato xml/TEI, csv o json con metadatos enriquecidos, así como un informe donde se expliquen tanto las sugerencias como los resultados.

Para el desarrollo del proyecto, se ha optado por una metodología *up-down*: primero, repasar tesauros y repertorios de donde se puedan recoger metadatos fiables; segundo, crear un modelo de metadatos y explorar sus posibilidades a partir de los corpus de los que disponemos con el fin de comprobar si esos metadatos concuerdan o no con lo que será nuestra ontología; ésta será la fase final del proyecto que consistirá en la construcción de una ontología con sus unidades y sus relaciones, que responda a nuestras necesidades de investigación.

La construcción de la herramienta utilizará técnicas de *machine learning*, NER (*Names Entities Recognition*), NEL (*Names Entities Linking*), clasificación de textos, etiquetados de secuencias y análisis de dependencias.

Así, por ejemplo, si el usuario está buscando información sobre la obra *Pyrrhus* de Thomas Corneille (1661), AMIS sugerirá un subtítulo, unos metadatos e incluso una clasificación del texto por su género; en todo momento, el usuario podrá aceptar o rechazar las sugerencias propuestas por la herramienta.

Por encontrarse el proyecto en sus primeros estadios (fue lanzado el 1 de abril de 2025), presentamos los primeros desafíos a los que nos enfrentamos y sus posibles respuestas: desde las dificultades en la gestión de fondos y el reclutamiento de personal en prácticas y contratado, hasta las carencias de

corpus de calidad en lenguas románicas frente a la abundancia de recursos anglosajones y la necesidad de ontologías y tesauri en el ámbito de nuestras ciencias, para llegar a señalar las prioridades de los técnicos que buscan robustez e interoperabilidad frente a los investigadores que reclaman herramientas fáciles de entender , bien organizadas, adaptables a una gran variedad de casos.

Unimos nuestra presentación a las que proponen nuestros compañeros en otros proyectos de financiación y alcance europeos como son Borja Navarro (Univ. de Alicante) por su participación en *ELTEC* (https://www.distant-reading.net/eltec/), Salvador Ros (UNED) con su experiencia en *CLS*-INFRA (https://clsinfra.io/), María Goicoechea (UCM) como investigador principal del DARIAH Project *Emerging Jobs Profiles in DH* (2023-2025). Queremos compartir nuestras experiencias sobre el terreno con la comunidad de HDH.

Bibliografía

AMIS (Advanced Metadata Intelligent System) (s. f.), recurso en línea. https://oscars-project.eu/projects/amis-advanced-metadata-intelligent-system

CONSORTIUM CAHIER, Groupe Typologies Textuelles (s. f.), recurso en línea. https://cahier.hypotheses.org/activites/groupe-typologie-textuelle

GALLERON, I., & IDMHAND, F. (2020b), «Why go from texts to data, or the Digital Humanities as a critique of the Humanities», *Word and Text: A Journal of Literary Studies and Linguistics*, X, pp. 53–69. ⟨halshs-03951355⟩

IDMHAND, F., & GALLERON, I. (2020a), *Guide pour la FAIRisation des données des corpus d'auteurs* [Groupe de travail Data_Cahier], Huma-Num. ⟨halshs-02889777⟩

MARTÍN-RODILLA, P., & GONZÁLEZ-PÉREZ, C. (2017), «A modelling language for discourse analysis in Humanities: Definition, design, validation and first experiences», *Revista de Humanidades Digitales*, 1, pp. 368–378. https://doi.org/10.5944/rhd.vol.1.2017.16133

DraMÆ (base de datos, estrategias dramatúrgicas y motivos): una nueva herramienta digital para el estudio del teatro de la memoria democrática en España

Jessica De Matteis
Università Roma Tre – Universitat de València

Palabras clave: Base de datos, teatro de la memoria, estrategias dramatúrgicas, teatro español contemporáneo, herramienta digital
Keywords: Database, memory theater, dramaturgical strategies, contemporary Spanish theater, digital tool

Introducción

El uso de bases de datos en las Humanidades Digitales ha transformado radicalmente la manera en que los investigadores recopilamos, organizamos, analizamos y compartimos información. Las bases de datos no solo sirven para conservar datos, sino que se convierten en herramientas de investigación activa, permitiendo realizar consultas complejas, combinar datos, generar modelos interpretativos y crear recursos duraderos, accesibles y reutilizables. En este sentido, no estamos ante *archivos estáticos* —si entendemos este término en el marco de las reflexiones en torno a los conceptos de archivo y repertorio en el ámbito de la memoria cultural, según Jan Assmann—, sino ante plataformas interactivas e interoperables que posibilitan nuevas formas de explorar, narrar y transformar el patrimonio cultural desde una perspectiva innovadora y dinámica.

Las Humanidades Digitales han superado el paradigma del documento lineal —pensemos, en este sentido, en los textos digitales— para adoptar modelos basados en bases de datos relacionales: los datos «atomizados» se refieren mediante claves a diferentes entidades (personas, lugares, acontecimientos), lo que permite consultas complejas y nuevas perspectivas.

En el ámbito de las Humanidades Digitales, la creación de bases de datos no se limita a la recopilación y conservación de información, sino que exige el cumplimiento de una serie de estándares internacionales que garanticen la interoperabilidad, la sostenibilidad, la accesibilidad y la reutilización de los datos a largo plazo, promoviendo un enfoque colaborativo y abierto en el tratamiento de los datos humanísticos, respaldado por el principio FAIR (Findable, Accessible, Interoperable, Reusable).

Estos estándares abarcan aspectos técnicos (como el uso de formatos abiertos y estructurados: XML, TEI, RDF), metodológicos (documentación

detallada, ontologías compartidas, metadatos normalizados) y éticos (acceso abierto, licencias claras, respeto a los derechos de autor y a la diversidad cultural). La adhesión a estos marcos comunes no solo permite compartir datos entre diferentes plataformas y disciplinas, sino que potencia nuevas formas de investigación crítica y creativa, haciendo de la base de datos un verdadero entorno de conocimiento activo.

Una base de datos relacional se estructura en torno al concepto de *tabla* (o relación), en la que los datos se organizan en filas (tuplas) y columnas (atributos). Cada tabla representa una entidad o concepto (por ejemplo, obras, autores, eventos), y las relaciones entre diferentes tablas se definen mediante claves primarias y claves foráneas (*primary key* y *foreign key*), lo que permite establecer conexiones lógicas entre los datos.

En el contexto de las Humanidades Digitales, una base de datos relacional bien diseñada debe cumplir con los siguientes criterios:

- Normalización de datos: evitar redundancias y asegurar la coherencia interna.
- Integridad referencial: garantizar que las relaciones entre tablas sean consistentes.
- Escalabilidad y sostenibilidad: facilitar la ampliación futura del sistema.
- Acceso mediante lenguaje SQL (Structured Query Language): para realizar consultas complejas y generar resultados significativos.
- Documentación clara del esquema: uso de diccionarios de datos, diagramas entidad-relación (ER) y modelos conceptuales.
- Exportación en formatos interoperables: como CSV, XML, o JSON, para facilitar el intercambio de datos con otras plataformas o herramientas analíticas.

La Base de datos DraMÆ

Pasando a nuestra base de datos, DraMÆ (acrónimo de Dramaturgia de la Memoria Actual en España) es pensada como una herramienta digital para el estudio del corpus de piezas que abarcan el tema de la memoria en España, centrándonos especialmente en la memoria de la II República, de la guerra civil y del franquismo, de la transición. El corpus que estamos recopilando incluye textos escritos a partir del año 2000. Se trata de un corpus en continua evolución y, por consiguiente, de una base de datos en construcción y, sin embargo, ya funcional para los objetivos que se propone.

La ideación y realización de esta base de datos surge del trabajo desarrollado por el grupo de investigación *Herencias – scritture di memoria e identità* de la Università Roma Tre, en el marco de su prolongada labor de recopilación, investigación y análisis de textos dramáticos de la memoria.

Su implementación concreta ha sido posible, sin embargo, solo más recientemente, gracias a la financiación obtenida con el Proyecto PRIN 2022 *Identity Heritage and Cultural Memory. The elaboration of the past through the*

theatre of democratic Spain (1975 to the present) (Prot. 202285ZT47 - CUP D53D23014880006), financiado por la Unión Europea - Next Generation EU, Mission 4, Component 2, Investment 1.1 *Research Projects of Significant National Interest*, bajo la supervisión de Simone Trecca, investigador principal del grupo Herencias y autor correspondiente del proyecto en Roma Tre. El desarrollo de esta base de datos por parte del equipo de la Università Roma Tre se inscribe entonces dentro de un proyecto más amplio que incluye otros centros (Universidades de Turín, Venecia Ca' Foscari coord. Veronica Orazi y Universidad de Pisa, coord. Enrico Di Pastena), cuyo IP es Veronica Orazi (Università Ca' Foscari, Venecia).

DraMÆ está dedicada a la memoria de Guillermo Heras y Jerónimo López Mozo, quienes generosamente contribuyeron, durante años, a la definición del corpus de obras de teatro de la memoria democrática por parte del equipo del proyecto Herencias.

La base de datos está concebida para distintos niveles de consulta y análisis, y está dirigida tanto al ámbito académico como profesional. Entre los principales tipos de uso, destacan:

- Uso informativo: permite localizar fácilmente textos teatrales que abordan el tema de la memoria, proporcionando información básica sobre autoría, sinopsis, año de publicación, y temas tratados.
- Uso dramatológico: ofrece un marco de análisis estructurado sobre las coordenadas espacio-temporales, niveles de dramatización, personajes y formas de representación del pasado en escena, útil para investigadores, docentes y estudiantes de teatro.
- Explotación cuantitativa de los datos: la organización de la información en categorías estructuradas permite realizar análisis estadísticos, detectar tendencias y visualizar la evolución de ciertos motivos o enfoques en el teatro de la memoria.
- Aplicación profesional: la base puede servir como herramienta para dramaturgos, directores, compañías teatrales y programadores interesados en incorporar textos que dialoguen con la memoria histórica, ofreciendo criterios comparativos y descriptivos claros.

Metodología

La creación de esta base de datos especializada en textos teatrales españoles relacionados con la memoria está siguiendo un proceso estructurado en varias fases. El trabajo combina herramientas propias de la dramatología con categorías específicas que el grupo de investigación ha construido para el análisis del teatro de la memoria.

La primera fase consiste en la identificación, recopilación y adquisición de obras teatrales que abordan explícitamente temas vinculados con la memoria española. Este proceso ha implicado la elaboración de un corpus inicial de títulos representativos, considerando criterios temáticos (memoria del exilio, de la

infancia, de las mujeres, etc.) y verificando su disponibilidad en ediciones impresas o digitales. Dado que algunas de estas obras aún no han sido publicadas, resulta imprescindible contar con la colaboración generosa de dramaturgos y dramaturgas, quienes están facilitando al grupo el acceso a textos inéditos para su estudio.

Una vez recopiladas las obras, se ha realizado un análisis dramatológico detallado de cada texto, basándose en textos teóricos como el de García Barrientos (2001). Este análisis no se limita al contenido temático, sino que aborda aspectos formales y estructurales clave para el teatro de la memoria.

Se analizan particularmente los personajes principales que activan o representan una forma específica de memoria —como el personaje homosexual, el niño/a, el maestro/a republicano/a, el ejecutado/a, etc.— y se investiga cómo el pasado se reactiva en escena a través de sus voces y acciones.

Además, entre las categorías fundamentales estudiadas se encuentra el tiempo, esencial en el teatro de la memoria. De hecho, este tipo de teatro, al buscar recuperar el pasado desde el presente, establece con frecuencia un diálogo entre distintos momentos temporales. Este diálogo puede adoptar diversas formas: alternancia de planos temporales, superposición o incluso fragmentación cronológica.

Otra categoría clave es el espacio, tanto desde una dimensión geográfica concreta (memorias localizadas en contextos como Cataluña, Euskadi, Galicia, etc.), como desde la representación de lugares de memoria (los Hogares de Auxilio Social, cunetas, campos de concentración, etc.).

Una vez analizadas las categorías espacio-temporales, se procede al estudio de los niveles de dramatización, entendidos como las coordenadas espacio-temporales que se representan efectivamente en escena y que generan los campos de fuerzas que vertebran las obras.

Resultados

El corpus de obras analizadas es amplio y en continua expansión, por lo que los resultados actuales deben entenderse como una primera fase de trabajo dentro de un proyecto en desarrollo. A pesar de ello, ya es posible obtener algunos datos reveladores.

En cuanto a la autoría, si consideramos el número total de personas que han escrito las obras del corpus:

- 36,5 % son autoras
- 63,5 % son autores

Sin embargo, si analizamos la autoría desde el punto de vista de las obras (y no de los individuos) —es decir, considerando cada pieza y su atribución autoral, incluyendo aquellas escritas en colaboración— los porcentajes cambian:

- 48 % de las obras tienen autoría femenina
- 55 % de las obras tienen autoría masculina (*estos porcentajes superan el 100% porque algunas obras tienen coautoría mixta*).

Este dato sugiere una tendencia interesante: aunque hay más autores que autoras en términos absolutos, las mujeres escriben proporcionalmente más obras sobre la memoria, lo que indica una participación especialmente activa en este campo temático.

Además, a partir del análisis del corpus recopilado, ha sido posible reconocer una serie de motivos temáticos que se repiten de forma significativa, así como otros de aparición más reciente, que reflejan la evolución y la diversificación del teatro de la memoria en España. Entre estos motivos emergentes se encuentran la resistencia antifranquista, el tardofranquismo, la Transición democrática, así como temas como las fosas comunes y cunetas, la memoria de la infancia, la represión de las disidencias sexuales, el exilio en el norte de África, las memorias familiares y el rescate escénico de figuras históricas reales. Estos nuevos enfoques responden tanto a un cambio generacional como a un contexto político-cultural que ha favorecido la pluralización de las memorias posibles.

Bibliografía

ALEXANDER, J. C. (2004), *Cultural Trauma and Collective Identity*, University of California Press.

ASSMANN, A. (2015), *Ricordare. Forme e mutamenti della memoria culturale*, Bologna, Il Mulino.

ASSMANN, J. (1997), *La memoria culturale. Scrittura, ricordo e identità politica nelle grandi civiltà antiche*, trad. F. De Angelis, Torino, Einaudi.

GARCÍA BARRIENTOS, J. L. (2001), *Cómo se comenta una obra de teatro*, Madrid, Síntesis.

HALBWACHS, M. (2024), *I quadri sociali della memoria*, eds. T. Grande & L. Migliorati, Roma, Meltemi.

LACAPRA, D. (2005), *Escribir la historia, escribir el trauma*, trad. H. Pons, Buenos Aires, Nueva Visión.

ORAZI, V. (coord.) (2025), monográfico «50 Years since the Transition. Memorial Patterns of the Spanish State Theatre (1975 to the Present)», *eHumanista IVITRA*, vol. 27. https://www.ehumanista.ucsb.edu/ivitra/volume/27

SCHREIBMAN, S., SIEMENS, R., & UNSWORTH, J. (eds.) (2015), *A New Companion to Digital Humanities*, Malden (MA), Wiley-Blackwell.

TRECCA, S. (coord.) (2019), monográfico «Los escenarios de la postmemoria en el teatro hispánico último (2000–2018)», *Orillas: Rivista d'Ispanistica*, vol. 8.

TRECCA, S. (2023), «El teatro español y la construcción cultural del trauma. Notas sobre la dramaturgia de la memoria actual», *Anales de la Literatura Española Contemporánea*, 48(2), pp. 139–163.

Hacia un paradigma dataísta de la memoria. Potencialidades de MNLAB en el almacenamiento, tratamiento y análisis de datos sobre la literatura de la memoria en España

José Martínez Rubio
Luz C. Souto
Universitat de València

Palabras clave: Humanidades Digitales; visualización de datos; Base de datos MNLAB; Guerra Civil española; literatura española
Keywords: Digital Humanities; data visualization; MNLAB database; Spanish Civil War; Spanish literature

Partiendo de la premisa que encabeza el *VII Congreso de la Sociedad Internacional de HUMANIDADES DIGITALES HISPÁNICAS, "Compromiso cívico y social de las Humanidades Digitales"*, la presente ponencia tiene como objetivo presentar el proyecto *Memory Novels Lab* (MNLAB), una base de datos interactiva y multilingüe dedicada al almacenamiento, tratamiento y análisis de datos sobre la narrativa española de la Guerra Civil y el franquismo.

La Base de Datos MNLAB, desarrollada por un equipo interdisciplinar y coordinada desde el Departamento de Filología Española de la Universitat de València, ofrece no solo un archivo ampliado y dinámico de obras literarias vinculadas a la memoria histórica, sino también una infraestructura conceptual, técnica así como un conjunto de datos que permite nuevas interpretaciones, particulares o de conjunto, sobre la narrativa del pasado reciente. En un contexto en el que la memoria democrática constituye un eje crucial de debate público y de intervención política, el proyecto MNLAB se presenta como una herramienta estratégica desde las Humanidades Digitales para democratizar el conocimiento, fomentar la investigación crítica, potenciar la capacidad de comprensión del fenómeno memorialístico y articular procesos de reparación simbólica y cultural.

En la primera parte de la intervención se introduce la relación entre las Humanidades Digitales y archivo literario. Las Humanidades Digitales se han consolidado en las últimas décadas como un campo interdisciplinar capaz de reformular los modos tradicionales de producción, conservación y circulación del conocimiento en las ciencias humanas. Este ámbito no solo supone una aplicación tecnológica sobre objetos culturales, sino una transformación epistemológica en el modo en que se construyen y comparten las humanidades. En este marco, el desarrollo de bases de datos literarias con capacidad analítica

se convierte en una práctica crítica y política que va más allá del mero almacenamiento documental, permitiendo investigaciones de conjunto con un alcance mucho más ambicioso y refinado.

En el caso español, esta transformación cobra especial relevancia cuando se cruza con el campo de los estudios de la memoria, donde el trabajo de archivo y el trabajo con datos permiten visibilizar voces silenciadas, indagar sobre las condiciones de generación de cánones culturales, posibilitar análisis de temáticas particulares, establecer comparativas en el uso de estrategias, recursos o temáticas de distintas generaciones literarias (Oleza, 2023), esclarecer hipótesis sobre las representaciones de las víctimas o victimarios que trascienden a debates éticos en la esfera internacional (Todorov, 1995; Filkenstein, 2001; Huyssen, 2002) y, en definitiva, responder desde el conocimiento académico a una demanda social y ética de la memoria. En este contexto, la investigación que atañe a los compromisos cívicos con el pasado y su conexión con las Humanidades Digitales permite darle forma al "trabajo de la memoria", es decir, "incorporar [la memoria] al quehacer que genera y transforma el mundo social" (Jelin, 2002: 14), sumar fuerza al impulso colectivo por dar sentido, justicia y visibilidad al pasado traumático.

La segunda parte de la intervención se centrará en la actividad del *Laboratorio de literatura de la memoria MNLAB*, cuya "base de datos de novelas sobre memoria histórica española" surge como respuesta a esta intersección entre memoria, literatura y tecnologías digitales. En su fase actual (2021–2025), el proyecto ha almacenado y etiquetado los datos principales de 1876 novelas publicadas entre 1990 y 2025, coincidiendo con la gestación y la consolidación del llamado boom de la memoria en la literatura española contemporánea, y ha analizado los componentes extraliterarios e intraliterarios de 287 de ellas.

Los objetivos principales de la Base de Datos pueden resumirse en:

1. Explorar, clasificar, preservar y difundir la narrativa sobre la Guerra Civil y el franquismo.
2. Crear un archivo digital abierto que permita búsquedas avanzadas por criterios temáticos, lingüísticos, temporales, genéricos y editoriales.
3. Activar nuevas formas de análisis del corpus narrativo que permitan detectar mutaciones temáticas, patrones de representación y zonas de silencio en el campo de la memoria.

El proyecto, financiado por convocatorias públicas de I+D+i (GV/2021/183, CIGE/2023/74, UV-INV-AE 3672384), cuenta con un equipo multilingüe e interdisciplinar que articula saberes provenientes de la filología, la historiografía, la biblioteconomía y la informática. El enfoque iberístico del proyecto garantiza la inclusión de novelas escritas en castellano, catalán, gallego, asturiano y euskera, así como un análisis comparado con obras en inglés, francés, alemán, italiano y portugués que abordan el conflicto desde una perspectiva internacional.

En cuanto a la metodología y explotación de datos, cabe señalar que uno de los valores añadidos de MNLAB reside en su capacidad para organizar el conocimiento literario como dato analizable, sin renunciar a la complejidad

hermenéutica propia del discurso literario. Los investigadores trabajan en la base de datos con fichas individualizadas para cada obra, con metadatos estructurados que permiten búsquedas cruzadas y análisis relacionales por: género del autor o autora, año de nacimiento, años de publicación, lengua original, estrategias narrativas, género y subgénero literario, líneas temáticas, tiempo de la narración vs. tiempo de la acción, topónimos, premios recibidos, tipología de personajes, etc.

Estos parámetros no solo ordenan el archivo, sino que activan nuevas preguntas críticas: ¿qué generaciones han producido más literatura de la memoria? ¿Cuál ha sido el papel de las mujeres en la configuración de este relato? ¿Cómo varían las representaciones según la lengua o el territorio? ¿Qué zonas del conflicto permanecen sin narrar o narradas deficientemente frente a polos geográficos predominantes?

La visualización de los datos (gráficos, mapas, líneas de tiempo) permite acceder de manera intuitiva a estos resultados. A modo de ejemplo, se aportan a continuación dos gráficos: El primero (Figura 1), se trata de un gráfico de líneas comparativo en el que se representa la evolución de las publicaciones en torno a la memoria histórica española, desde 1990 a 2022, periodo en el que ya se han volcado los corpus completos. La línea azul evidencia la producción masculina y la violeta las obras escritas por mujeres.

El segundo gráfico (Figura 2) es circular y arroja los porcentajes de la escritura masculina y femenina sobre la memoria, teniendo en cuenta que de un corpus de 1876 obras introducidas en la BBDD, solo 453 están escritas por mujeres, es decir, el 24,1 %.

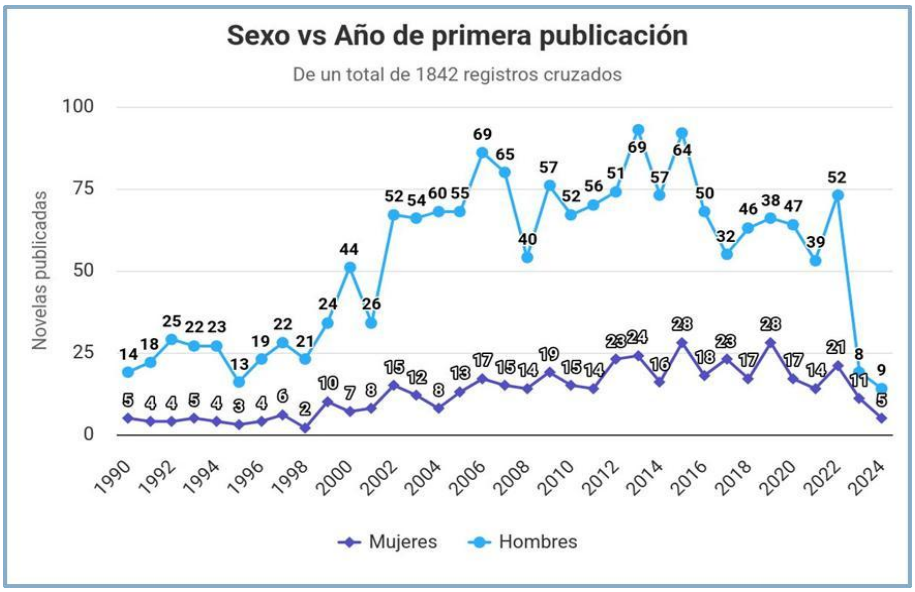


Figura 1. Datos preliminares de la evolución de las narrativas de la memoria desde 1990 a 2022, atendiendo a un criterio de género

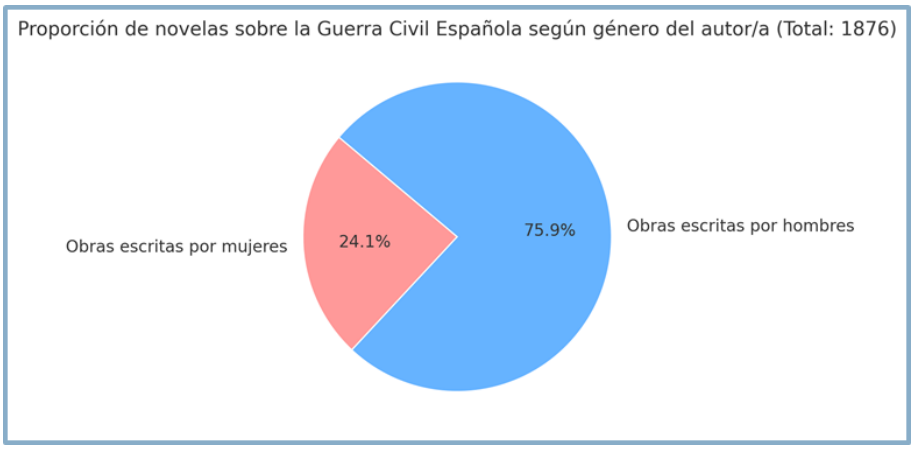


Figura 2. Datos preliminares de proporción de novelas sobre la Guerra Civil española escrita por hombres y mujeres

Los resultados que se pueden presentar son preliminares, y responden al trabajo en marcha que está realizando el equipo de MNLAB con una planificación pormenorizada para los próximos dos años (2025 y 2026). De entre estos resultados, podremos lanzar hipótesis muy plausibles en torno a: la configuración de género en el fenómeno literario de la memoria y la primacía de las voces masculinas en esta tendencia literaria; el comportamiento editorial ante el fenómeno literario; el papel de las lenguas cooficiales y su evolución ante el interés por la memoria; la respuesta de la industria literaria o de la esfera institucional validando mediante premios y galardones este tipo de literatura; las distintas representaciones que se han dado de forma diacrónica en los últimos 35 años de memoria.

En la ponencia se incidirá, asimismo, en los aportes a la memoria democrática que proporciona el desarrollo de bases de datos literarias abiertas como MNLAB y la capacidad de la transferencia social del conocimiento. Entre las acciones desarrolladas se explicará, por un lado, la colaboración con la Biblioteca de Humanidades Joan Reglà para enriquecer el catálogo con metadatos MNLAB. Por otro lado, se explicará el proyecto de creación de vídeo-reseñas de novelas escritas por mujeres que se vincula a la BBDD, una apuesta a la participación estudiantil nacional e internacional que busca la trasmisión intergeneracional en la memoria y la formación de los y las estudiantes en el manejo de herramientas digitales para la investigación.

Finalmente, la intervención analizará de qué modo el proyecto MNLAB ha respondido a los tres objetivos principales de las Humanidades Digitales:

“1. Crear bases de datos con recursos digitales relevantes para las Humanidades. Esto incluye la captura, estructuración, documentación, preservación y diseminación de los datos. 2. Desarrollar metodologías que permitan generar nuevos elementos derivados de estos datos. 3. Generar investigación y conocimiento para incrementar nuestra comprensión en las Humanidades” (Galina Russell, 2011: 3).

Bibliografía

FINKELSTEIN, N. (2001), *The Holocaust Industry: Reflections on the Exploitation of Jewish Suffering*, London, Verso.

GALINA RUSSELL, I. (2011), «¿Qué son las Humanidades Digitales?», *Revista Digital Universitaria*, 12(7).

HUYSSEN, A. (2002), *En busca del futuro perdido. Cultura y memoria en tiempos de globalización*, México, Fondo de Cultura Económica.

JELIN, E. (2002), *Los trabajos de la memoria*, Madrid, Siglo XXI.

OLEZA, J. (2023), *Claves ibéricas de la Guerra Civil: memorias y narrativas*, Sevilla, Renacimiento.

TODOROV, T. (1995), *Les abus de la mémoire*, París, Arléa.

GEXEL Digital: nuevos enfoques sobre la literatura del exilio republicano español de 1939

Lucía Cotarelo Esteban
Universidad Autónoma de Barcelona

Palabras clave: GEXEL Digital; Exilio republicano; bases de datos; visualización de redes sociales
Keywords: GEXEL Digital; Republican exile; databases; social media visualization

Durante la última década, la revisión crítica del panorama literario español del siglo XX y consecuente renovación de los estudios de la literatura del exilio republicano de 1939 ha sido profunda. Por un lado, se ha expandido el objeto de estudio de la literatura del periodo a través del interés por una producción hasta el momento marginal. Esto es: la otredad dentro de la otredad que es el exilio, si la consideramos desde el punto de vista de la Historia de la Literatura Española más oficial que siempre opera, en tanto que heredera del canon nacional franquista, con afán organizativo reduccionista y sin dar espacio a las complejas "líneas de fuga" que el fenómeno exiliar produce (Balibrea, 2017).

Por otro lado, se está realizando un importante esfuerzo teórico para superar esta misma historia oficial también en sus métodos desde, por ejemplo, los estudios culturales (Balibrea, 2017), sobre emociones (Rodríguez López y Ventura Herranz, 2014) y de redes sociales (Sánchez de Madariaga, 2017; Cotarelo Esteban, 2022). Algunas de las nuevas propuestas trazan metodologías para la organización historiográfica que sobrepasan los tradicionales sistemas cronológicos o generacionales en favor de sistemas basados en categorías conceptuales específicas, eventos internacionales y relaciones transatlánticas.

Creemos que la historiografía literaria, como método analítico, debería estar abierta a la incorporación de nuevas técnicas que permitan abarcar un objeto de estudio mayor – visibilizando colectivos, géneros, etc.–. Asimismo, que esta incorporación debería llevarse a cabo desde la perspectiva ecléctica y de fluidez transdisciplinar (Max-Neef, 2004) –enfoque holístico que supere las disciplinas estancas y la incomunicación entre academia y sociedad– propia de los estudios culturales, que ofrezca nuevas teorías y metodologías desde las que releer los periodos literarios. Contar con nuevas materias de estudio y, en consecuencia, con nuevos modos de trabajo adaptados a ellas, supone necesariamente comenzar a generar procesos de conocimiento también nuevos. ¿Qué papel pueden jugar los medios digitales en esta transformación? ¿Qué sería interesante implementar a la hora de abordar el exilio republicano, tanto general como específicamente literario, desde las humanidades digitales?

Siguiendo las propuestas de Bocanegra (2017) y Fèrriz Roura (2017), los proyectos de humanidades digitales, planteados y creados por equipos transdisciplinares son nuestra apuesta para encontrar una metodología y un espacio de reflexión y divulgación adaptados a nuestro objeto de estudio. Este enfoque nos resulta básico a la hora de empezar a plantear un proyecto de humanidades digitales sobre el exilio republicano español, materia en sí misma de naturaleza híbrida, atravesada por factores sociales, culturales, políticos, históricos, sociológicos, antropológicos, filosóficos, etc. Nos resulta imprescindible que, además, estos proyectos partan de una base ética de restitución, conservación y difusión del patrimonio cultural. También lo es que se configuren como laboratorios para la experimentación de los investigadores especializados y que, para ello, sean creadores de productos digitales.

Concretamente, abogamos por webs desarrolladas sobre bases de datos relacionales cuyos datos sean interoperables y procedan de fuentes diversas –bibliotecas y repositorios digitales, y la investigación realizada por especialistas, quienes siempre deben de estar detrás de su curación–, de las que beban colecciones específicas de objetos diseñadas a través del modelado de metadatos. Deben ser bases extensibles –que sigan ampliándose mediante las nuevas incorporaciones–, cuyos datos se estructuren en base a una forma estándar de expresión –normalizadas– que permitan la interoperatividad, el cruce de datos (Romero, Bueren y Gayoso, 2017). Deben, asimismo, estar implementadas con motores de búsqueda avanzados que permitan explotar toda esa información obteniendo resultados novedosos, y mapas de información que permitan visualizar y explorar el contenido de forma alternativa y no lineal.

Como respuesta a estas necesidades nace GEXEL Digital (https://gexel.graphery.com/welcome/), una nueva propuesta para la historiografía literaria del exilio republicano de 1939 atenta a todos los requisitos mencionados que se desarrolla actualmente en el seno del Grupo de Estudios del Exilio Literario (GEXEL) de la Universidad Autónoma de Barcelona. Parte de la ingente labor desarrollada durante años por el equipo para la concepción del *Diccionario biobibliográfico de los escritores, editoriales y revistas del exilio republicano de 1939* (Aznar Soler y López García, 2016, cuatro tomos) y, de esta manera, GEXEL Digital se alimenta íntegramente de la base de datos sobre el exilio republicano literario más amplia que existe.

Los metadatos en los que se estructura su información, combinados, permiten realizar búsquedas complejas. GEXEL Digital ha sido financiada por un proyecto del Ministerio de Ciencia y Educación y otro de la Generalitat de Catalunya (SGR) y está desarrollada por un equipo de filólogos especialistas en humanidades digitales e informáticos que son, asimismo, humanistas en distintas áreas. Se ha aprovechado la información del mencionado *Diccionario* y, a través de él, se ha desarrollado una base de datos programada en PostgreSQL que ya recoge 1.687 entradas.

Esta base relacional ha dado lugar a dos tipos de visualizaciones o itinerarios de lectura (una primera lineal y textual y una segunda abierta,

construida sobre hipervínculos en forma de red, un mapa relacional) y espera dar lugar a una tercera muy pronto (visualización de mapas geográficos). El Front se encuentra programado con HTML y el entorno Node.JS. Los metadatos iniciales de las entradas "persona", más sencillos de deducir, han sido extraídos de modo automatizado mediante chatGPT en formato JSON: tipo de elemento, nombre, apellidos, sexo, profesión, nacimiento (país, fecha, comunidad, municipio o provincia de nacimiento, defunción y exilio). La red dinámica de la segunda visualización mencionada ha sido programada desde observablehq.com.

Los metadatos del exilio que estructuran la información de GEXEL Digital aspiran a conformar una propuesta metodológica en sí misma, en tanto que: (1) son propuestos como resultado de una revisión bibliográfica e historiográfica del periodo; (2) tratan de dar cuenta de las necesidades específicas de una historiografía que debe atender a las discontinuidades y la heterogeneidad propias del fenómeno del exilio; (3) pretenden superar las categorías canónicas en que se ha estructurado tradicionalmente el conocimiento y abordar cuestiones transversales propias de los estudios culturales; (4) y orientan al investigador y al interesado, en fin, hacia una lectura guiada a través de cuestiones que, desde la perspectiva de un especialista, se consideran relevantes.

El primer tipo de visualización se alcanza a través de "Explora todo el contenido", y permite el acceso a entradas de tres categorías que podemos filtrar: personas, publicaciones y editoriales. Se han implementado unos primeros metadatos para personas: sexo, géneros y área profesional. A estos metadatos se espera añadir otros específicos del exilio republicano como son "Paso por campos de concentración", "Afiliación política" o "Regresos". Cada entrada ofrece metadatos a la izquierda, texto en el centro y un navegador por el texto a la derecha: ofrece, de entrada, dos formas de acceder a la información: esquemática y rápida (metadatos) y pausada y más apegada a lo analógico (texto). Entre los metadatos encontramos dos imprescindibles: "Enlaces externos", que nos llevan a las bases de datos de PARES y BNE; y "Redactores", al que próximamente acompañará una identificación DOI.

El segundo tipo de visualización sigue una concepción de "constelaciones culturales" y se realiza a través del menú "Explorar red". Aparece un panel con todos los autores ordenados alfabéticamente, modelo que supone un índice enriquecido que aporta una visión con valor propio. A la derecha, la posibilidad de aplicar, de momento, un único filtro (por sexo) y los distintos tipos de relación que identificamos, categorías basadas en expresiones regulares contenidas en PARES: amistad, familiar, colaboración, pareja/matrimonio, otras relaciones y aprendizaje. Esta red se ha confeccionado mediante los datos de PARES que bebieron, asimismo, del *Diccionario* de GEXEL.

La página dispone de tres perfiles (visitante, editor y administrador) y posibilita: (1) consultas abiertas para cualquier usuario visitante, no registrado; y (2) acceso a los editores al panel de edición a través de un login controlado para que un equipo reducido del grupo GEXEL pueda introducir información, mejorar,

ampliar o corregir. Pasando cualquier entrada al “Modo edición”, algo posible para editores, todos los elementos son editables de un modo sencillo con un panel similar a Word: nombre, fechas, texto, metadatos, datos sobre las obras… Asimismo, podemos introducir nuevos elementos: nuevas obras, prólogos, bibliografía… de este modo, GEXEL Digital se convierte en un modelo extensible y actualizable: una base de datos en constante crecimiento y renovación.

Con la creación de GEXEL Digital abogamos, en fin, por webs desarrolladas sobre bases de datos relacionales cuyos datos procedan de fuentes diversas, bases extensibles, normalizadas, implementadas con motores de búsqueda avanzados y con mapas de información. Se superan así la perspectiva eminentemente nacional, generacional y cronológica de las historiografías tradicionales, y se proponen modos específicos – interdisciplinares, internacionales, relacionales– para acceder a la información sobre el exilio.

Bibliografía

AZNAR SOLER, M., & LÓPEZ GARCÍA, J. R. (coords.) (2016), *Diccionario biobibliográfico de los escritores, editoriales y revistas del exilio republicano de 1939*, Sevilla, Renacimiento.

BALIBREA, M. P. (coord.) (2017), *Líneas de fuga. Hacia otra historiografía cultural del exilio republicano español*, Madrid, Siglo XXI de España Editores.

BOCANEGRA BARBECHO, L. (2017), «Identidad y pertenencia del exilio republicano español en la red internet: generando memoria y cultura a través de las plataformas sociales digitales y el uso del Open Data», en C. OLIVIERI & A. ORTEGA SANTOS (eds.), *Decolonizando identidades. Pertenencia y rechazo de/desde el Sur Global*, Granada, Instituto de Migraciones, pp. 211–221.

COTARELO ESTEBAN, L. (2022), «Mujeres intelectuales españolas y sus redes sociales en el exilio estadounidense», *Journal of Spanish Cultural Studies*, 23(1), pp. 13–33.

FÈRRIZ ROURA, T. (2017), «Redes del exilio de 1939 en internet», *TSN*, 4 (julio–diciembre), pp. 59–63.

GEXEL DIGITAL (s. f.), recurso en línea. https://gexel.graphery.com/welcome/

MAX-NEEF, M. A. (2004), *Fundamentos de la transdisciplinaridad*, Valdivia, Universidad Austral de Chile.

RODRÍGUEZ-LÓPEZ, C., & VENTURA HERRANZ, D. (2014), «De exilios y emociones», *Cuadernos de Historia Contemporánea*, 36, pp. 113–138.

ROMERO LÓPEZ, D., BUEREN GÓMEZ-ACEBO, J. L., & GAYOSO CABADA, J. (2017), «Modeling *Colecciones de Literatura de Quiosco* for Mnemosine Digital Library», en J. ZAMOSTNY & S. LARSON (eds.), *Kiosk Literature of Silver Age Spain. Modernity and Mass Culture*, Bristol, Intellect, pp. 397–418.

SÁNCHEZ DE MADARIAGA, E. (2017), «Escritura epistolar y redes sociales. Pilar de Madariaga, Vassar College y el exilio», *Ayer*, 105(1), pp. 129–154.

Humanismo digital y lengua latina: el latín como espacio de reflexión crítica y metalingüística

Germán González Muñoz
IES Guillem D'Alcalà

Palabras clave: Latín; humanismo digital; metalingüística; enseñanza activa; tecnologías educativas.
Keywords: Latin; digital humanism; metalinguistics; active teaching; educational technologies.

En el contexto contemporáneo, profundamente marcado por la digitalización del conocimiento, la automatización de procesos cognitivos y la transformación del lenguaje mediada por tecnologías inteligentes, el papel de las lenguas clásicas, como el latín, puede parecer, a primera vista, marginal o incluso obsoleto. Sin embargo, esta comunicación defiende con vehemencia que el latín —lejos de ser una lengua muerta y meramente objeto de estudio filológico— ofrece un marco pedagógico y epistemológico privilegiado para el desarrollo de la conciencia metalingüística, la formación del pensamiento lógico-crítico y el ejercicio de una reflexión profunda y matizada sobre el lenguaje y la cultura en la sociedad actual. Argumentamos que, en un ecosistema comunicativo cada vez más dominado por la inmediatez, la fragmentación y la superficialidad, la estructura inherente y la riqueza semántica del latín proporcionan las herramientas necesarias para cultivar una comprensión más profunda de cómo funciona el lenguaje, cómo se construye el significado y cómo los discursos influyen en la percepción de la realidad.

Partimos de una concepción del humanismo digital que trasciende la visión instrumental de la tecnología como una mera herramienta para el estudio de las humanidades. En su lugar, lo entendemos como un espacio de diálogo epistemológico y ético fundamental entre la rica tradición humanística y la cultura tecnológica contemporánea. Este enfoque no solo busca adaptar las humanidades a los desafíos y oportunidades de la era digital, sino que también aspira a infundir los valores humanísticos —como la crítica, la reflexión, la contextualización y la ética— en el propio desarrollo y uso de las tecnologías. En este marco dialógico, el latín se presenta no solo como una lengua histórica, sino como un sistema lingüístico intrínsecamente estructurado, preciso y excepcionalmente formativo. Su estudio riguroso fomenta una comprensión profunda de la gramática, la sintaxis, la lógica y la etimología, habilidades cognitivas que son altamente transferibles y vitales para cualquier disciplina académica o profesional. Consecuentemente, su enseñanza, concebida desde

una perspectiva activa y comunicativa, permite repensar radicalmente las metodologías pedagógicas tradicionales y, de igual modo, reconsiderar la función esencial del lenguaje en la formación intelectual y ciudadana del siglo XXI. El objetivo es formar individuos capaces de discernir, analizar críticamente y crear significado en un entorno comunicativo cada vez más complejo y saturado de información.

Metodología: El método "Sequere me!" y la Integración Digital

Esta propuesta pedagógica se apoya en el método innovador "Sequere me!", una metodología integral desarrollada por el autor y aplicada con éxito durante varios años en Educación Secundaria Obligatoria y Bachillerato. Dicho método concibe el latín como una lengua viva y dinámica, superando la aproximación puramente gramatical-traductora que ha caracterizado tradicionalmente su enseñanza. En su lugar, "Sequere me!" emplea enfoques pedagógicos comunicativos, inductivos y contextuales, buscando replicar, en la medida de lo posible, los procesos naturales de adquisición de una lengua.

El método se estructura en tres fases interconectadas, diseñadas para construir progresivamente la competencia lingüística y metalingüística:

1. Input Comprensible en Contextos Significativos: Esta fase inicial se centra en la exposición abundante y significativa al latín. Los estudiantes no aprenden reglas aisladas, sino que se sumergen en textos y diálogos latinos cuidadosamente adaptados y graduados, presentados en contextos culturalmente relevantes. Se utilizan materiales auténticos o semi-auténticos (como adaptaciones de fábulas, mitos o pasajes históricos) acompañados de ayudas visuales, gestos y explicaciones en latín simplificado. El objetivo es que los alumnos comprendan el mensaje general sin necesidad de traducir cada palabra, desarrollando una intuición por la estructura y el significado. Recursos digitales como plataformas de aprendizaje online permiten la creación de bibliotecas de textos graduados con glosarios interactivos, grabaciones de audio de hablantes nativos o expertos, y vídeos con animaciones que ilustran el contenido.
2. Interacción Oral y Escrita Significativa: Una vez que los estudiantes han recibido suficiente input comprensible, se les anima a interactuar activamente con la lengua. Esta fase implica actividades comunicativas donde el latín se utiliza como medio real de expresión. Esto incluye diálogos guiados, role-playing sobre situaciones cotidianas o históricas, debates sencillos, descripciones de imágenes, y ejercicios de preguntas y respuestas. En el ámbito escrito, se fomenta la redacción de frases, párrafos o incluso cartas cortas, siempre con un propósito comunicativo claro. Las herramientas digitales facilitan la interacción asincrónica a través de foros de discusión en latín, herramientas colaborativas de escritura (wikis) y plataformas de videoconferencia

para prácticas orales. La gamificación se integra aquí a través de desafíos de comunicación y recompensas por la participación activa.

3. Producción Creativa y Reflexiva: La fase final busca que los estudiantes no solo reproduzcan, sino que generen contenido original en latín, aplicando su comprensión y desarrollando su propia voz y pensamiento crítico. Esto puede incluir la composición de pequeñas historias, poemas, diálogos originales, o incluso la adaptación de textos modernos al latín. Crucialmente, esta fase también incorpora la reflexión metalingüística: los alumnos analizan las estructuras gramaticales, las elecciones léxicas y las figuras retóricas en los textos latinos, comparándolas con su propia lengua y reflexionando sobre el impacto del lenguaje en el pensamiento. Aquí es donde la inteligencia artificial generativa (IAG) juega un papel innovador. La IAG se utiliza para crear ejercicios personalizados, generar variaciones de textos para diferentes niveles, o incluso para simular conversaciones en latín, proporcionando un entorno de práctica ilimitado. Sin embargo, esta integración tecnológica no posee un valor meramente instrumental; se enmarca dentro de una estrategia de alfabetización crítica digital en la que los estudiantes no solo consumen información generada por IA, sino que son capacitados para cuestionar activamente los usos del lenguaje en el ámbito digital, analizar críticamente los discursos mediáticos predominantes y comprender las estructuras de poder que los sustentan. El latín, con su rigor gramatical, su rica tradición retórica y su historia como lengua de poder y conocimiento, proporciona un marco incomparable para desentrañar las complejidades del lenguaje en la era de la información.

Resultados y conclusiones

Los resultados preliminares, observados tras varios años de aplicación sistemática del método “Sequere me!” en la Educación Secundaria Obligatoria y Bachillerato, son notablemente positivos y alentadores, respaldando la tesis central de esta comunicación.

En primer lugar, se ha constatado una mejora significativa en la motivación del alumnado. Los estudiantes, liberados de la memorización árida, perciben el latín como una materia relevante, dinámica y aplicable, no como una carga académica distante. La aproximación comunicativa y el uso de tecnologías lo hacen más accesible y atractivo, transformando la experiencia de aprendizaje en un proceso de descubrimiento activo.

En segundo lugar, se ha observado un incremento tangible en su capacidad de análisis lingüístico. Al comprender el latín desde una perspectiva funcional y estructural, los alumnos desarrollan una agudeza metalingüística que les permite abordar cualquier idioma —incluido el español— con mayor perspicacia, identificando patrones, comprendiendo la formación de palabras y

desentrañando la lógica subyacente a la gramática. Esta habilidad se transfiere a otras áreas del conocimiento, mejorando su capacidad de razonamiento abstracto y resolución de problemas.

En tercer lugar, y quizás lo más importante, se ha producido una transformación profunda en la percepción del latín. De ser considerada una lengua "muerta" y relegada a un nicho académico, el latín transita a ser un espacio vibrante de pensamiento y creación, una herramienta viva para la reflexión crítica y la expresión personal. Los estudiantes no solo aprenden sobre latín, sino que aprenden *a través* del latín.

Finalmente, en el profesorado, se ha evidenciado una marcada apertura hacia metodologías activas y un uso más reflexivo y pedagógicamente consciente de la tecnología en el aula. Se ha superado la resistencia inicial a la innovación, reconociendo la tecnología no como un mero sustituto del contacto humano o de la enseñanza tradicional, sino como un aliado estratégico para potenciar el aprendizaje, personalizar la instrucción y fomentar la autonomía del estudiante.

En conclusión, este trabajo subraya con rotundidad el valor inestimable del latín como herramienta metalingüística y formativa esencial en el entorno digital contemporáneo, así como su potencial intrínseco para contribuir activamente a la construcción de un humanismo digital verdaderamente crítico y reflexivo. Lejos de oponer resistencia a la inevitable digitalización, las lenguas clásicas pueden y deben participar en ella de manera proactiva. Al hacerlo, ofrecen marcos rigurosos para la reflexión profunda, una profundidad analítica indispensable para comprender el mundo complejo en el que vivimos y un rigor ejemplar en el uso del lenguaje. Esto es crucial en un panorama donde la precisión, la comprensión crítica, la capacidad de discernir entre la información y el ruido, y la formación de ciudadanos conscientes y éticos son más necesarias que nunca. El latín, al fomentar estas capacidades cognitivas y críticas, se posiciona no como un vestigio del pasado, sino como una piedra angular para el desarrollo de ciudadanos digitales plenamente competentes y conscientes.

Bibliografía

BURDICK, A., DRUCKER, J., LUNENFELD, P., PRESNER, T., & SCHNAPP, J. (2012), *Digital Humanities*, MIT Press.

TERRAS, M., NYHAN, J., & VANHOUTTE, E. (eds.) (2013), *Defining Digital Humanities: A Reader*, Ashgate Publishing.

DRUCKER, J. (2013), *SpecLab: Digital Libraries and the Digital Humanities*, Harvard University Press.

KRASHEN, S. D. (1985), *The Input Hypothesis: Issues and Implications*, Longman.

VANPATTEN, B. (2017), *While We're Talking: A Practical Guide for Applying Second Language Acquisition Research to the Classroom*, ACTFL.

ØRBERG, H. H. (2003), *Lingua Latina per se Illustrata*, Focus Publishing.

POLIS, C. (ed.) (2019), *Teaching the Latin Language: Reflections and Responses*, University of Michigan Press.

SELWYN, N. (2016), *Education and Technology: Key Issues and Debates*, Bloomsbury Academic.

HOLMES, W., PERSSON, M., & WASSON, B. (eds.) (2022), *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*, Routledge.

O'NEIL, J. (2020), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown.

Inquietar inteligencias, despertar conciencias: humanos artesanos. La IA en los procesos de formación de docentes colaborando en la construcción de competencias humanizadoras.

Patricia Bibiana Ramírez Lombardi
Universidad Católica de Santa Fe

Palabras clave: Formación de competencias; Inteligencia Artificial; Enfoque Humanizador de la Tecnología; Perspectivas académicas; Creatividad
Keywords: Skills development; Artificial Intelligence; Humanizing Approach to Technology; Academic perspectives; Creativity

Introducción

Según el diccionario de la Real Academia Española, “humanizar” se define como un verbo “*Que humaniza. Se pretende una educación humanizadora, que forme ciudadanos creativos, responsables, críticos y respetuosos*”. Y la educación contemporánea se enfrenta al desafío de integrar las tecnologías emergentes, especialmente la inteligencia artificial (IA), en un sistema que históricamente ha priorizado la uniformidad y la repetición. La IA, aunque vista con recelo por algunos educadores, presenta un potencial transformador que puede complementar y enriquecer el proceso educativo. Esta investigación reciente (no acabada) explora cómo la IA puede ser utilizada como una herramienta para humanizar la educación universitaria, enfocándose en el desarrollo de habilidades esenciales y la promoción de un aprendizaje significativo. El germen de esta investigación se encuentra en la resonancia de un cuento particularmente interpelante, titulado “Alma, despierta”, que obtuvo el segundo puesto en el certamen literario universitario del año 2024. Su autor, Franco Dumont, un estudiante avanzado y reflexivo del tercer año del profesorado en Ciencias de la Educación de nuestra propia institución, logró plasmar en su narrativa una exploración profunda y conmovedora de la intrincada relación entre el avance tecnológico y la experiencia humana en contextos hibridados. El relato nos movilizó y de allí surgió el interés en investigar cuánto afecta, positiva o negativamente, el uso de la IA en las aulas universitarias de formación de docentes.

El desafío de la automatización

La automatización, impulsada por la IA, ha permeado diversos aspectos de la vida cotidiana, desde los dispositivos móviles hasta la gestión del tráfico urbano. En el ámbito educativo, existe la preocupación de que la IA pueda reemplazar el toque humano esencial para una educación integral. Sin embargo, la propuesta pedagógica aquí presentada aboga por utilizar la IA como un apoyo al docente, no como un sustituto. Se subraya la importancia de que el educador guíe y motive a los estudiantes, fomentando habilidades como la empatía, la comunicación asertiva y la resolución creativa de problemas.

La esencia humana en la era de la IA

La educación humana se distingue por su naturaleza dinámica, emocional y social. A diferencia de la IA, los humanos experimentamos una amplia gama de emociones y nos comunicamos a través de lenguajes ricos en matices y gestos. La IA puede ser una herramienta poderosa, pero no debe reemplazar la conexión humana y la importancia del desarrollo de la inteligencia emocional si queremos que los futuros docentes sean verdaderos artesanos de dispositivos y procesos para una educación integral, protagónica y resolutiva.

Aprendizaje significativo y resiliencia

El aprendizaje significativo, trascendiendo la mera memorización, se cimienta en experiencias humanas auténticas y relevantes, constituyendo un pilar para el desarrollo integral del estudiante. Si bien la Inteligencia Artificial (IA) ofrece herramientas valiosas para complementar este proceso, facilitando la personalización y la retroalimentación adaptativa, es crucial no subestimar la gestión activa de la frustración y la internalización del fracaso como instancias intrínsecas y necesarias para el crecimiento personal y académico.

En un panorama global crecientemente complejo e incierto, cultivar la resiliencia y una mentalidad de crecimiento en los estudiantes emerge como un imperativo en la educación universitaria. La capacidad de adaptarse a la disrupción, aprender de los errores y obstáculos y perseverar ante la adversidad son habilidades fundamentales para afrontar los desafíos futuros con confianza y determinación. Implementada con, por y desde una visión didáctico-pedagógica humanista, la IA posee el potencial de ofrecer una retroalimentación formativa continua y personalizada, optimizando los procesos de aprendizaje de cada estudiante. Esto adquiere una relevancia particular en contextos donde las demandas académicas y personales pueden generar una sensación de sobrecarga y desánimo, subrayando la necesidad de un apoyo que combine la eficiencia tecnológica con la comprensión de las dinámicas emocionales humanas.

Ética y responsabilidad en la era de la IA

El uso ético de la IA en la educación es fundamental. Se debe promover la reflexión crítica sobre las implicaciones de esta tecnología, incluyendo la responsabilidad en su aplicación y la conciencia sobre los posibles riesgos. La formación de usuarios críticos de la tecnología es esencial para discernir entre información relevante y no relevante, y para utilizar la IA de manera que potencie y enriquezca el aprendizaje en lugar de deshumanizarlo. Lo aceptemos o no, los estudiantes de esta era universitaria ya se han plegado a las improntas de la IA, sobre todo por aquellas de opción generativa que les facilitan el surgimiento de ideas para textos ricos, la ubicación de documentos bibliográficos publicados en idiomas que no conocen pero que son valiosos y la generación de mapas o imágenes que son respuesta automática a los prompts que están aprendiendo a formular cada día con mayor complejidad.

El humano artesano: imperfección y trascendencia

La imperfección es una característica intrínseca de la condición humana, y es precisamente esta imperfección la que nos impulsa a crecer y aprender. La IA puede ofrecer eficiencia y rapidez, pero no puede replicar la chispa de la creatividad humana ni la capacidad de encontrar significado en la imperfección. La educación debe promover el desarrollo integral de los estudiantes, incluyendo su dimensión ética, emocional, social y espiritual, ayudándolos a encontrar un propósito en sus vidas y a conectar el aprendizaje con valores humanos fundamentales. Si solo buscan rapidez o exactitud en las aplicaciones de la IA vigentes, entonces estarán tercerizando en la robótica sus propias capacidades de inventiva. Aquí es donde radica, quizás, uno de los mayores desafíos de la educación con tecnologías desde un enfoque humanizador. El artesano no es perfección sino que busca también en las utopías y vericuetos imaginativos de la bella imperfección.

El papel de la IA en la educación del futuro

El debate sobre el papel de la IA en la educación está en sus inicios. Urge reflexionar sobre cómo integrar la IA de manera que se desarrollen y fortalezcan las competencias humanas sustanciales promoviendo el desarrollo integral de los estudiantes. Se identifican competencias humanas primordiales en la era de la IA, como el pensamiento crítico, la resolución de problemas, la colaboración, la empatía y la comunicación asertiva. La IA puede ser una herramienta valiosa para desarrollar estas competencias, pero no se logra porque sí; su implementación debe ser cuidadosamente planificada y guiada por principios éticos que impregnen las planificaciones disciplinares de emoción, vincularidad, sentidos de comunidad, bien común, respeto por la inteligencia personal y colectiva, amor por la historia de cada pueblo y de cada persona, entre muchos otros sentidos que han que considerar en el enfoque humanista.

Metodología

Se adoptó un enfoque metodológico mixto, priorizando la etnografía para una inmersión profunda en el grupo de estudiantes del profesorado de Ciencias de la Educación de la UCSF, con el docente investigador como observador participante. Esto permitió comprender sus prácticas y dinámicas en un entorno de confianza, utilizando diversas técnicas de recolección de datos cualitativos.

Para una comprensión más amplia, se realizó una encuesta cerrada, anónima y voluntaria, difundida entre estudiantes de diferentes profesorados en Santa Fe, generando datos cuantitativos. El muestreo fue no probabilístico por conveniencia, basado en la participación voluntaria de estudiantes de profesorados actuales o recientemente egresados.

Las preguntas de la encuesta fueron claras y concisas para facilitar respuestas rápidas y sencillas, buscando garantizar la validez y confiabilidad de los datos. El formulario de Google incluyó 12 ítems: tres sobre el perfil del participante (carrera, edad, género -opcional-), ocho de opción múltiple o selección de lista y una pregunta abierta y opcional sobre experiencias personales con la IA.

El objetivo principal es comprender en profundidad las complejas intervenciones y prácticas con IA de los estudiantes del profesorado en sus procesos de aprendizaje personales y colectivos. La investigación se centra en el significado que otorgan a sus prácticas académicas cotidianas desde una perspectiva humanizadora en la era de la IA, identificando patrones de interacción dentro del grupo con honestidad conceptual y procedimental, reconociendo la influencia de las percepciones subjetivas y las emociones. La elaboración del informe requirió un arduo trabajo interpretativo a la luz de la bibliografía para lograr una comprensión coherente de la realidad áulica, buscando fomentar una cultura resolutiva y una formación en competencias en escenarios educativos altamente humanizadores con la IA como eje transversal.

Resultados recogidos

Hasta la fecha, los datos recogidos son suficientes para estimar y definir líneas de diálogo y reflexión significativos para el moderno maridaje entre IA y sistema educativo universitario.

Aun cuando los docentes no mencionan y/o priorizan en sus prácticas (de manera explícita o visible) la IA, el cien por ciento de los estudiantes actuales respondieron que son usuarios frecuentes de aplicaciones. Los usos que ellos realizan están íntimamente relacionados a sus propios procesos de aprendizaje, puesto que la mayoría respondió que utilizan la IA para la producción de trabajos prácticos solicitados por las cátedras, sobre todo cuando urge el tiempo para las calificaciones o acreditaciones de los mismos.

Un alto porcentaje de los estudiantes han respondido que se debería propiciar una reflexión ética sobre los usos actuales de la IA pero no demuestran en el aula, en el hacer cotidiano, que les preocupe lo suficiente como para

mesurar la elección y usos de las aplicaciones que implementan.

La vinculación entre IA y ética, producción artesanal y exactitudes tecnológicas, eficiencia automática e imperfecciones humanas no son aún, para muchos de ellos, compromiso o responsabilidad. Los usos gratuitos de la IA priman sobre la reflexión ética y emocional que deberían considerar al momento de delegar en la tecnología el proceso creativo personal de una decisión, exposición o punto de vista.

Como docente a cargo, me animo fogonear el cambio. Considero, desde los datos recogidos, que es momento de animar los escenarios de reflexión académica no solo para decidir sobre el sí o el no de la IA en la práctica universitaria sino para, fundamentalmente, promover en los futuros docentes el compromiso indelegable de la impronta humanista de cualquier tecnología que sea transversal a los contenidos de una cátedra. Solo así se mantiene viva la chispa de la inteligencia curiosa, crítica, empática, emocional, humana y artesana.

Tecnología y preservación del patrimonio cultural en contextos de emergencia: herramientas y desafíos contemporáneos, el caso de la Dana en Valencia

Alejandra Nieto Villena
Álvaro Solbes García
Carmen Cano Sola
Universitat de València

Palabras clave: Fotografía; emergencia; preservación; patrimonio cultural; digitalización.
Keywords: Photography; emergency; preservation; cultural heritage; digitization.

Casos de emergencia climática como las inundaciones provocadas por la DANA (Depresión Aislada en Niveles Altos) el pasado 29 de octubre de 2024 en la provincia de Valencia, han demostrado que representan una amenaza significativa para el patrimonio cultural, tanto material como inmaterial. En este contexto, las tecnologías de la imagen, incluyendo el registro fotográfico o la digitalización mediante el uso de equipos portátiles y accesibles, aparecen como herramientas esenciales para la conservación y preservación de bienes culturales afectados. Esta comunicación persigue evaluar cómo estas tecnologías fueron empleadas en el rescate y salvaguarda del patrimonio cultural tras las inundaciones en Valencia en octubre de 2024, destacando iniciativas para la protección del patrimonio familiar y otras acciones colaborativas impulsadas por instituciones académicas y culturales como la Universitat de València.

Impacto de la DANA en el patrimonio cultural de Valencia

La DANA del pasado octubre causó devastadoras inundaciones en la Comunidad Valenciana, afectando gravemente a numerosos municipios de l'Horta Sud así como en algunas áreas de la Ribera Alta. Las precipitaciones intensas provocaron la crecida de ríos y la anegación de áreas urbanas, resultando en la destrucción de viviendas, infraestructuras y numerosos bienes culturales. Además del patrimonio cultural resguardado por instituciones públicas o privadas, el patrimonio familiar quedó arrasado por la enorme destrucción de viviendas, llevándose consigo innumerables tesoros familiares entre los que se encontraban los álbumes y fotografías de tantas generaciones de vecinos afectados. Este patrimonio se encuentra fuera del radar de protección

institucional, a pesar de ser considerado como una valiosa representación de la memoria colectiva y la identidad cultural de valencianos y valencianas. Ante el potencial riesgo de que las familias se deshicieran de todo, la iniciativa impulsada por la Universitat de València, *Salvem les fotos* ofreció apoyo para el rescate de estos objetos personales a través de una red de instituciones culturales que de forma colaborativa ofrecieron todos sus recursos.

Durante el proceso de rescate y salvaguarda de cientos de miles de objetos y fotografías llevados a cabo estos últimos meses, las tareas de gestión y construcción de infraestructura de emergencia ha ocupado mucho tiempo y esfuerzo. Paralelamente, las tareas de estabilización, limpieza y secado han permitido preservar una parte importante de ese patrimonio. El contexto de emergencia requirió el diseño de estrategias para optimizar las acciones de rescate donde los recursos tecnológicos se volvieron un instrumento fundamental del proceso.

Tecnologías de la imagen en la recuperación del patrimonio cultural

Ante la magnitud de los daños y con la idea de mejorar los protocolos de intervención diseñados en una primera fase, se fueron incorporando diversas estrategias basadas en tecnologías de la imagen para la recuperación del patrimonio cultural.

El primer recurso para optimizar el registro de los datos de las familias fue trasladarlos a un gestor de ítems que nos permitió trabajar de forma más coordinada y accesible. La creación de fichas centralizadas se tradujo en un registro mucho más ágil, independiente y completo. Desde el gestor de datos, podemos acceder cada miembro del laboratorio con nuestro móvil o tablet y registrar los avances de diagnóstico del estado de conservación, de intervención, el registro fotográfico, entre otros. Al inicio, al tener la información bajo un mismo procesador de datos se hacía más complejo el manejo y este modelo sirvió para enriquecer el registro y facilitar el acceso.

Por otro lado, de la mano de estos registros y una vez superada la primera fase más crítica, se comenzó a emplear la digitalización como recurso para la conservación de la imagen. La fotografía en su dualidad material y documental nos permite utilizar la digitalización de la imagen como recurso para preservar el contenido cuando la conservación del soporte y materiales originales peligra o no es posible llevarlo a cabo. Ante el estado de deterioro que presentan muchas fotografías, encontramos emulsiones desvanecidas que son ya irrecuperables y otras emulsiones estables que habitualmente se han salvado por encontrarse protegidas por el sobre o recubrimiento plástico de la hoja del álbum. La cantidad de agua y barro que impregnara esos soportes, el tiempo que ha permanecido húmedo, el tipo de emulsión fotográfica, la clase de soporte de papel, y otros factores extrínsecos como la temperatura y humedad ambiental, la condensación por almacenaje, entre otras causas, provocaron deterioros de innumerables alcances.

El volumen de material que se ha recogido en relación con los recursos humanos y la infraestructura disponible ha hecho complicado el poder recurrir a todo tipo de sistemas, herramientas y materiales. Sin embargo, un trípode o mesa de reproducción, una cámara fotográfica y un sistema de almacenamiento digital es suficiente para asegurar que al menos, ante casos críticos donde las limitaciones del proyecto no nos permiten consolidar una emulsión o restaurar la fotografía debido a las condiciones y tiempos limitados, podemos asegurar la preservación de la imagen *latente* antes de que siga deteriorándose y termine perdiéndose por completo. Un caso bastante común ha sido encontrar fotografías que durante el proceso de secado se han adherido en parte al film protector de la página del álbum fotográfico y otra parte ha permanecido en el soporte de papel, al despegarlas resulta muy complejo recomponer la emulsión sobre un solo soporte, pero al superponer las dos partes como capas, podemos visualizar parte de la imagen o en algunos casos, su totalidad. Para estas ocasiones, digitalizar ambas partes juntas y separadas (ambas capas) nos permitirá emplear los sistemas de retoque digital para recomponer la imagen y en el caso que sea necesario, reimprimirla y recuperar su materialidad.

Además del uso de estos recursos tecnológicos, se han utilizado pequeños equipos portátiles de laboratorio para examinar fotografías y otros objetos patrimoniales. Es el caso de la fotografía multibanda que solo necesita además de la cámara digital, el uso de luces especiales que permiten profundizar en el conocimiento tanto del estado de conservación como de la materialidad original de la obra. Es el caso de colecciones particulares de soportes pictóricos que también sufrieron daños y donde el empleo de estos equipos ha sido decisivo a la hora de realizar diagnósticos más detallados que nos ayudan, por ejemplo, a evidenciar los daños a los seguros que se encargarán de cubrir la restauración.

La reflectografía infrarroja ha sido otra de las tecnologías que se ha empleado en algunos casos críticos. Para la realización de estos estudios, tuvieron que trasladarse las obras al laboratorio de Análisis y diagnóstico de bienes culturales de la Facultad de Geografía e Historia, para poder así llevar a cabo un conocimiento aún mayor de la estructura estratigráfica de determinadas piezas. Se encontraron algunos resultados inesperados que ayudaron a comprender el proceso de manufactura del autor y poder relacionarlo con el fenómeno de deterioro.

Conclusiones

Los dispositivos empleados han permitido realizar un trabajo rápido, eficaz, preciso y sobre todo no invasivo tanto con las fotografías como con el resto de las piezas. Éstas, han facilitado la conservación de muchas fotografías a través del respaldo digital que podrá devolver al menos la imagen a las familias. Si pensamos en los sistemas tradicionales, las herramientas digitales nos han ayudado a reducir de forma significativa el tiempo necesario para el registro de grandes volúmenes de objetos como los que encontramos en situaciones de

emergencia y grandes sucesos devastadores. También podemos destacar el nivel de detalle y la calidad de los resultados que nos ofrecen estos equipos y que son cruciales para las tareas de restauración digital.

Podemos también destacar la facilidad de acceso a estos procesados técnicos, ya que la mayoría no requieren de conocimientos altamente especializados, lo que permitirá en un futuro que estudiantes y voluntarias de nuestros laboratorios puedan participar activamente en estas tareas de conservación preventiva, sobre todo ante estas situaciones de emergencia que requieren de más y óptimos recursos.

El uso de recursos colaborativos como gestores de datos compartidos o almacenamientos en la nube, hace más accesibles los resultados obtenidos, facilitando la colaboración entre instituciones y profesionales, así como pudiendo trabajar con esos datos de manera remota.

En suma, todas estas herramientas tecnológicas no solo optimizan procesos en términos de tiempo y eficiencia, sino que también contribuyen a la resiliencia de las instituciones patrimoniales, ampliando sus capacidades de respuesta y protección ante situaciones críticas.

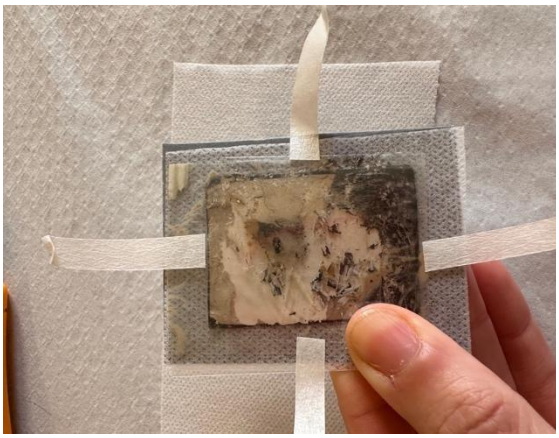


Figura 1. Montaje tipo sándwich de una fotografía con emulsión dañada (e irreparable).

Figura 2. Caja de fruta con álbumes fotográficos embarrados.

Figura 3. Voluntarias sacando fotografías de álbumes afectados.

Figura 4. Ejemplo de fotografía dañada por el barro y el agua, preparada para digitalizar.

Bibliografía

CONSERVATION CENTER FOR ART AND HISTORIC ARTIFACTS (2024), «Cuidado de los tesoros familiares», [Students in the Winterthur/University of Delaware Program in Art Conservation].

HESS NORRIS, D. (1998), *Disaster Recovery: Salvaging Photograph Collections*, Conservation Center for Art and Historic Artifacts.

SHIRAIWA, Y. (2015), «Rescuing tsunami-damaged photographs in Japan», *Topics in Photographic Preservation*, 16, pp. 315–323.

El Archivo Histórico Riva-Agüero (PUCP) y el acceso digital a colecciones patrimoniales privadas

Magally Alegre Henderson
Instituto Riva-Agüero –
Pontificia Universidad Católica del Perú

Palabras clave: Digitalización; archivos en peligro; AtoM; accesibilidad digital; patrimonio documental
Keywords: Digitization; archives in danger; AtoM; digital accessibility; documentary heritage

A partir de la experiencia del proyecto de investigación *Ai2 Archiving: Validación de un kit de herramientas especializado en catalogación de archivos históricos con Inteligencia Artificial*, financiado por el Vicerrectorado de Investigación de la PUCP, a través del Concurso Anual de Innovación 2024 (CAP-Inn-2024-PI1150), esta presentación analiza los retos que enfrentan en el Perú los procesos de digitalización de colecciones patrimoniales (GLAM en inglés por: Galerías, Bibliotecas, Archivos y Museos). La mayor parte de ellos no cuentan con bases de datos estandarizadas, ni con un sistema integrado de catalogación que permita la gestión de colecciones digitales y responda a las necesidades ciudadanas de accesibilidad digital al patrimonio cultural. El impacto de esta brecha de acceso a la información es muy alto, en particular en el desarrollo de las humanidades digitales. Ello que se evidencia en una importante participación en proyectos de digitalización con financiamiento internacional, como el *Endangered Archives Programme* que gestion la Biblioteca Británica con el auspicio de Arcadia y en el que el Perú tiene casi el doble de proyectos que cualquier otro país de la región.

En términos generales, los archivos históricos y otros tipos de repositorios de colecciones patrimoniales en el Perú, enfrentan una problemática similar en relación a la gestión de sus colecciones: la dificultad al interior de las instituciones para movilizar recursos económicos para el desarrollo e implementación de softwares de catalogación y la dificultad de movilizar recursos humanos suficientes para emprender procesos de catalogación intensiva de sus acervos documentales y patrimoniales. A nivel país, el Perú tampoco cuenta con un buscador nacional que integre los catálogos y objetos digitales de colecciones patrimoniales custodiadas por el Estado, ni con una política de catalogación o de digitalización nacional que se oriente a la interoperatividad de esta data. En ese sentido, proponemos que, a partir del uso de metodologías de Humanidades Digitales, la archivística universitaria puede contribuir a dar respuesta a esta

problemática, al abordar las necesidades de accesibilidad digital que garantizan el derecho al patrimonio cultural.

Desde el Archivo Histórico Riva-Agüero–AHRA, que forma parte del Instituto Riva-Agüero de la PUCP, y que conserva el principal acervo documental del país en manos privadas, nuestro compromiso principal es garantizar la accesibilidad digital plena, de modo que nuestro valioso patrimonio documental esté al servicio de la comunidad académica dentro y fuera del país. Para ello, desarrollamos una intensa actividad en humanidades digitales con 3 pilares fundamentales: 1) Capacitación 2) Diseño de proyectos y 3) Alianzas internacionales.

Así, en febrero de este año, desarrollamos el Curso de Verano en Humanidades Digitales –PREFALC CONTAMAZ - *Contar la Amazonía,* junto con la Universidad Marie y Louis Pasteur (antes Universidad del Condado Franco, Francia) y la Universidad Nacional de Colombia (UNC). Gracias al auspicio del Fondo PREFALC – Programa Regional Francia-América Latina-Caribe, los becarios de la PUCP y de la UNC tuvieron la oportunidad de aprender el uso de herramientas de humanidades digitales para el análisis de una fuente inédita, el conjunto de 24 cuadernos de notas de Marcos Cavero, un administrador de la Aduana Fluvial de Iquitos, escritos a inicios del siglo XX. Los cuadernos, pertenecientes a la colección Félix Denegri Luna, han sido digitalizados gracias al proyecto Peruvian Legacy, financiado por el Endangered Archives Programme de la Biblioteca Británica (EAP1495). Para su procesamiento digital, se usaron herramientas de extracción de datos como *Transkribus,* herramientas para el análisis de datos como *Voyant Tools*, *Iramuteq*, *uMap* y otros desarrollos para edición de textos TEI/XML. En agosto de 2025 se realizará la segunda edición de este curso en la Universidad Nacional de Colombia, dedicada al procesamiento digital de *La Vorágine* de José Eustasio Rivera.

A su vez, desde 2023, el AHRA actúa como punto de enlace latinoamericano del Programa de Archivos en Peligro de la Biblioteca Británica, que promueve la digitalización de patrimonio documental en peligro. Así, como parte de esta alianza, preparamos para julio de 2025, nuestro primer taller de digitalización de patrimonio documental para América Latina, con la participación de expertos del Equipo de Imágenes de la Biblioteca Británica. Este taller gratuito, ha recibido una enorme acogida, lo que permitirá crear una red de archivos, públicos y privados, con los cuales desarrollar en común estándares de digitalización y principios de interoperatividad en la catalogación de acervos digitales.

A estas iniciativas de capacitación se suma el proyecto *AI-2 Archiving*, con el objetivo de implementar un software de catalogación de código abierto (AtoM) para el registro de las colecciones documentales en el AHRA. Esta implementación conlleva el desarrollo de un kit de herramientas de Inteligencia Artificial, para la estandarización y limpieza de sus bases de datos. La diversidad de soportes y tipos de documentos que albergan las 86 colecciones de este Archivo, cuyos documentos más antiguos datan del siglo XVI, lo convierte en un

piloto excepcional para el desarrollo de un sistema de catalogación que permita su uso en otras colecciones documentales y patrimoniales universitarias, con la expectativa de poder implementarse en todo tipo de colecciones patrimoniales peruanas. *Ai2 Archiving* se encuentra en la fase de diseño de un prototipo de software basado en IA para crear un sistema de catalogación personalizado, diseñado para satisfacer las necesidades específicas de los repositorios del sector GLAM (Galerías, bibliotecas, archivos y museos, por sus siglas en inglés). Este prototipo consiste en un kit de herramientas que incluye manuales, prompts y entrenamiento en español, enfocado en la catalogación de patrimonio documental.

Ai2 Archiving mejora la accesibilidad de los archivos históricos al acelerar los procesos de catalogación archivística, procesos claves en las funciones de conservación y accesibilidad de los archivos históricos y documentales, que sin embargo, se caracterizan por ser procesos laboriosos y lentos. Debido a la cantidad de tiempo de procesamiento que se requiere para catalogar cada documento, en el caso de grandes volúmenes de documentos, se necesita contar con un ejército de personas. Esto es aún más intensivo en el caso de documentos históricos manuscritos, que requieren de un proceso de transcripción de grafías antiguas que demandan un tiempo adicional de procesamiento, o el uso de tecnologías de HTR (Handwritten Text Recognition).

Para enfrentar los desafíos de estandarización (que garantiza la interoperatividad de la data) en los procesos de catalogación de colecciones documentales y otros repositorios de patrimonio cultural es indispensable llevar a cabo procesos manuales de limpieza de datos y de categorización que demandan un considerable número de horas y un esfuerzo intensivo. Aunque la inteligencia artificial (IA) ha comenzado a usarse en la catalogación, facilitando la automatización de tareas repetitivas y la organización de grandes volúmenes de información, su aplicación en experiencias específicas de catalogación en archivos peruanos es verdaderamente incipiente. Sin tesauros u ontologías específicas para la catalogación de patrimonio cultural peruano, esta carencia resalta la necesidad urgente de desarrollar y adaptar tecnologías de IA que no solo comprendan, sino que también procesen eficientemente la extracción y sistematización de data, optimizando así la accesibilidad digital y conservación del patrimonio archivístico del Perú.

A través de estas acciones, el Archivo Histórico Riva-Agüero promueve, desde el ámbito de los archivos universitarios, la accesibilidad de fuentes documentales en tres estrategias interconectadas: capacitación, desarrollo de proyectos y alianzas internacionales, que responden a las dificultades de acceso a fuentes digitales y a las necesidades de interoperabilidad de los repositorios digitales disponibles para el estudio de la historia y la cultura peruanas.

La Red Goya: Humanidades Digitales para mapear la recepción de Goya en el siglo XIX

Paula Fayos Pérez[1]
Anita Verö[2]

Palabras clave: Francisco de Goya; Visualización de redes; Colaboración interdisciplinar; Visualización Interactiva; Historia del arte
Keywords: Francisco de Goya; Network Visualization; Interdisciplinary Collaboration; Interactive Visualization; Art History

Introducción

A lo largo de los siglos XVIII y XIX, las redes de artistas, escritores, coleccionistas e intelectuales fueron fundamentales para la producción artística, así como para la difusión y comprensión de figuras históricas. El mapeo de estas conexiones nos permite comprender mejor la dinámica de la sociabilidad y la circulación de ideas. En los últimos años, proyectos en el ámbito de las Humanidades Digitales han demostrado el potencial del análisis de redes para arrojar luz sobre los entornos artísticos históricos, como el Archivo de Cartas de Thorvaldsen, Mapping the Republic of Letters (Universidad de Stanford) o Early Modern Letters Online (EMLO) (Universidad de Oxford). El desarrollo de las HD ha superado la fase inicial de digitalización para adentrarse en nuevas formas de estructurar el conocimiento. Las redes visuales permiten representar la complejidad histórica desde perspectivas relacionales, mostrando cómo los sujetos, las ideas y los objetos culturales se entrelazan en el tiempo y el espacio.

La Red Goya o Goya Network (www.goyanetwork.com) responde a esta transformación, ofreciendo una interfaz interactiva que reconstruye las conexiones de uno de los artistas más influyentes del ámbito hispánico, Francisco de Goya. Traza las relaciones sociales y profesionales que rodeaban al artista, centrándose en sus conexiones con artistas, críticos y coleccionistas de España y Francia en el siglo XIX. A diferencia de los estudios tradicionales sobre la obra de Goya, este proyecto hace hincapié en el contexto relacional que dio forma a su legado, ofreciendo una herramienta novedosa para el estudio de las redes sociales de las figuras históricas. Como base de datos “viva”, facilita la

[1] Investigadora postdoctoral en Historia del Arte. Beca Leonardo 2024, Fundación BBVA. Doctora en Historia del Arte (Universidad de Cambridge, 2019). Miembro externo de los grupos SU+MA (Universidad Complutense de Madrid) y ARCHE (Universidad de Estrasburgo).
[2] Informática especializada en aprendizaje automático (Machine Learning). Doctora en Inteligencia Artificial (PhD in Computer Science, Universidad de Cambridge, 2022). Fundadora de ConsonantAI (www.consonant-ai.com).

colaboración interdisciplinar y sirve tanto como herramienta de investigación como recurso público, proporcionando información sobre el alcance de la influencia de Goya y la recepción de su obra.

Origen

La Red Goya un proyecto digital basado en la investigación desarrollado por Paula Fayos-Pérez como parte de su trabajo postdoctoral sobre las redes sociales y profesionales de Goya. Es una continuación de su trabajo de doctorado (Universidad de Cambridge, 2018) (repositorio), que posteriormente se ha publicado en forma de libro: Goya's Caprichos in Nineteenth-Century France. Politics of the Grotesque (CEEH, 2024). Con el apoyo de una Beca Leonardo (Fundación BBVA), la parte digital del proyecto comenzó a desarrollarse en septiembre de 2024 con la colaboración de Anita L Verő, doctora en Inteligencia Artificial (Universidad de Cambridge, 2022), informática y fundadora de ConsonantAI. Se tomó como referencia el código abierto de un proyecto para visualizar la red de artistas en una lista de reproducción de Spotify, desarrollado por Verónica Espinoza en GitHub.

Metodología: infraestructura técnica y despliegue abierto

La Red Goya (fig. 1) es una plataforma interactiva desarrollada con tecnología web abierta y desplegada a través de GitHub y Vercel, que permite explorar las conexiones entre artistas, coleccionistas, mecenas y críticos en la España y Francia del siglo XIX. Esta comunicación presenta una filosofía de diseño colaborativo y un modelo semántico de datos e integración de contenidos.

Frente a los modelos institucionales cerrados, La Red Goya ha sido concebida como un proyecto de código abierto, fácilmente replicable. La página web utiliza HTML y JavaScript, con datos estructurados en JSON. Se desarrolla en Visual Studio Code, con control de versiones e implementación a través de GitHub, y se despliega de manera continua mediante Vercel, un sistema que ha sido implementado con la ayuda de ChatGPT (OpenAI). Esta arquitectura facilita tanto la transparencia como la sostenibilidad técnica a largo plazo. También se han utilizado herramientas de IA para generación de imágenes cuando no existen retratos históricos documentados. Por otro lado, la metodología para la investigación en la parte de Historia del Arte del proyecto está basada en el estudio de fuentes bibliográficas y documentos de archivo. Solemos pensar en las redes como ilustraciones visuales, pero esta metodología puede ir mucho más allá, ya que las redes son formas de investigación en sí mismas. Ver un nodo aislado, un edge inesperado o una constelación densa no es sólo informativo sino que es además generador de nuevas preguntas y, junto a las contribuciones de los usuarios, creador de conocimiento.

Resultados y experiencia de usuario

En la Red Goya, cada nodo (individuo) y cada edge (conexión entre dos individuos) contiene metadatos (fig. 2). En los nodos, los usuarios pueden ver una imagen del individuo, información biográfica, así como la producción artística/literaria, actividad profesional, colección de arte, afiliaciones, maestros y discípulos, viajes, vínculo con la obra de Goya (menciones, colección, copias), bibliografía, y otros. El tamaño de los nodos es proporcional al número de conexiones (edges) que contienen. Cada nodo puede además citarse académicamente y se ofrece exportación en formatos como CSV o RIS, y existe también la opción de compartir en redes sociales.

En los edges entre dos individuos se puede ver el tipo de relación, correspondencia, colaboraciones, retratos, parentesco, y otros. Los edges de color gris/plateado son conexiones directas (amigos, colegas, familiares, conocidos, etc.), mientras que los edges de color dorado son conexiones secundarias (por ejemplo, artistas y escritores que copiaron o escribieron sobre Goya, pero nunca lo conocieron en persona). Al hacer clic sobre un nodo o en un edge, se abre un panel lateral todos los metadatos contenidos en ellos. La interfaz incluye además dos barras de búsqueda con desplegables para filtrar por profesión (pintor, dibujante, editor, coleccionista, político, jurista, etc.) y por país (Francia, España, Alemania, Reino Unido, Irlanda, etc.), y muestra automáticamente hipervínculos entre nodos relacionados.

El diseño busca un equilibrio entre rigor académico y accesibilidad, y es plenamente funcional en dispositivos móviles. Toda la visualización está optimizada para facilitar el razonamiento visual por parte del usuario.

Colaboración y apertura

Aunque nace como un proyecto personal de investigación, la Red Goya se concibe como una base de datos colaborativa con el objetivo de convertirse en una herramienta útil para los profesionales de museos y patrimonio cultural, académicos, estudiantes y el público en general interesado en Goya.

Este enfoque fomenta una filosofía de ciencia abierta (open science) y participación colectiva: los usuarios pueden enriquecer la red, señalando errores, aportando referencias o proponiendo nuevos datos. A través de la sección de Contributions que envía a un formulario GoogleDoc, son bienvenidas las contribuciones de historiadores del arte, historiadores, hispanistas, profesores y conservadores especializados en Goya, la España del siglo XIX, la Francia del siglo XIX y cualquiera de los miembros de la red (por ejemplo, Federico de Madrazo, Prosper Mérimée, Eugène Delacroix, etc.). En la página se acredita explícitamente a los contribuyentes, a no ser que prefieran mantenerse en el anonimato (lo que es una opción en el formulario).

Potencial y futuras funciones previstas

Actualmente trabajamos en nuevas funcionalidades que permitirán una lectura temporal, comparativa y temática de la red:

- Una barra temporal (time slider) para explorar la red por periodos históricos.
- Identificación automática de clusters o constelaciones dentro de la red, que podrían estudiarse por separado (por ejemplo, redes femeninas, grupos editoriales, la familia Madrazo o circuitos de traducción).
- Exploración de patrones a partir de coincidencias en los metadatos.
- Implementar una nueva barra de búsqueda para filtrar por sexo (hombres y mujeres).

Estas herramientas permiten una lectura relacional más rica, fomentando la interpretación visual crítica como forma de generar nuevas hipótesis.

Un modelo reproducible: más allá de Goya

Más allá del caso concreto de esta red, proponemos este modelo como un marco adaptable a otros proyectos de Humanidades, promoviendo una metodología visual y reproducible que conciba las redes como infraestructuras epistemológicas compartidas. El código y los datos están disponibles para su reutilización, de modo que investigadores de otras disciplinas pueden generar fácilmente sus propias redes a partir de esta estructura.

Proponemos por tanto que La Red Goya puede convertirse en una plantilla adaptativa para proyectos de Humanidades Digitales que quieran mapear relaciones históricas complejas.

Algunos ejemplos de posibles adaptaciones:

- Redes intelectuales de mujeres en el siglo XIX
- Cartografías de traducción literaria entre lenguas ibéricas
- Redes de coleccionismo o mecenazgo transnacionales
- Comunidades epistolares en contextos de exilio o censura

La clave del modelo es su flexibilidad y escalabilidad: permite añadir nuevas categorías, nuevos tipos de relaciones, y soporta integración con otros estándares como Wikidata.

Derechos de autor y licencia:

La Red Goya está protegida por escritura notarial (Madrid, 2025), que documenta el proceso de investigación y desarrollo desde 2018 hasta 2025. Su código y estructura están licenciados bajo CC BY-NC 4.0, lo que permite su uso académico no comercial con la debida atribución al autor.

Conclusión: redes como infraestructuras humanísticas

La Red Goya es a la vez una herramienta de investigación y una propuesta metodológica. Muestra que es posible construir un entorno digital riguroso y

accesible desde una lógica colaborativa, modular y abierta. La clave no es el volumen de datos, sino la calidad semántica y la capacidad de adaptación al ser un continuo work in progress. Al combinar precisión técnica (clustering, filtros, IA) con sensibilidad humanística (conocimiento del contexto histórico, análisis de documentos), La Red Goya demuestra que las Humanidades pueden beneficiarse de metodologías híbridas sin renunciar a la profundidad crítica.

Frente a la centralización tecnológica, apostamos por un modelo que permita a investigadores en cualquier contexto crear, modificar y compartir sus propias redes. Nuestra propuesta es que pensemos las redes no como productos finales, sino como espacios en construcción donde Historia y tecnología pueden encontrarse. Esto queda reflejado en nuestro artículo recientemente aceptado en Leonardo (MIT Press), una revista dedicada a la intersección entre ciencias y Humanidades.

Invitamos a la comunidad de Humanidades Digitales Hispánicas a imaginar cómo este modelo puede desplegarse, diversificarse y escalarse en nuevas direcciones colectivas.

Figuras, tablas, mapas o gráficas

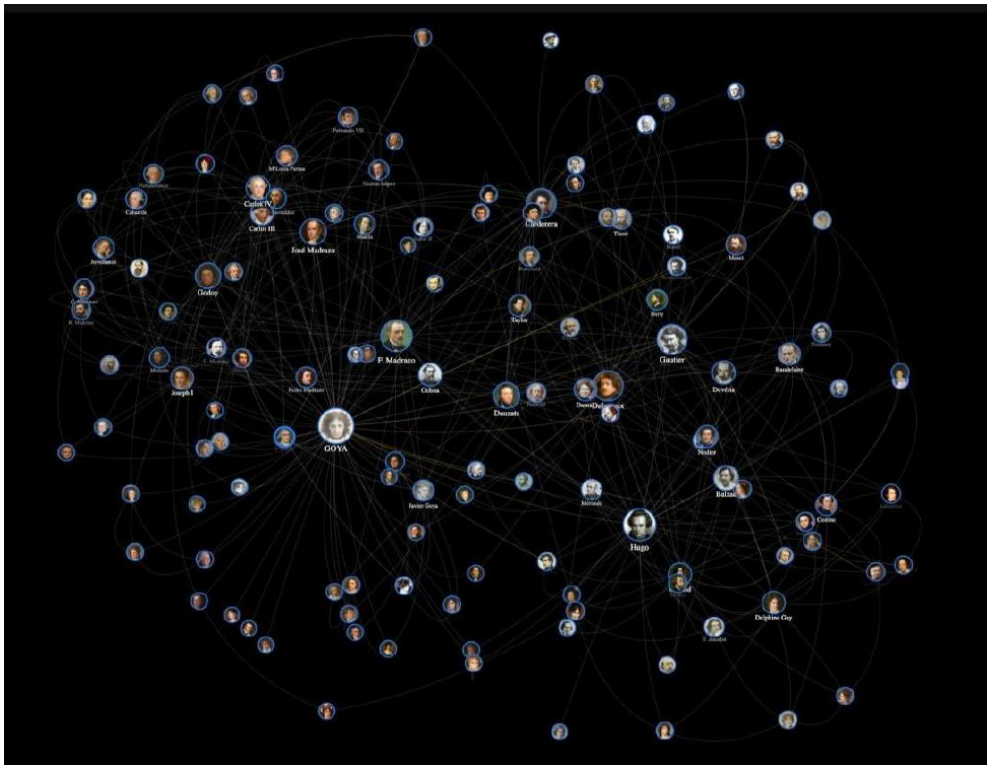


Figura 1. Visualización de la Red Goya completa (junio 2025)

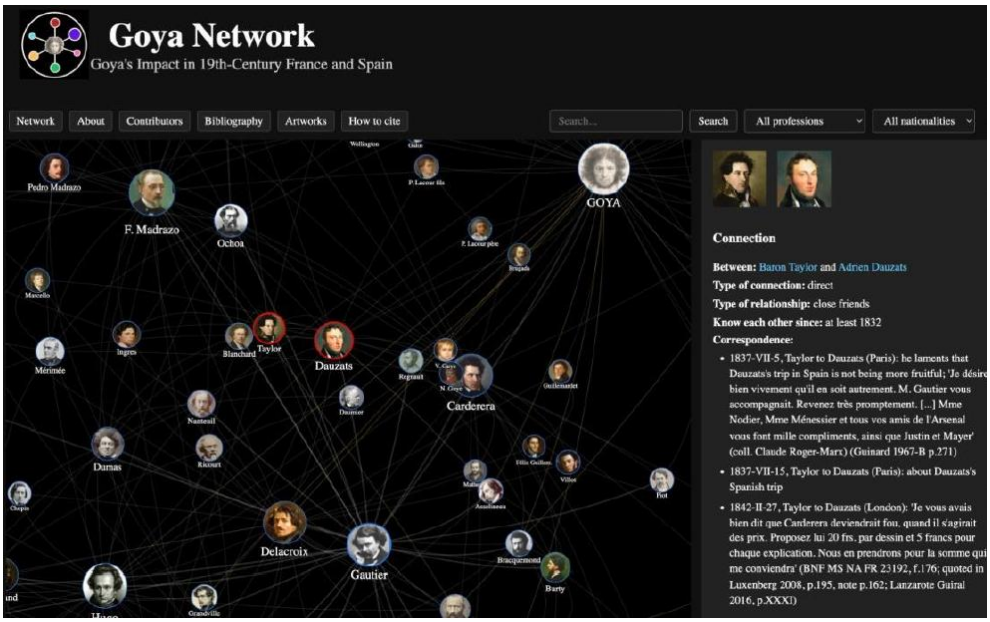


Figura 2. Visualización de la página web Goya Network, incluyendo las diferentes pestañas, barras de búsqueda y sección lateral para mostrar la información de los nodos y edges (junio 2025).

Bibliografía

FAYOS PÉREZ, P., & VERŐ, A. (en prensa, 2025), «The Goya Network: Mapping Goya's Impact in 19th-century France & Spain», *Leonardo*.

FAYOS PÉREZ, P. (2024), *Goya's Caprichos in 19th-Century France: Politics of the Grotesque*, Madrid, Centro de Estudios Europa Hispánica (CEEH).

GILARRANZ IBÁÑEZ, A. (2020), «No solo es arte: las redes de Federico de Madrazo durante su pensionado en París y Roma (1837–1842)», *Bulletin hispanique*, 122(1), pp. 277–292.

JAILLANT, L., & REES, A. (2023), «Applying AI to digital archives: Trust, collaboration and shared professional ethics», *Digital Scholarship in the Humanities*, 38, pp. 571–585.

Soluciones y nuevos modelos para la edición crítica digital de la obra de Calderón

Candela Iglesias Balsa
Universidade de Santiago de Compostela – GIC

Palabras clave: Edición digital; crítica textual; Calderón; teatro del Siglo de Oro
Keywords: Digital edition; textual criticism; Calderón; Golden Age theater

La elaboración de ediciones digitales en el ámbito del teatro del Siglo de Oro posee ya un recorrido reseñable que se dirige progresivamente hacia la creación de un espacio de trabajo digital que suponga una verdadera diferencia con respecto de las ediciones en papel.

En 2020, con el inicio del proyecto CALDERON DIGITAL (PID2020-114711GB-100), el Grupo de Investigación Calderón (USC) puso en marcha la creación de una biblioteca digital, vinculada a un corpus con el interactuase bidireccionalmente, que permitiese consultar en línea los textos de las comedias calderonianas en ediciones críticas. El modelo de edición digital filológica confeccionado inicialmente, y sobre el que se articulan las ediciones elaboradas hasta el momento, sigue las pautas establecidas por ejemplos previos como el de *La dama boba*, bajo la dirección de Marco Presotto—que continúa el modelo de edición académica iniciado por las Internet Shakespeare Editions— o *La Estrella de Sevilla*, nacida en el marco de la tesis doctoral de Nàdia Revenga. Este modelo, que tuvo su primera manifestación en la edición digital de *Las manos blancas no ofenden*, a cargo de Verónica Casais, permite la consulta de un texto crítico, depurado de errores e intervenciones no autoriales, así como de las digitalizaciones de los diversos testimonios que se han tenido en cuenta para la elaboración de la edición y que pueden alinearse con el texto crítico para visualizarse de manera simultánea. El texto guía se ofrece anotado: los pasajes con notas o variantes aparecen en la interfaz subrayados y sombreados, respectivamente y al clicar sobre ellos se despliega una ventana emergente en la que leer la nota o seleccionar el testimonio de interés para consultar su variante.

Sin embargo, este prototipo de edición crítica digital creado inicialmente no tiene en cuenta todos los escenarios textuales que se producen en la transmisión del corpus calderoniano y a medida que avanzaba el proceso de confección de ediciones para la biblioteca, surgió la necesidad de reflexionar acerca de nuevos modelos y posibilidades para enfrentar y representar digitalmente los retos ecdóticos impuestos por la historia textual de las comedias de Calderón. La

ausencia de un paradigma verdaderamente digital y único con el que afrontar los diferentes problemas ecdóticos que presenta el teatro del Siglo de Oro dio lugar a la necesidad de ir solventando progresivamente los obstáculos que se iban planteando a la vez que se procuraba idear nuevos tipos de edición crítica que se amoldasen a las diferentes características de la transmisión textual que puede presentar una obra: procesos de reescritura, dobles versiones, piezas breves… Por lo tanto, el obstáculo fundamental, sustentando en la dificultad de crear un modelo de edición crítica digital compatible con todas las circunstancias ecdóticas presentes en la transmisión del teatro aurisecular se atomiza en retos diferentes que se expondrán en las líneas que siguen. Para su solución, fue preciso aunar conocimientos de crítica textual tradicional con las posibilidades ofrecidas por el lenguaje XML y las soluciones de programación convenidas con NLPgo, la empresa informática que asesora al grupo en el proyecto.

Uno de los primeros retos fue el de las comedias con una transmisión textual manuscrita que presenta intervenciones de varios copistas o modificaciones de carácter textual o escénico, sean o no autógrafas. Es el caso de piezas como *El monstruo de los jardines* o *El mayor encanto, amor* en las que es interesante poder reflejar los cambios experimentados por un mismo testimonio derivados del uso de sus poseedores, que pueden llegar a ser clave en la filiación de testimonios.

En relación con estos ejemplos también están los casos de comedias que presentan pasajes notablemente reescritos mediante grandes omisiones o adiciones de texto. Es reseñable el caso del final de *El mayor encanto*, *amor* o del autógrafo de *El secreto a voces,* que ha sido sometido a una revisión de Calderón en la que se atajan numerosos versos que dan lugar a versiones largas y breves de algunas escenas y presenta intervenciones de otras manos. La edición digital de esta comedia exige poder identificar los versos descartados por mano ajena y poner a disposición del usuario los que la comedia ofrecía antes de la revisión autorial. El problema de estas comedias no reside tanto en el añadido de versos, como en la representación de omisiones largas, porque el esquema no ofrece la posibilidad de marcar todos los versos omitidos bajo una misma etiqueta de «referencia», sino que cada verso ha de ser etiquetado individualmente. Marcar cada verso completo genera un visionado en la interfaz que no permitiría distinguir en ese pasaje las variantes presentes en los testimonios que sí hubiesen transmitido los versos, puesto que todo el fragmento se representaría sombreado en la interfaz.

En el caso de que el grado de reescritura o las diferencias entre testimonios posean la envergadura suficiente como para alcanzar la consideración de que existen dos versiones de la misma comedia, lo fundamental será mantener la centralidad de ambas, en la medida en que se asume la existencia de dos originales. Por lo tanto, la disposición de las ediciones críticas en la biblioteca digital no debe dar a entender la primacía de un texto sobre otro, pero ambas ediciones deberían poder interactuar entre ellas porque comparten un vínculo textual. El modelo actual está muy lejos de cubrir las necesidades que presenta

la edición de este tipo comedias y la solución pasaría por definir un nuevo modelo que pueda convivir en la biblioteca con el anterior u optar por medidas que aprovechen el ya existente.

Otro problema similar es el de la edición de fiestas cortesanas que se han trasmitido con loas de autoría calderoniana. Comedias como *Los tres mayores prodigios* (1636) o *El golfo de las sirenas* (1657) constituyen ejemplos en los que la loa y la pieza principal comparten un estrecho vínculo que exige interpretar todos los elementos de la fiesta como partes de un todo de las que no se puede prescindir [4]. La dificultad fundamental está en editar la loa de manera que preceda a su comedia correspondiente y se pueda relacionar con ella, pero que mantenga su independencia en el corpus y en la numeración de los versos.

La resolución de estos problemas no ha visto todavía resultados uniformes en cuanto a su calidad; en algunos casos se han implementado soluciones sencillas, mientras que, para los obstáculos más complejos, se han empezado a esbozar posibles vías para desarrollar ediciones digitales que presenten esas casuísticas. Los resultados obtenidos hasta el momento se exponen a continuación.

Para representar las modificaciones de los manuscritos se resolvió añadir notas explicativas a las variantes. Se trata de un proceso de anidación que el esquema XSD de partida contemplaba, pero que requería de una serie de modificaciones en el procesado para mostrarse adecuadamente en la interfaz. Con esta solución, las variantes que presentan correcciones o reescrituras figuran subrayadas en la ventana emergente y su consulta ofrece una explicación o una transcripción detallada de los cambios experimentados por ese lugar crítico.

De cara a la representación de omisiones largas en el texto crítico, se está trabajando en la introducción en el esquema XSD de un atributo «resaltado» que en el nivel del procesado de los ficheros por parte del programa diseñado por NLPgo permita marcar los versos omitidos con un color o sombreado diferente al de las variantes habituales. Esto haría posible que los pasajes ausentes en algún testimonio se detectaran a simple vista.

La solución idónea a la edición de comedias con doble versión sería crear un nuevo modelo que permitiese alinear los pasajes de dos textos críticos de la misma manera que se alinean con las imágenes de los testimonios. No obstante, la elevada dificultad de esta propuesta obliga a plantear una solución más simple, aunque no ideal, que consistiría en editar cada versión por separado y servirnos del atributo «resaltado» para marcar los versos reescritos o presentes en una versión y no en otra. El vínculo entre las piezas se indicaría no solo mediante el título o la mención a un doble título, sino también a través de la inclusión en la cabecera de la pieza de un enlace a la otra versión.

Finalmente, a la hora de elaborar ediciones de fiestas cortesanas o comedias precedidas de loa, la propuesta consiste en elaborar la edición de las piezas breves como si se tratase de una comedia de una sola jornada. A continuación, se modificaría el esquema para poder incluir elementos como

«loa» o «entremés» dentro del elemento «contenido». Estos elementos tendrían un atributo «ref» para indicar el lugar en el que está en realidad el XML independiente que acoge el texto marcado. La visualización en la interfaz permitiría la vinculación entre ambos textos mediante un enlace en los paratextos.

Bibliografía

AICHINGER, W., KROLL, S., & RODRÍGUEZ-GALLEGO, F. (eds.) (2015), *Pedro Calderón de la Barca, El secreto a voces*, Kassel, Reichenberger.

CALDERÓN DE LA BARCA, P. (s. f.), *Las manos blancas no ofenden*, ed. crítica y digital de V. CASÁIS VILA. Disponible en: https://www.calderondelabarca.org/ediciones_criticas [Acceso: 03/10/2024].

CASAÍS VILA, V., & FERNÁNDEZ MOSQUERA, S. (2022), «Hacia una biblioteca digital de las comedias de Calderón: el proceso del GIC», *Hipogrifo: Revista de literatura y cultura del Siglo de Oro*, 10(1), pp. 373–384.

FARRÉ VIDAL, J. (2013), «La mecánica imprecisa de las loas palaciegas», *Anuario Calderoniano*, 1 (Extra), pp. 147–161.

IGLESIAS BALSA, C. (en prensa), «Correcciones, dobles versiones y piezas breves: algunos retos pendientes en la edición digital de la obra dramática de Calderón», *Archivum Calderonianum*, s. p.

INTERNET SHAKESPEARE EDITIONS (s. f.), University of Victoria, recurso en línea. https://internetshakespeare.uvic.ca/ [Acceso: 28/04/2025].

PRESOTTO, M. (dir.) (s. f.), *Lope de Vega, La dama boba: edición crítica y archivo digital*, con la colaboración de S. BOADAS, E. MAGGI, & A. PESSARRODONA.

REVENGA GARCÍA, N. (2021), *«La Estrella de Sevilla» y las potencialidades de la edición digital* [Tesis doctoral], València, Universitat de València.

Primeras calas en *Valor, ingenio y fineza*: edición digital y análisis de redes

Carles Márquez Molins
Universitat de València

Palabras clave: Teatro del Siglo de Oro; Francisco de la Torre y Sevil; edición digital; análisis de redes; interacción lingüística
Keywords: Golden Age Theater; Francisco de la Torre y Sevil; digital publishing; network analysis; linguistic interaction

Es de sobra conocida la importancia de las Humanidades Digitales en la investigación sobre el teatro del Siglo de Oro, que puede comprobarse en la trayectoria de grupos de investigación de excelencia como ARTELOPE, DICAT o Manos Teatrales. La presente ponencia parte de esta trayectoria consolidada para ejemplificar un caso práctico centrado en el estudio de la obra dramática de Francisco de la Torre y Sevil y la edición de una de las obras que pudo haber escrito: *Valor, ingenio y fineza*. Alvar López (1947; 1985; 1986; 1987; 1990) estudió la figura de Francisco de la Torre y Sevil en su vertiente poética, pero hasta ahora no se le ha dedicado la atención que se merece a su producción como autor teatral. La única comedia de Torre y Sevil que se ha editado recientemente es *La Azucena de Etiopía* (Giovanni Cara, 2006), quedando así sin editar las otras obras que conforman el corpus teatral de este autor y que no son colaboradas.

Valor, ingenio y fineza es una zarzuela atribuida a Francisco de la Torre y Sevil en el único manuscrito que de ella se conserva, que se encuentra albergado en la Hispanic Society of America (B2533-2). Dicha obra no se ha estudiado ni editado, por lo que creemos oportuno ofrecer una edición digital anotada que se base en los métodos ecdóticos tradicionales, pero que esté codificada en XML-TEI y haga un buen uso de las potencialidades del medio digital, como pueden ser la creación de gráficos a partir de la marcación de la obra o la anotación de determinados elementos, como las notas de vocabulario o aquellas con contenido explicativo. Asimismo, hemos de decir que la edición digital nos parece tanto un punto de llegada como un punto de partida para realizar otro tipo de análisis. De ese modo, no podrá entenderse en su totalidad la edición de *Valor, ingenio y fineza* sin su estudio introductorio, y será bueno ampliar la mirada sobre el objeto de estudio desde otros enfoques, como el análisis de redes, en el que también se ha trabajado. La razón por la que nos hemos planteado aplicar la metodología del análisis de redes estriba en que elaborar una red de interacciones lingüísticas entre los personajes, nos permite estudiar la relación

entre ellos a partir de un criterio lo más objetivo posible y ver, de ese modo, qué personajes interactúan más entre ellos, si sobresale la interacción de algunos de ellos, etc. Al tratarse de una obra de la que no tenemos confirmada la autoría, nos planteamos, además, compararla con otras obras del autor para ver si existen o no semejanzas en los modelos que muestran las redes de interacciones lingüísticas del conjunto de la producción dramática de Torre y Sevil.

En cuanto a la metodología, creemos conveniente señalar que para la marcación de la obra se ha empleado el estándar XML-TEI y se ha seguido la Guía de marcación XML y el Tutorial para la reutilización del código elaborados por Nàdia Revenga García como material complementario a su edición de *La estrella de Sevilla*. En ese sentido, queremos poner en valor la elaboración de estas guías, puesto que su publicación se alinea idóneamente con los principios de las Humanidades Digitales. Así, el hecho de que el código esté en abierto ha permitido que podamos reutilizarlo para la edición en curso de *Valor, ingenio y fineza*, que hubiese resultado mucho más costosa de no estar disponible dicho código. Podemos decir, por ello, que nuestro trabajo se beneficia de las herramientas ya construidas en los grupos o proyectos de investigación. La reutilización del código se ha llevado a cabo empleando el software Oxygen XML Editor, que es aquel para el que está pensada dicha reutilización. Asimismo, también se ha empleado Visual Studio Code para editar o ampliar algunas partes del código (por ejemplo, en aquellas escritas en JavaScript).

En esta ponencia se explicará igualmente el trabajo que se ha realizado con el análisis de redes y cómo se han elaborado las redes lingüísticas basadas en la interacción de los personajes. En ese sentido, se explicarán las decisiones tomadas sobre el cómputo de las interacciones y los procedimientos seguidos para crear las redes. Por ejemplo, se profundizará en la manera en que hemos elaborado las tablas de nodos y matrices de adyacencias, que se han generado en formato CSV para importarlas al programa Gephi, donde hemos empezado a trabajar con ellas. Allí, hemos aplicado una distribución determinada a las redes (en nuestro caso Force Atlas 2) y hemos explorado las distintas opciones para preparar las redes para la visualización.

Para finalizar la exposición, se dará cuenta de los resultados obtenidos que se centran en la edición en curso en web y seis redes sobre las interacciones lingüísticas de las distintas obras de Francisco de la Torre y Sevil. La edición en curso pone de manifiesto la capacidad del código de Revenga García para ser reutilizado al tiempo que da cuenta de que la edición de *Valor, ingenio y fineza*, aunque en curso, es ya un hecho, lo cual no es baladí porque, por una parte, nos pone en el camino de editar una obra desconocida hasta hace bien poco y que pudo haber escrito un dramaturgo cuya producción teatral no ha estado cuidada desde el punto de vista de la edición y de los estudios críticos. Por otra parte, esa edición es, como hemos señalado, una edición digital, y pretendemos que esté integrada no solo por la estructura que Revenga García proponía junto con el contenido de la obra que editamos, sino que creemos que dicha estructura

puede ampliarse y enriquecerse, de manera que la edición contenga también referencias a métricas provenientes del análisis de redes.

Por lo que respecta a las redes, hemos de decir que las interacciones lingüísticas que visualizamos en estas y los patrones que se repiten en ellas de una manera más o menos constante nos permiten reflexionar sobre el conjunto de la producción de Torre y Sevil y hacer una hipótesis, aunque provisoria, sobre la similitud entre el conjunto de las obras de atribución segura al dramaturgo y *Valor, ingenio y fineza*. En concreto, el análisis de las interacciones lingüísticas en las redes nos ha mostrado que existen ciertos patrones en dichas redes, que tienen que ver con qué personajes interactúan más y qué tipos de personaje son los que más interactúan.

Lo expuesto anteriormente permite afirmar que este trabajo se enmarca en la línea de las investigaciones que se sirven de las Humanidades Digitales para estudiar el teatro del Siglo de Oro. El propósito de dicho trabajo es, en ese sentido, fundamentalmente humanístico, y es que su fin no es otro que arrojar luz sobre una parcela del teatro áureo (en concreto el de Francisco de la Torre y Sevil y la obra atribuida a dicho autor, *Valor, ingenio y fineza*). Además, partimos de presupuestos humanísticos para la elaboración de las redes y la edición, pues cuestiones como las decisiones sobre cómo establecer el cómputo de las intervenciones de los personajes o cómo fijar la estructura del documento XML-TEI no responden únicamente a razones propias del medio digital, sino que forman parte de un modelado humanístico de los datos; esto es, del juicio con el que hemos procedido a lo largo del trabajo. Así, creemos que dicho fundamento humanístico justifica el empleo y la elaboración de recursos digitales, como puede ser determinado software (Oxygen XML Editor, Gephi) o los productos fruto de la investigación (la edición en curso y las redes). Todo ello nos lleva a recordar que las Humanidades Digitales no solo han de tener presente lo humanístico en sus premisas, sino también en sus métodos, conclusiones y productos, por lo que será necesario ser coherente con el modelo de datos en abierto, pues dichos datos no solo han de ser reutilizables, sino también accesibles. Con ello queremos poner en valor el trabajo que presentamos, ya que en un futuro pretendemos poner al servicio de la comunidad investigadora y del usuario en general la edición web completa, que integrará también elementos provenientes del análisis de redes. Asimismo, también deseamos que al menos una parte de los datos (la relativa a la edición) sea reutilizable, de manera que pueda servir a futuros investigadores que estén interesados en elaborar ediciones digitales de teatro del Siglo de Oro. No hay que olvidar, por otro lado, la labor interdisciplinar que subyace a todo trabajo de Humanidades Digitales, y que pone en contacto los métodos humanísticos tradicionales con los informáticos, de manera que se genera una suerte de simbiosis donde ambas partes se benefician de la unión. Así pues, consideramos que el trabajo presentado responde a dicha interdisciplinariedad. Para concluir, quisiéramos señalar que la edición en curso tiene como objetivo formar parte del corpus de ediciones digitales de teatro del Siglo de Oro, contribuyendo a ampliar el conjunto

de obras editadas y, de esa manera, enriqueciendo el panorama, tanto por lo que tiene que ver con la ecdótica en el sentido tradicional como por lo relacionado con la aportación a las Humanidades Digitales centradas en el teatro del Siglo de Oro.

Figuras, tablas, mapas o gráficas

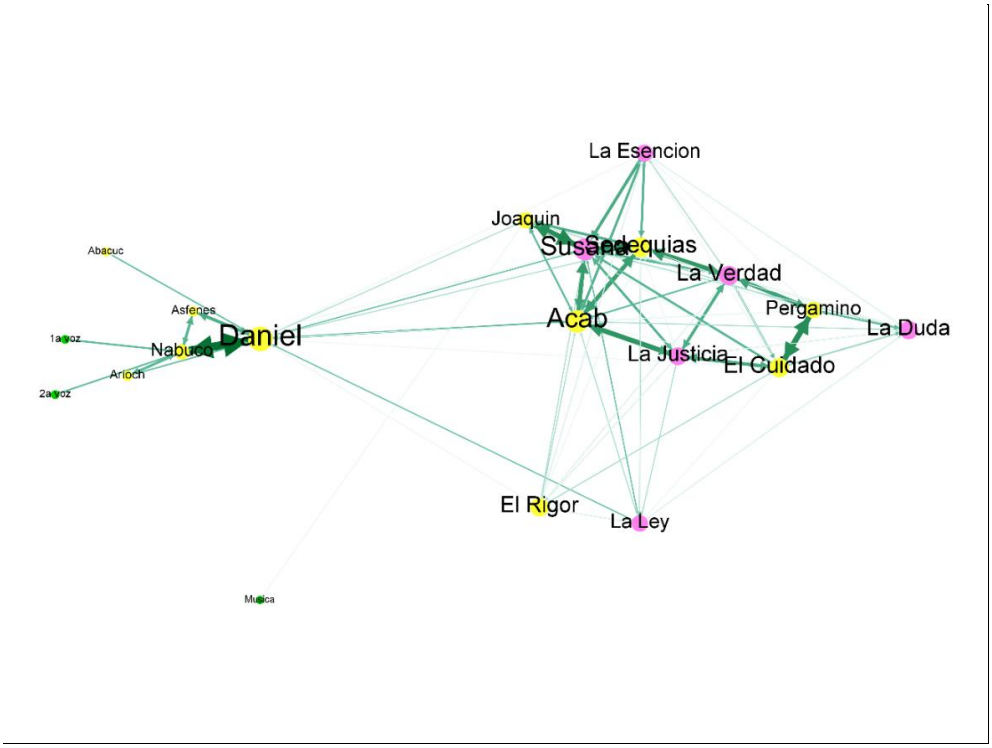


Figura 1. Red de interacciones lingüísticas de *La justicia y la verdad*

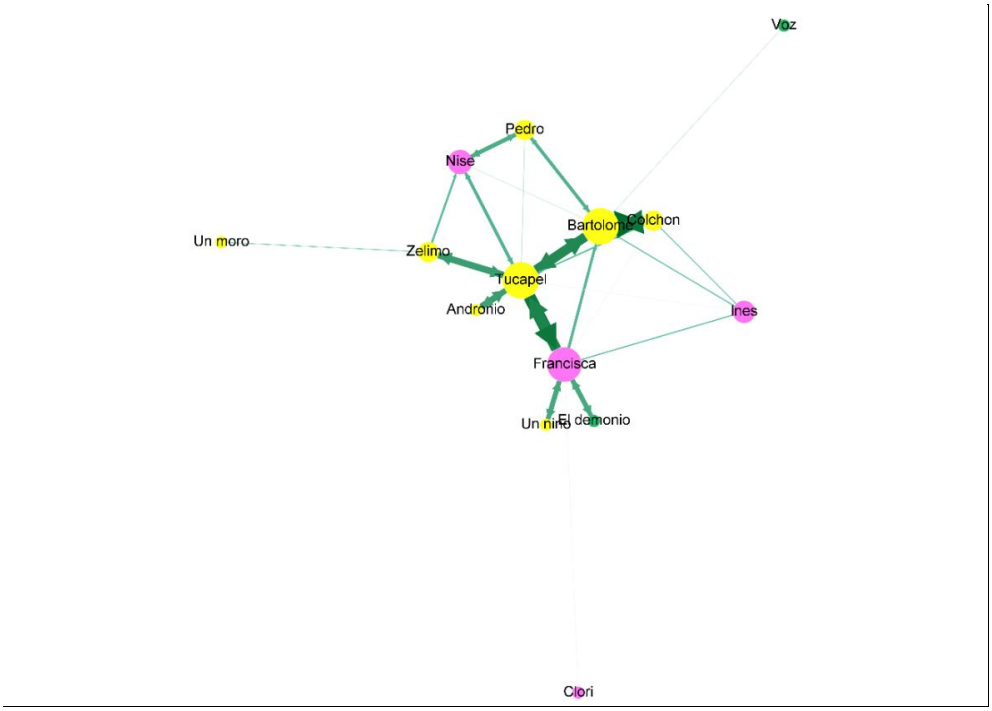


Figura 2. Red de interacciones lingüísticas de *La confesión con el demonio*

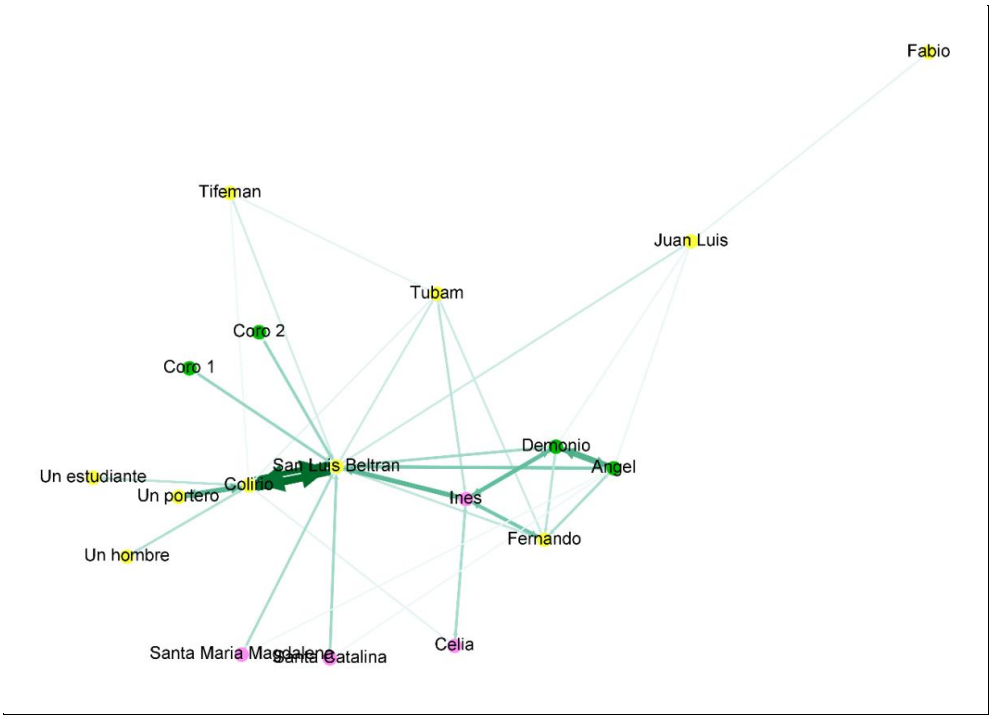


Figura 3. Red de interacciones lingüísticas de *La batalla de los dos. Comedia de san Luis Beltrán*

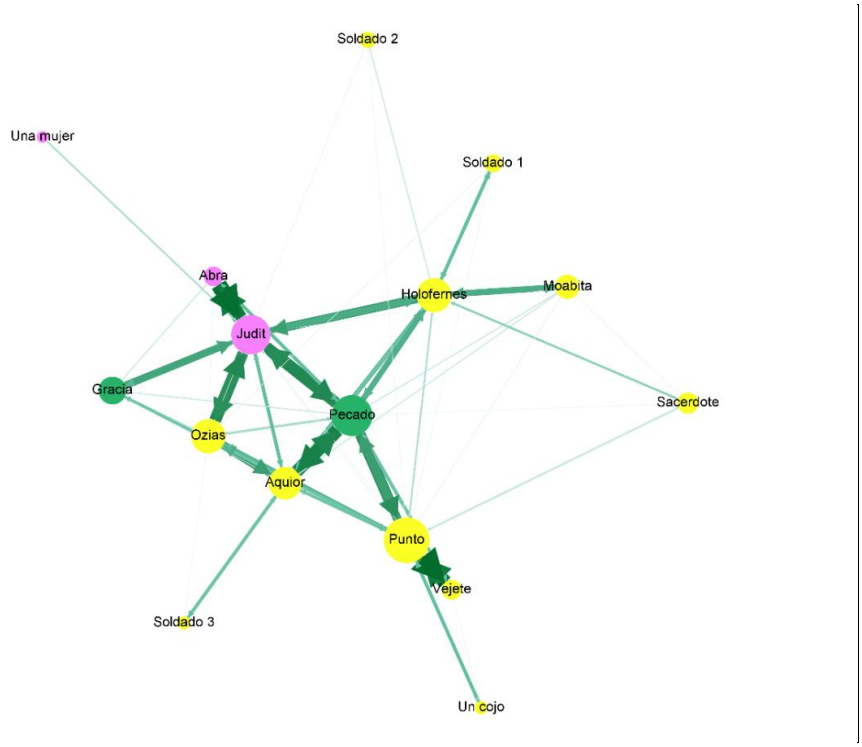


Figura 4. Red de interacciones lingüísticas de *Triunfar antes de nacer*

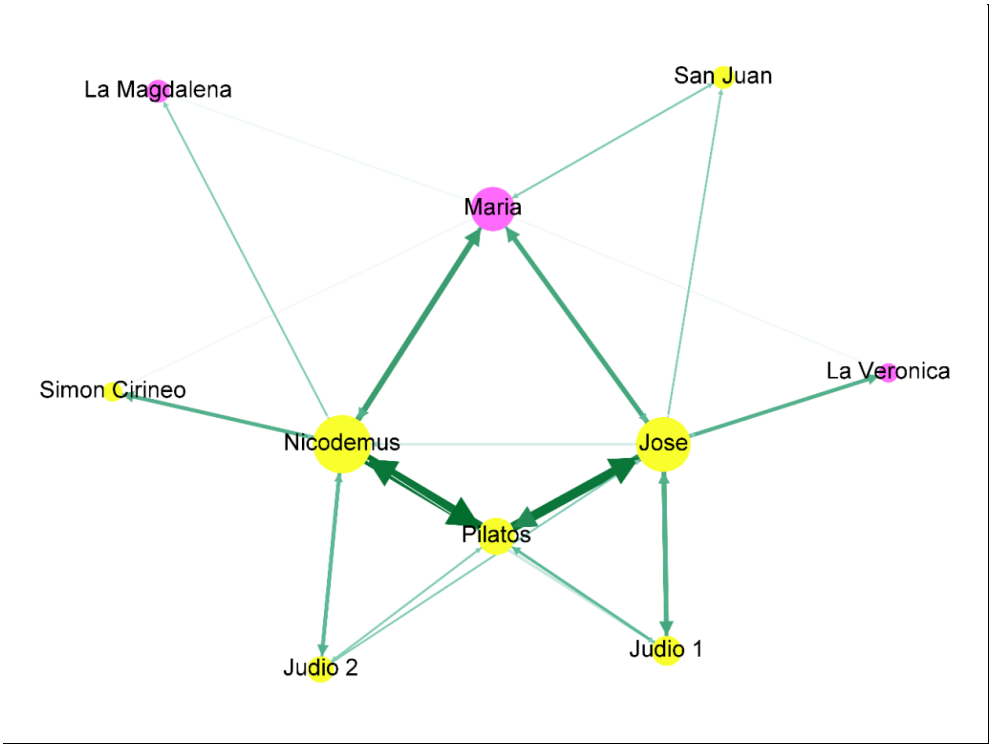


Figura 5. Red de interacciones lingüísticas de *El descendimiento de la cruz*

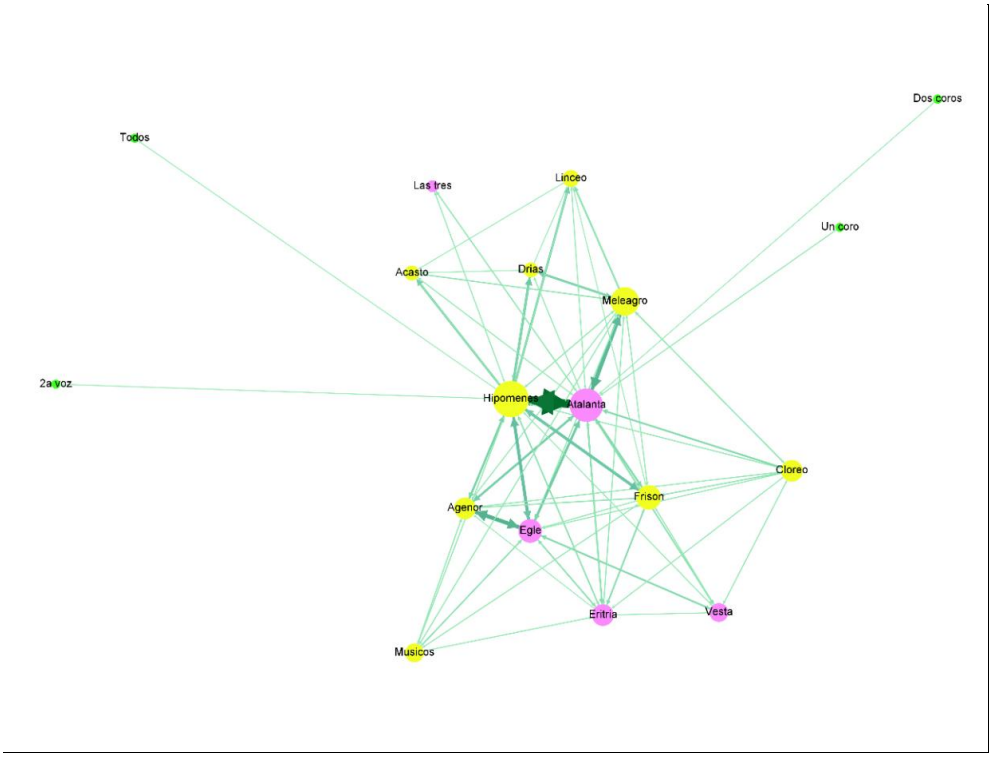


Figura 6. Red de interacciones lingüísticas de *Valor, ingenio y fineza*

Bibliografía

MERINO RECALDE, D. (2023a), «Análisis de redes sociales de personajes teatrales (parte 1)», *Programming Historian en español*, 7. https://doi.org/10.46430/phes0064

MERINO RECALDE, D. (2023b), «Análisis de redes sociales de personajes teatrales (parte 2)», *Programming Historian en español*, 7. https://doi.org/10.46430/phes0065

REVENGA GARCÍA, N. (s. f.), *La Estrella de Sevilla. Edición digital*, recurso en línea. https://regarna.blogs.uv.es/

TORRE Y SEVIL, F. de la (1664), *La justicia y la verdad. Comedia de Susana*, Valencia, Gerónimo Vilagrasa.

TORRE Y SEVIL, F. de la (1678), «La confesión con el demonio», en *Comedias escogidas, xliv. Parte cuarenta y cuatro de comedias nuevas, nunca impresas, escogidas de los mejores ingenios de España*, Madrid, Roque Rico de Miranda, pp. 197–242. En línea: http://bdh.bne.es/bnesearch/detalle/bdh0000073419 [Acceso: 16/07/2025].

TORRE Y SEVIL, F. de la (s. a.), *La batalla de los dos. Comedia de San Luis Beltrán. Primera parte de su vida*, en línea: http://bdh.bne.es/bnesearch/detalle/bdh0000216900 [Acceso: 16/07/2025].

TORRE Y SEVIL, F. de la (s. a.), «Triunfar antes de nacer», en *Poesías castellanas varias*, vol. 2, fols. 5r–78r. En línea: http://bdh.bne.es/bnesearch/detalle/bdh0000174000 [Acceso: 16/07/2025].

TORRE Y SEVIL, F. de la (s. a.), *Valor, ingenio y fineza*, s. p.

TEI CONSORTIUM (eds.) (s. f.), *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, recurso en línea, versión [número], última modificación [fecha]. http://www.tei-c.org/Guidelines/P5/ [Acceso: 16/07/2025].

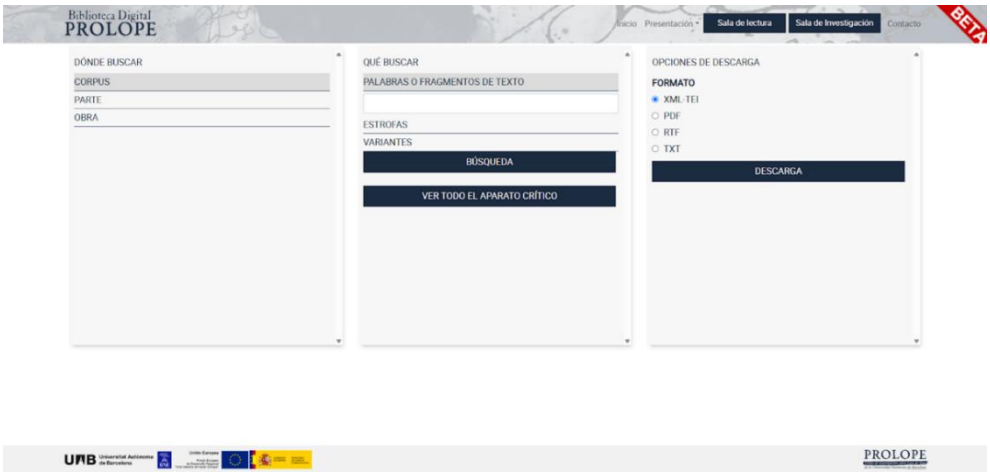
Una estrategia sostenible para la edición crítica digital del teatro del Siglo de Oro: el caso de PROLOPE

Anna Abate
Università di Trento
Emanuele Leboffe
David Merino Recalde
Universitat Autònoma de Barcelona

Introducción

Desde su concepción, la Biblioteca Digital PROLOPE ha tenido como objetivo difundir en formato digital las ediciones críticas del teatro de Lope de Vega realizadas por el grupo PROLOPE de la Universitat Autònoma de Barcelona. En más de tres décadas de labor filológica, el grupo ha publicado más de 200 comedias, consolidándose como referente internacional en los estudios teatrales áureos y la crítica textual (Valdés 2014; Pontón 2022). Su versión beta (2024) permite consultar ediciones en XML-TEI con aparato crítico, notas explicativas y funcionalidades de búsqueda avanzadas orientadas al análisis ecdótico y a la reutilización de datos.

El modelo adoptado en la primera fase del proyecto apostó por la codificación detallada de las ediciones críticas, lo que permite la búsqueda avanzada en múltiples niveles y convierte la Biblioteca en una auténtica herramienta de investigación textual orientada al análisis ecdótico que va más allá de un archivo de ediciones digitales.



Vista de la ‘Sala de investigación’, espacio dedicado al buscador avanzado.

No obstante, la complejidad de este enfoque —que exige una codificación manual en XML-TEI y un tratamiento minucioso del aparato crítico— dificulta la incorporación masiva de nuevas ediciones a la plataforma.

Para garantizar la sostenibilidad del proyecto y acelerar la disponibilidad digital del repertorio de PROLOPE, se ha impulsado una estrategia complementaria de codificación semiautomática, con el fin de ofrecer a los editores filológicos un instrumento para la generación de archivos XML-TEI sin conocimientos técnicos avanzados. Este artículo presenta sus fundamentos prácticos y editoriales, así como los primeros resultados en el marco de la Biblioteca Digital PROLOPE.

Estado de la cuestión

En una fase preliminar se analizaron herramientas de codificación automatizada aplicadas al teatro en verso, como *txt2tei* (Sanz-Lázaro 2023) y *EzDrama* (Skorinkin 2024), ambas basadas en símbolos insertos en el texto y su posterior transformación a XML-TEI mediante scripts Python. Sin embargo, ninguna cubre todos los elementos necesarios para una edición crítica en PROLOPE. Además, el proceso de revisión y validación, al requerir verificación manual de los símbolos, incrementa el margen de error y el tiempo de edición.

También son precedentes relevantes los proyectos ARTELOPE (Oleza 2013; Muñoz Pons 2013) y Gondomar Digital (Presotto 2022), que implementan sistemas de codificación semiautomática de obras del Siglo de Oro para su publicación digital. Aunque ambos convierten los textos a XML-TEI desde sus bases de datos, difieren en la preparación de los archivos: ARTELOPE recurre a etiquetas simbólicas en textos planos ({P}, {AC}, {V}), luego enriquecidos mediante FileMaker; Gondomar Digital, en cambio, usa OpenOffice con estilos de párrafo, notas al pie y comentarios.

Metodología

Ante esta situación, se valoró ampliar modelos abiertos existentes mediante un script en Python que mantuviera la lógica simbólica, añadiendo nuevos elementos. Sin embargo, la proliferación de símbolos dificultaba el trabajo editorial y agravaba los problemas de revisión. Además, las soluciones de ARTELOPE y Gondomar Digital, desarrolladas ad hoc por informáticos, estaban ligadas a infraestructuras técnicas propias, lo que dificultaba su adopción externa.

Dado que ninguna de estas alternativas respondía adecuadamente a nuestras necesidades, optamos por explorar una vía propia, basada en nuestras rutinas editoriales ya consolidadas: los editores entregan sus materiales para la publicación en papel en archivos Word separados (prólogo, texto crítico, aparato, notas, etc.), estructurados mediante una plantilla DOCX con algunos estilos de párrafo personalizados para diferenciar sus distintas secciones.

A partir de esta premisa, nuestra propuesta consiste en asignar estilos de párrafo específicos a cada uno de los elementos textuales del texto crítico. Estos estilos, predefinidos en una plantilla DOCX, permiten marcar con precisión las unidades textuales del drama: *Verso* o *Partido_inicial/medio/final* para los versos, *Prosa* para los pasajes no versificados, *Personaje* para los nombres de los hablantes, *Acot* para las acotaciones escénicas y *ACTO* para los encabezados de acto. También se incluyen estilos específicos para el reparto (*DRAMATIS_LIST*), los epígrafes (*Epigr_final*, *Epigr_Dramatis*, etc.), las citas y otros elementos estructurales.

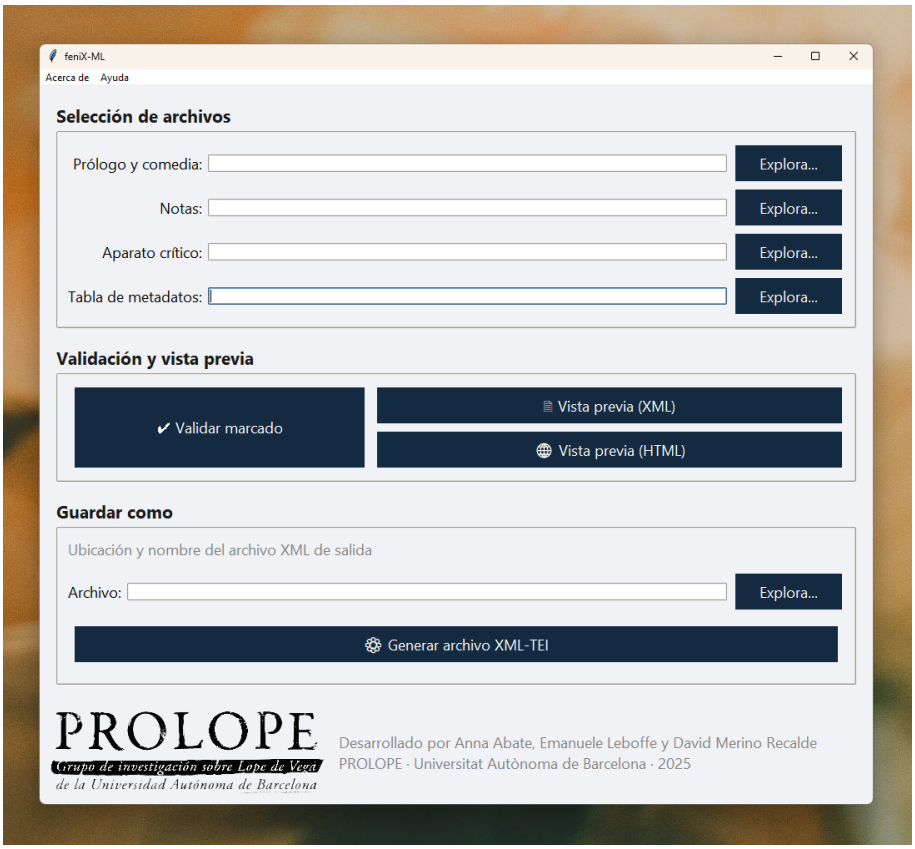
Captura de pantalla de los estilos creados para marcar el texto crítico.

El sistema de símbolos, sin embargo, no se ha eliminado por completo, sino que se mantiene de forma residual para casos puntuales, como la identificación de secciones del prólogo (#) o la notación de formas estróficas ($redondilla). Por su parte, las notas explicativas y variantes se registran en un archivo separado, asociadas al número de verso correspondiente. Para posibilitar su transformación automática a XML, se ha optado por una necesaria simplificación del modelo original de marcado para las notas filológicas y el aparato crítico: en lugar de usar estructuras diferenciadas como <app>, <lem>, <rdg> y <note>, ambos se codifican mediante este último elemento, distinguido semánticamente mediante el atributo @subtype, con valores ‘nota’ o ‘aparato’. Esta solución, aunque se aparta del modelo ortodoxo propuesto por las *Guidelines TEI*, ha permitido un equilibrio razonable entre fidelidad estructural y viabilidad operativa. Todos estos elementos (estilos, símbolos, notas y variantes) son reconocidos, vinculados y transformados automáticamente en un único archivo XML gracias a un script en Python, lo que acelera de manera significativa el proceso de elaboración de la edición digital.

Resultados

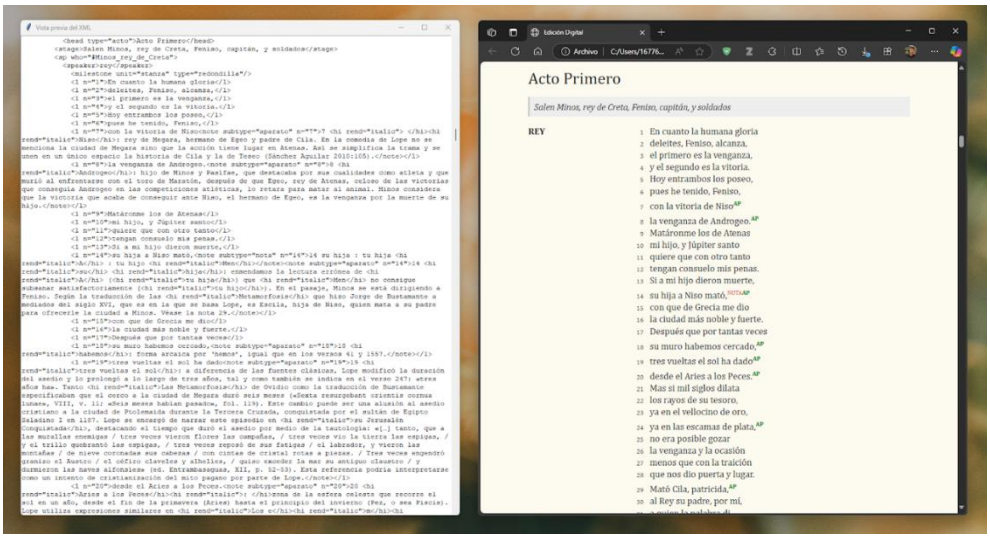
Actualmente se está finalizando el desarrollo del *script* en Python para la transformación automatizada, acompañado por un ejecutable con interfaz gráfica (GUI). Esta herramienta, bautizada “feniX-ML” en honor a Lope de Vega, está concebida para facilitar la accesibilidad del procedimiento incluso a usuarios sin experiencia técnica, permitirá llevar a cabo de forma intuitiva tanto la conversión como la revisión de los archivos XML generados a partir de los materiales editoriales.

La herramienta permite cargar directamente los archivos con símbolos y estilos definidos, y generar con un solo clic el archivo XML-TEI correspondiente. Incorpora, además, un sistema de validación que detecta posibles inconsistencias, como la presencia de texto sin estilo aplicado en el texto crítico o errores en el formato de las notas. Este sistema responde directamente a los problemas detectados en los modelos analizados en la fase de revisión.



Captura de pantalla de la herramienta en desarrollo.

Otra funcionalidad destacada es la previsualización, que permite al usuario consultar el archivo XML-TEI resultante y visualizar una edición generada en HTML gracias a la integración de la librería JavaScript CETEIcean y una hoja de estilos personalizada. Estas previsualizaciones facilitan la detección temprana de errores antes de proceder a la exportación definitiva del archivo XML, especialmente para usuarios poco familiarizados con XML.



Captura de pantalla de las dos previsualizaciones del XML-TEI generado, previo a la exportación.

De cara a la Biblioteca Digital PROLOPE, el equipo de investigación está realizando el marcado de las ediciones ya publicadas, aunque el objetivo es que cada editor crítico aplique directamente los estilos al preparar su edición, incorporando desde el inicio los requisitos necesarios para la conversión digital. La herramienta está concebida tanto para los editores que colaboran con PROLOPE como para cualquier investigador interesado en adaptar sus propias ediciones críticas a este modelo semiautomático, ya que se acompaña de una guía detallada y de plantillas DOCX que permiten una implementación autónoma. El ejecutable, los scripts, así como las guías y plantillas en Word, están en acceso abierto en GitHub.

Conclusiones

La estrategia adoptada busca un equilibrio razonable entre la aplicación rigurosa de modelos teóricos y las necesidades concretas del proyecto, apostando por una mayor sostenibilidad del flujo de trabajo sin renunciar a la calidad científica. Este enfoque ha permitido al grupo alcanzar autonomía técnica y establecer un *iter* operativo que facilita la incorporación continua de nuevas obras a la Biblioteca Digital. En una primera fase se transformarán todas las ediciones críticas ya publicadas por PROLOPE; una vez completada, será posible avanzar hacia una codificación más exhaustiva del aparato crítico y activar todas las funcionalidades avanzadas del sistema.

En definitiva, este modelo, basado en herramientas desarrolladas por el propio grupo de investigación, apuesta por una lógica de escalabilidad y mejora continua, donde un flujo de trabajo claro y el control interno del proceso editorial constituyen los pilares de un proyecto científico sostenible y en evolución constante.

Bibliografía

ITALIA, P. (2020), *Editing Duemila. Per una filologia dei testi digitali*, Roma, Salerno Editrice.

LEBOFFE, E. (en revisión), «La codificación en XML-TEI del teatro español del Siglo de Oro: problemas, soluciones, ortodoxia y heterodoxia», *Journal of the Text Encoding Initiative*.

MUÑOZ PONS, C. (2013), «El proceso de edición digital en Artelope y CTCE», *Teatro de Palabras: Revista sobre Teatro Áureo*, 7, pp. 483–496.

PIERAZZO, E. (2016), *Digital Scholarly Editing: Theories, Models and Methods*, London, Routledge.

DEL RIO RIANDE, G. (2024), «Acerca de la sostenibilidad, archivo y preservación de proyectos de edición digital de textos hispánicos», *Philologia Hispalensis*, 38(2), pp. 41–57. https://doi.org/10.12795/PH.2024.v38.i02.02

SANZ-LÁZARO, F. (2023), «Del fonema al verso: una caja de herramientas digitales de escansión teatral», *Revista de Humanidades Digitales*, 8, pp. 74–89. https://doi.org/10.5944/rhd.vol.8.2023.37830

SKORINKIN, D. (2024), «EasyDrama: A lightweight solution for encoding plays in TEI-XML», en G. DEL RIO RIANDE (ed.), *Text, Languages and Communities. Vigésimo cuarto Congreso de la Text Encoding Initiative*, Buenos Aires, pp. 38–39. https://doi.org/10.5281/zenodo.13883242

OLEZA, J. (2013), «A description of ARTELOPE: A database containing information, plots, and texts related to Lope de Vega's theater», *Journal for Early Modern Cultural Studies*, 13(4), pp. 150–151. https://doi.org/10.1353/jem.2013.0041

PONTÓN, G. (2022), «Editar el teatro de Lope de Vega: de la práctica al método (y viceversa)», *Ecdotica*, 19, pp. 103–118. https://doi.org/10.7385/107685

PRESOTTO, M. (2022), «L'edizione del teatro classico spagnolo e le tecnologie digitali», en N. BENEDETTO, S. GRECO, & P. LASKARIS (eds.), *Saberes humanísticos, ciencia y tecnología en la investigación y la didáctica del hispanismo*, Roma, AISPI Edizioni, pp. 150–164. https://cvc.cervantes.es/literatura/aispi/biblioteca_07.htm

VALDÉS GÁZQUEZ, R. (2014), «Anhelos, realidades y sueños ante la perspectiva y urgencia de la edición crítica digital. Reflexiones desde un grupo de investigación», *Anuario Lope de Vega. Texto, Literatura, Cultura*, 20. https://doi.org/10.5565/rev/anuariolopedevega.86

Proyecto SCORE: investigación musicológica, edición digital y modelo ontológico a partir de la correspondencia de Kurt Schindler (1882-1935)[1]

Sara Escuer Salcedo
Universidad Internacional de La Rioja –
GIR IHMAGINE (Universidad de Salamanca)

Palabras clave: Kurt Schindler; cartas; redes; edición digital; ontología
Keywords: Kurt Schindler; letters; networks; digital publishing; ontology

La correspondencia conservada del musicólogo, director y compositor Kurt Schindler (Berlín, 1882; Manhattan, 1935) se encuentra dispersa en diversos archivos y bibliotecas de España y EE. UU., según se recoge en diversas publicaciones de la investigadora Matilde M. Olarte Martínez, una de las máximas especialistas en la figura de Schindler y en los proyectos llevados a cabo por musicólogo alemán en España en las décadas de 1920 y 1930, financiados por la Universidad de Columbia. Los archivos y bibliotecas que albergan dichos fondos documentales son los siguientes: en Manhattan, la Biblioteca Pública de Nueva York, la Hispanic Society of America y el Instituto Hispánico de la Universidad de Columbia; en España, la Fundación Juan March, el Archivo Manuel de Falla, la Biblioteca de Catalunya y la Biblioteca Valenciana (Olarte Martínez 2012, 556).

Son en total más de 730 cartas que ofrecen una ventana al pensamiento de sus redactores, a sus métodos de trabajo, sus intereses y preocupaciones expresados en primera persona. Este gran corpus documental representa una fuente primaria de gran importancia para contextualizar la vida musical española en el primer tercio del siglo XX, así como para trazar redes de contactos y analizar el impacto de todo ello en la musicología y la etnomusicología de dicha centuria.

En 2010, la investigadora María Enriqueta Frontera alertó sobre el grave estado de deterioro en el que se encontraba, ya entonces, el archivo personal de Kurt Schindler, a causa de la desidia y el desinterés, en algunos casos, de las instituciones que lo custodian (Frontera Zunzunegui 2010). A esta situación se suma el alto interés que presentan un conjunto de cartas de la magnitud de las conservadas en los archivos anteriormente citados, de cuyo contenido se puede

[1] La propuesta y resultados que aquí se presentan forman parte del recién depositado Trabajo fin de máster «Proyecto SCORE: la correspondencia de Kurt Schindler como fuente de estudio musicológico desde las Humanidades Digitales» realizado en el marco del Máster en Humanidades Digitales de la Universidad Internacional de La Rioja, y cuya defensa se prevé para el mes de julio (Escuer Salcedo 2025).

inferir valiosa información en relación con la personalidad del propio Schindler, acerca de sus relaciones con otros investigadores, compositores o artistas de su época, sobre la actividad musical que llevó a cabo en sus primeros años de estancia en EE. UU. y los posteriores trabajos de campo desarrollados en la España rural de las décadas de 1920 y 1930. En definitiva, el conjunto de cartas localizadas, son una fuente de gran valor documental para la contextualización de la vida musical, la musicología y la etnomusicología en España en particular, pero también en Europa y Norteamérica (Olarte Martínez 2012; Olarte Martínez y Montoya Rubio 2012).

El actual estado de conservación de las fuentes primarias, sumado al interés que su estudio puede suponer para la comunidad científica, no solo en el ámbito musicológico, sino también etnográfico, antropológico o sociológico, pone de manifiesto la necesidad de abordar cuanto antes un trabajo de recuperación, edición digital y difusión en acceso abierto en la web semántica de esta documentación con el fin de facilitar su estudio y su difusión, a nivel científico, académico y social.

Las más de 730 cartas localizadas –escritas o recibidas por Kurt Schindler– se localizan en los siguientes archivos y bibliotecas de EE. UU. y España:

- The New York Public Library. Music Division, Sección: Arts Division del Lincoln Center. Fondo: Kurt Schindler's Papers 1882-1946 (Manhattan, EE. UU.): 694 cartas.
- Biblioteca y Archivo del Hispanic Institute at Columbia University (Manhattan, EE. UU.): 16 cartas.
- Hispanic Society of America (Manhattan, EE. UU.): 3 cartas.
- Fundación Juan March (Madrid, España): 6 cartas.
- Archivo Manuel de Falla (Granada, España): 16 cartas.
- Biblioteca Valenciana (Valencia, España): al menos 3 cartas.
- Biblioteca de Catalunya (Barcelona, España): 1 carta.

A nivel metodológico, este trabajo combina investigación musicológica tradicional con herramientas propias de las humanidades digitales para realizar un estudio transdisciplinar que se ha desarrollado en tres fases:

- FASE 1: En la primera fase del trabajo se ha llevado a cabo la edición digital de una selección de cartas, mediante Lenguaje de Marcado Extensible (XML-TEI) y la posterior visualización de la edición digital de dichas cartas de forma simultánea a los documentos digitalizados –algunos de ellos inéditos hasta el momento–, facilitados para esta investigación por la autora de las fotografías, Matilde M. Olarte Martínez.
- FASE 2: En segundo lugar, se ofrece un prototipo de ontología que permita visualizar las relaciones entre los diferentes elementos presentes en los documentos (personas, lugares, instituciones, publicaciones, trabajos de campo, eventos, etc.).
- FASE 3: Las tareas desarrolladas en las fases 1 y 2 han desembocado en la creación de un portal web en el que se ofrece, por una parte, la edición digital de las cartas junto a los documentos originales digitalizados, y por otra, la explicación del modelo ontológico desarrollado.

Resultados

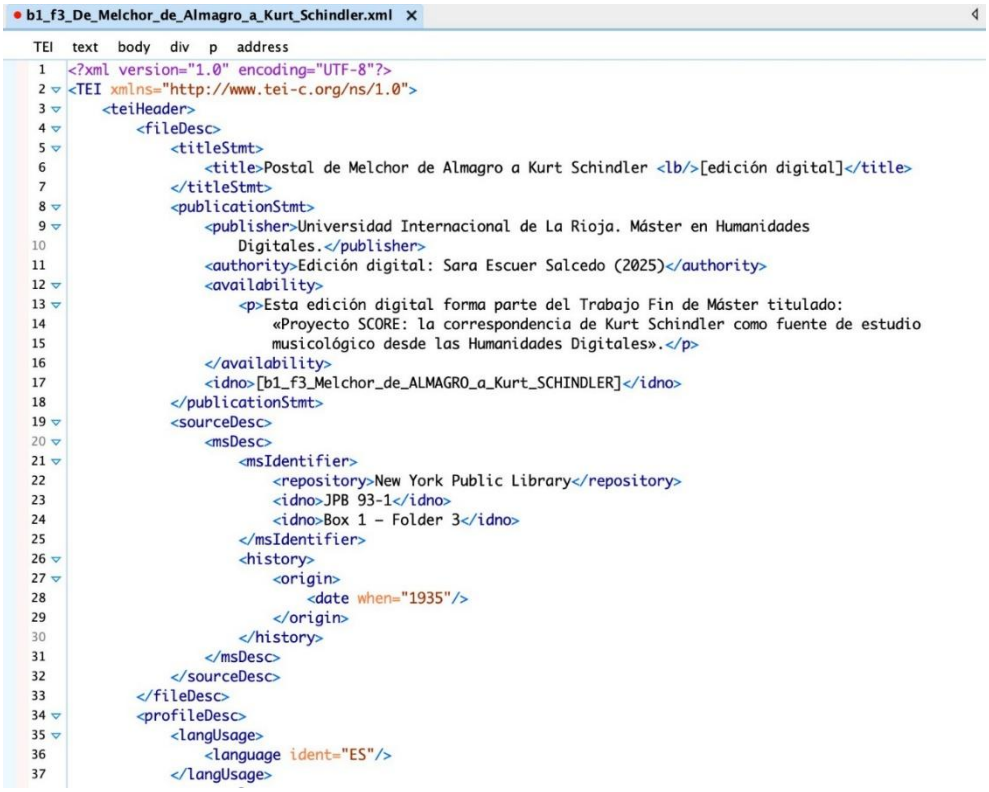
EDICIÓN CRÍTICA DIGITAL

Se han editado un total de 20 de las 694 cartas conservadas en The New York Public Library. Music Division, Sección: Arts Division del Lincoln Center. Fondo: Kurt Schindler's Papers 1882-1946. En la siguiente tabla se ofrecen algunos datos identificativos de las mismas, así como el enlace URL para acceder a la edición digital de las mismas.

Signatura NYPL	Remitente URL (XML) URL (visualización)	Destinatario	Fecha (dd/mm/aaaa)
b1 f3 (MAI 5143)	Melchor de Almagro	Kurt Schindler	07/03/1935
	https://unirnet-my.sharepoint.com/b1_f3		
b1 f4 (MAI 5146)	Norberto Almandoz Mendizábal	Kurt Schindler	14/04/1920
	https://unirnet-my.sharepoint.com/b1_f4		
b1 f5 (MAI 5149)	Marguerite D'Álvarez	Kurt Schindler	__/__/1920
	https://unirnet-my.sharepoint.com/b1_f5		
b2 f30 (MAI 5394)	Caridad Rodríguez Castellano	Kurt Schindler	11/02/1932
	https://unirnet-my.sharepoint.com/b2_f30		
b1 f52 (MAI 5394)	Delfín González	Kurt Schindler	31/03/1923
	https://unirnet-my.sharepoint.com/b1_f52		
b1 f70 (MAI 5526)	Aurelio de Llano	Kurt Schindler	25/08/1932
	https://unirnet-my.sharepoint.com/b1_f70 https://proyectoscore.org/edicion-critica-digital/b2-f70/		
b2 f76 (MAI 5546)	Pilar de Madariaga	Kurt Schindler	22/12/1931
	https://unirnet-my.sharepoint.com/b2_f76 https://proyectoscore.org/edicion-critica-digital/b2-f76/		
b2 f88a (MAI 5562)	Francesch de Millet	Kurt Schindler	22/10/1930
	https://unirnet-my.sharepoint.com/b2_f88a		
b2 f88b (MAI 5562)	Lluis Millet	Kurt Schindler	__/08/1917
	https://unirnet-my.sharepoint.com/b2_f88b		
b2 f94a (MAI 5569)	Federico de Onís	Kurt Schindler	29/01/1932
	https://unirnet-my.sharepoint.com/b2_f94a		
b2 f94b (MAI 5569)	Federico de Onís	Kurt Schindler	15/02/1932
	https://unirnet-my.sharepoint.com/b2_f94b		
b2 f102a (MAI 5577)	Merle Proctor	Kurt Schindler	28/06/1930
	https://unirnet-my.sharepoint.com/b2_f102a		
b2 f110a (MAI 5585)	Ángel del Río	Kurt Schindler	22/10/1930

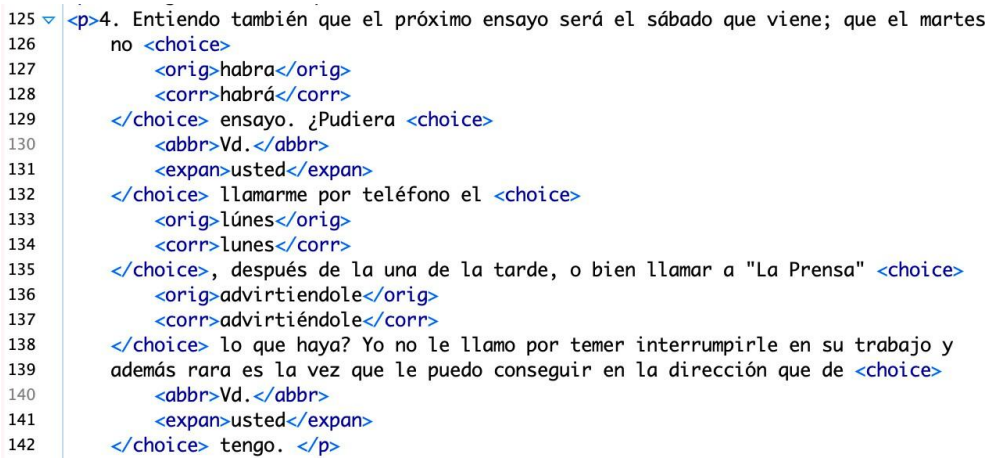
Tabla 1. Relación de cartas editadas digitalmente

Se muestran a continuación algunos extractos representativos del trabajo de edición digital realizado.



```
b1_f3_De_Melchor_de_Almagro_a_Kurt_Schindler.xml
TEI  text  body  div  p  address
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
    <teiHeader>
        <fileDesc>
            <titleStmt>
                <title>Postal de Melchor de Almagro a Kurt Schindler <lb/>[edición digital]</title>
            </titleStmt>
            <publicationStmt>
                <publisher>Universidad Internacional de La Rioja. Máster en Humanidades
                    Digitales.</publisher>
                <authority>Edición digital: Sara Escuer Salcedo (2025)</authority>
                <availability>
                    <p>Esta edición digital forma parte del Trabajo Fin de Máster titulado:
                        «Proyecto SCORE: la correspondencia de Kurt Schindler como fuente de estudio
                        musicológico desde las Humanidades Digitales».</p>
                </availability>
                <idno>[b1_f3_Melchor_de_ALMAGRO_a_Kurt_SCHINDLER]</idno>
            </publicationStmt>
            <sourceDesc>
                <msDesc>
                    <msIdentifier>
                        <repository>New York Public Library</repository>
                        <idno>JPB 93-1</idno>
                        <idno>Box 1 - Folder 3</idno>
                    </msIdentifier>
                    <history>
                        <origin>
                            <date when="1935"/>
                        </origin>
                    </history>
                </msDesc>
            </sourceDesc>
        </fileDesc>
        <profileDesc>
            <langUsage>
                <language ident="ES"/>
            </langUsage>
```

Figura 1. Encabezado TEI (fragmento). Fuente: Oxygen XML Editor 27.0. Carta b1_f3_De_Melchor_de_Almagro_a_Kurt_Schindler (elaboración propia).



```
<p>4. Entiendo también que el próximo ensayo será el sábado que viene; que el martes
    no <choice>
        <orig>habra</orig>
        <corr>habrá</corr>
    </choice> ensayo. ¿Pudiera <choice>
        <abbr>Vd.</abbr>
        <expan>usted</expan>
    </choice> llamarme por teléfono el <choice>
        <orig>lúnes</orig>
        <corr>lunes</corr>
    </choice>, después de la una de la tarde, o bien llamar a "La Prensa" <choice>
        <orig>advirtiendole</orig>
        <corr>advirtiéndole</corr>
    </choice> lo que haya? Yo no le llamo por temer interrumpirle en su trabajo y
    además rara es la vez que le puedo conseguir en la dirección que de <choice>
        <abbr>Vd.</abbr>
        <expan>usted</expan>
    </choice> tengo. </p>
```

Figura 2. Edición de las cartas (corrección de términos <orig>). Fuente: Oxygen XML Editor 27.0. Carta b1_f3_De_Melchor_de_Almagro_a_Kurt_Schindler (elaboración propia).

```
105  and in the evening to Mallorca- Did you get "home to <placeName><choice>
106          <sic>Haca</sic>
107          <corr>Jaca</corr>
108      </choice></placeName>" happily -but doesn't the <placeName>hotel La
109      Paz</placeName> seem a little too peaceful without us? I hope at least a
110  dozen widows all bedecked with dollars and kiddies arrive to study there for the
111      <time>summer</time>. If not -join us in <placeName>Paris</placeName> July
112  4th- or better still in <placeName>New York</placeName> not too much later.</p>
113 .oser>
114   <salute>Thanks mille jois for your charming and inimitable guidance in Piedrás
115      gorges and <placeName>Jaca</placeName>'s fascinations… somewhere along the
116      path you won the love of both.</salute>
```

Figura 3. Marcado semántico: <placeName> y <time>. Fuente: Oxygen XML Editor 27.0. Carta b2_f102a_De_Merle_Proctor_a_Kurt_Schindler (elaboración propia).

Modelo ontológico

El modelo ontológico propuesto se ha realizado a través del *software* Protégé 5.6.4, desarrollado por la Universidad de Standford.

El esbozo de ontología que se aquí presenta se basa en la información contenida en 174 cartas (de las más de 730 localizadas) que fueron recibidas o enviadas por Kurt Schindler entre 1912 y 1935, relacionadas, mayoritariamente, con su actividad profesional. La ontología se ha creado desde cero, al no encontrar ninguna que pudiese servir como modelo. Este prototipo de ontología –ampliable en un futuro– busca representar conceptualmente los elementos que protagonizaron el desarrollo profesional y artístico de Schindler: personas, lugares, eventos científicos o artísticos, trabajos de campo, investigaciones, publicaciones y otros conceptos tratados en dicho corpus documental. Por otra parte, mediante la consulta de bibliografía complementaria o de diferentes bases de datos se ha recopilado información de carácter biográfico de las personas que protagonizan esta ontología, como fecha y lugar de nacimiento o muerte, vinculación profesional y cargos profesionales.

El dominio de esta ontología se enmarca en el ámbito de la investigación etnomusicológica y la práctica artística –especialmente por agrupaciones corales–, así como las redes de intercambio artístico, científico e intelectual entre EE. UU. y España en el primer tercio del siglo XX. El alcance del modelo incluye tanto elementos formales extraídos de la correspondencia (fecha, remitente, destinatario, lugares de envío y de destino) como los contenidos tratados en las cartas, telegramas o cables analizados. Asimismo, el análisis del contenido de la correspondencia ha prestado especial atención a las relaciones entre los diferentes elementos.

Se muestran a continuación algunas evidencias del trabajo realizado.

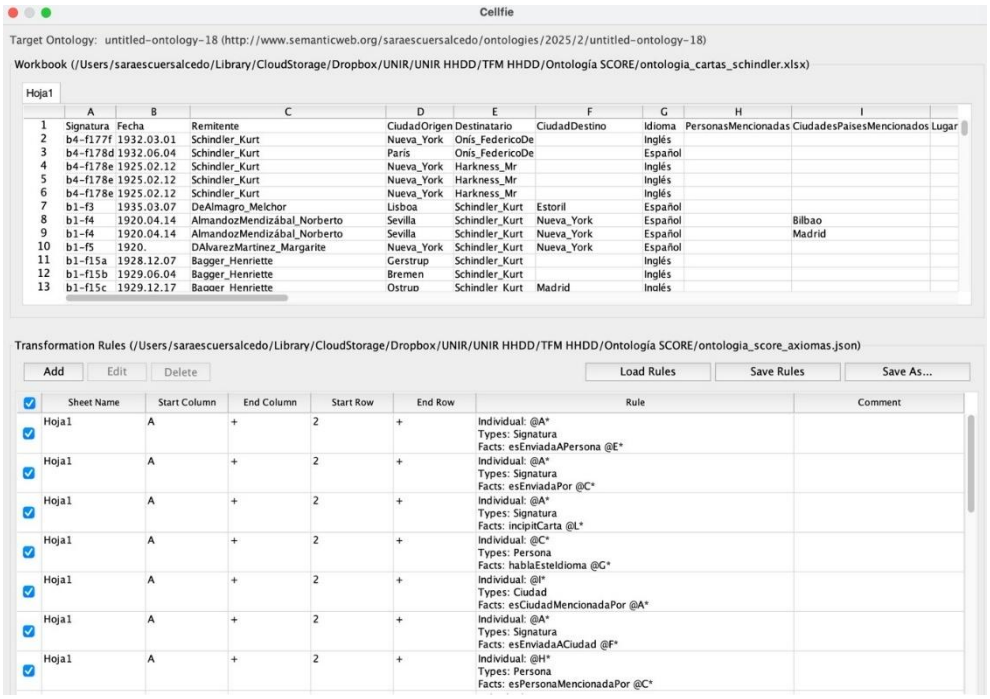


Figura 4. Ontología SCORE. Axiomas (fragmento). Fuente: Protégé. Ontologia_score.ofn (elaboración propia).

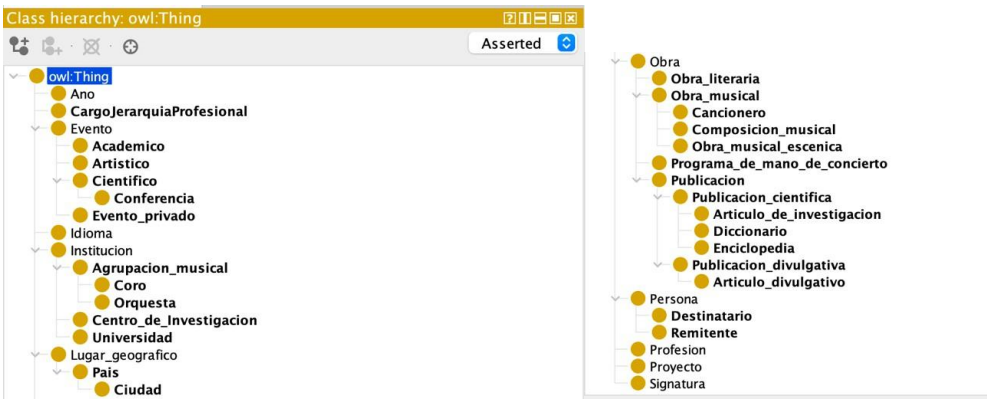


Figura 5. Ontología SCORE. Jerarquía de clases. Fuente: Protégé 5.6.4. Ontologia_score.ofn (elaboración propia).

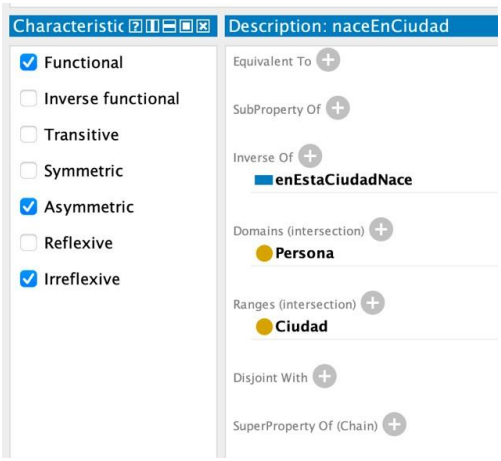
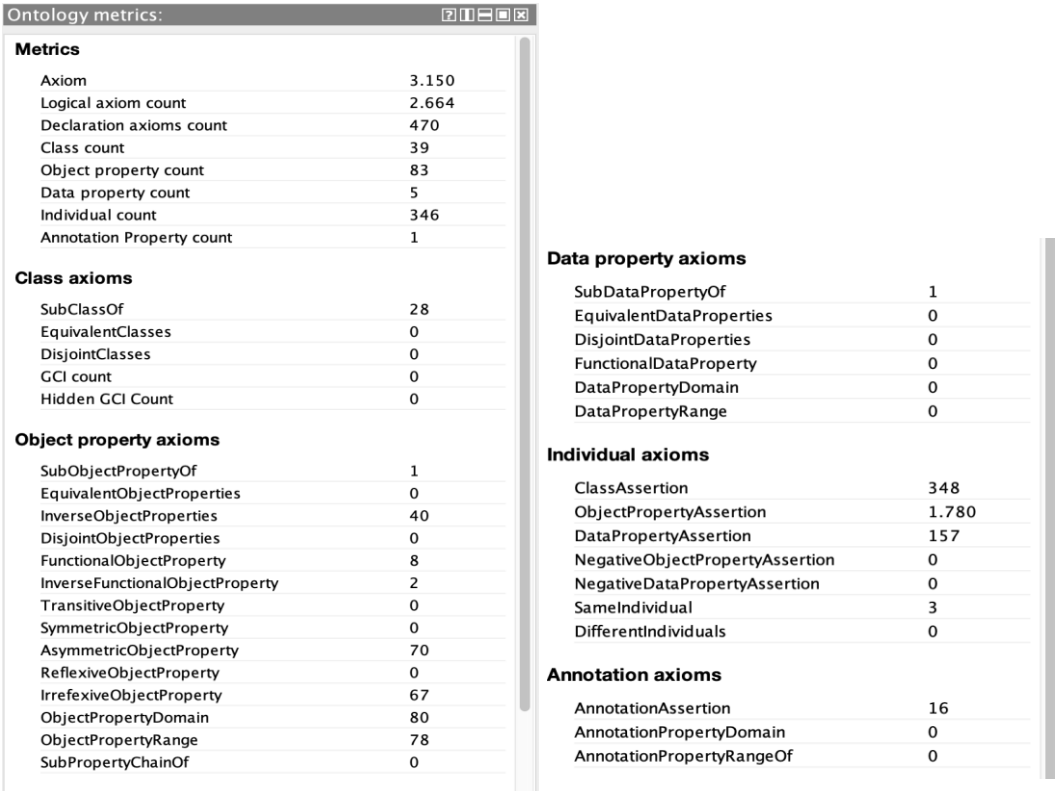


Figura 6. Ontología SCORE. Ejemplo de cardinalidad (1). Fuente: Protégé 5.6.4. Ontologia_score.ofn (elaboración propia).

Figura 7. Ontología SCORE. Propiedades de objeto (fragmento). Fuente: Protégé. Ontologia_score.ofn (elaboración propia).



Ontology metrics:

Metrics

Axiom	3.150
Logical axiom count	2.664
Declaration axioms count	470
Class count	39
Object property count	83
Data property count	5
Individual count	346
Annotation Property count	1

Class axioms

SubClassOf	28
EquivalentClasses	0
DisjointClasses	0
GCI count	0
Hidden GCI Count	0

Object property axioms

SubObjectPropertyOf	1
EquivalentObjectProperties	0
InverseObjectProperties	40
DisjointObjectProperties	0
FunctionalObjectProperty	8
InverseFunctionalObjectProperty	2
TransitiveObjectProperty	0
SymmetricObjectProperty	0
AsymmetricObjectProperty	70
ReflexiveObjectProperty	0
IrrefexiveObjectProperty	67
ObjectPropertyDomain	80
ObjectPropertyRange	78
SubPropertyChainOf	0

Data property axioms

SubDataPropertyOf	1
EquivalentDataProperties	0
DisjointDataProperties	0
FunctionalDataProperty	0
DataPropertyDomain	0
DataPropertyRange	0

Individual axioms

ClassAssertion	348
ObjectPropertyAssertion	1.780
DataPropertyAssertion	157
NegativeObjectPropertyAssertion	0
NegativeDataPropertyAssertion	0
SameIndividual	3
DifferentIndividuals	0

Annotation axioms

AnnotationAssertion	16
AnnotationPropertyDomain	0
AnnotationPropertyRangeOf	0

Figura 8. Ontología SCORE. Métricas. Fuente: Protégé. Ontologia_score.ofn (elaboración propia).

Web del proyecto SCORE

Con la finalidad de ofrecer los resultados digitales de este trabajo fin de máster, se ha creado un portal web en WordPress. Teniendo en cuenta que este trabajo se ampliará en el futuro, previsiblemente al amparo de un grupo de investigación vinculado a una universidad, se optó inicialmente por un plan económico para dar visibilidad de forma provisional a los resultados obtenidos hasta el momento. Si bien el plan contratado ofrece algunas limitaciones, se considera que cumple con los objetivos indicados en relación con la exposición de este trabajo fin de máster.

La web del proyecto SCORE (*Schindler's Correspondence*, *Ontology, Research and Ethnomusicology*) está estructurada en tres grandes apartados: presentación del proyecto, edición digital y ontología (figura 61).

Figura 9. *Página de inicio del proyecto SCORE.* Fuente: Web provisional del proyecto SCORE (https://proyectoscore.org) (elaboración propia)

Bibliografía

ESCUER SALCEDO, S. (2025), «Proyecto SCORE: la correspondencia de Kurt Schindler como fuente de estudio musicológico desde las Humanidades Digitales», *Trabajo Fin de Máster*, Universidad Internacional de La Rioja.

FRONTERA ZUNZUNEGUI, M. E. (2010), «El archivo personal de Kurt Schindler: una propuesta de organización», *Etno-folk: Revista de Etnomusicología*, 16–17, pp. 15–34. https://gredos.usal.es/handle/10366/83446

FRONTERA ZUNZUNEGUI, M. E. (2014), *Una figura señera en el campo de la música tradicional española en el primer tercio del siglo XX: el polifacético músico alemán Kurt Schindler (1882–1935)* [Tesis doctoral], Universidad de Salamanca. https://gredos.usal.es/handle/10366/125418

LOZANO TELLO, A. (2020), «Ontologías en la Web Semántica», en *I Jornadas de Ingeniería Web*, pp. 1–4. http://eolo.cps.unizar.es/docencia/MasterUPV/Articulos/Ontologias%20en%20la%20Web%20Semantica.pdf

OLARTE MARTÍNEZ, M. (2012), «La correspondencia inédita de Kurt Schindler como una fuente directa para contextualizar la vida musical del primer tercio del siglo XX», en M. OLARTE MARTÍNEZ (ed.), *Fuentes documentales interdisciplinares para el estudio del Patrimonio y la Oralidad en España*,

Baiona, Dos Acordes, pp. 553–652. https://gredos.usal.es/handle/10366/153581

OLARTE MARTÍNEZ, M., & MONTOYA RUBIO, J. C. (2012), «Cuando los epistolarios eran redes sociales. Aproximación a las conexiones de Kurt Schindler a través de sus corresponsales», en M. OLARTE MARTÍNEZ (ed.), *Fuentes documentales interdisciplinares para el estudio del Patrimonio y la Oralidad en España*, Baiona, Dos Acordes, pp. 652–682. https://gredos.usal.es/handle/10366/153584

W3C (2004), «Vista general del Lenguaje de Ontologías Web (OWL)», recurso en línea. https://www.w3.org/2007/09/OWL-Overview-es.html#s1.2 [Acceso: 09/04/2025].

Lo que aprendimos sobre el procesamiento de entrevistas de historia oral en español chileno para TF-IDF y análisis de sentimientos

Nicole Larrondo
Laboratorio de Patrimonio Documental y Humanidades Digitales – Pontificia Universidad Católica de Chile
Eduardo Cerda
Diseñador de Software

Palabras clave: Historia oral; Analítica cultural; Procesamiento de lenguaje natural; TF-IDF; Análisis de sentimientos
Keywords: Oral history; Cultural analytics; Natural language processing; TF-IDF; Sentiment analysis

Introducción

El presente trabajo propone una exploración metodológica basada en herramientas de analítica cultural computacional para complementar los análisis cualitativos de entrevistas de historia oral, enfocándose en la Villa Olímpica de Santiago de Chile. Esta investigación surge de la necesidad de diversificar los enfoques interpretativos tradicionales y aprovechar las potencialidades de los métodos de lectura lejana en el campo de las humanidades digitales. A través de quince entrevistas a residentes de la Villa Olímpica, se ilustra cómo técnicas de procesamiento de lenguaje natural (PLN) —en particular Term Frequency - Inverse Document Frequency —identificación de términos de alto peso léxico— (en inglés, TF-IDF) y análisis de sentimientos —clasificación de polaridades emocionales— pueden aportar nuevas perspectivas al estudio de las memorias barriales, detectando patrones discursivos que escapan al procesamiento manual.

Problema de investigación

La historiografía oral en Chile ha profundizado en las narrativas individuales y colectivas de los barrios, permitiendo reconstruir procesos sociales, políticos y arquitectónicos desde la experiencia de quienes los habitaron. No obstante, los marcos metodológicos comunes tienden a reproducir una estructura reiterativa: contextualización histórica y arquitectónica, presentación de testimonios seleccionados y análisis cualitativo centrado en pasajes ejemplares. Esta práctica limita la capacidad de identificar sistemáticamente tendencias

discursivas y variaciones semánticas en grandes volúmenes de texto. Asimismo, las tecnologías de PLN están diseñadas, en su mayoría, para el inglés y presentan sesgos al aplicarlas a variedades no hegemónicas como el español chileno. Por ello, este estudio se propone responder a dos preguntas centrales:

1) ¿Cómo integrar métodos computacionales en el análisis de entrevistas de historia oral en español chileno sin sacrificar sus matices culturales y sociales?; y

2) ¿Qué nuevos hallazgos narrativos y temáticos emergen de la aplicación combinada de TF-IDF y análisis de sentimientos en estos testimonios?

Metodología

Selección y transcripción de testimonios

Se trabajó con un corpus de quince entrevistas a residentes actuales o antiguos de la Villa Olímpica, obtenidas mediante criterio de diversidad generacional y tiempo de residencia. Las entrevistas se realizaron durante la pandemia el año 2021 bajo protocolos sanitarios y fueron conducidas por un equipo de campo local con experiencia en trabajo comunitario. Esta decisión obedeció a las restricciones de movilidad y a la necesidad de mantener la calidad del diálogo.

Procesamiento de audio a texto

Todas las grabaciones se procesaron con el paquete SpeechRecognition en Python, configurado para la variante “es-CL” (español chileno). Esta configuración permitió capturar modismos locales, entonaciones y pausas características. Seguidamente, se aplicó una etapa de limpieza y normalización textual: eliminación de ruido (símbolos no reconocidos, puntuación excesiva), corrección manual de errores de transcripción y normalización ortográfica para homogeneizar formatos.

Preparación del corpus para PLN

Una vez normalizado, el texto fue convertido a un formato plano y segmentado en unidades de discurso correspondientes a respuestas individuales. Se conservó la identificación del entrevistado y el contexto de la pregunta para análisis diferenciado.

Técnicas de analítica cultural utilizadas

1. **TF-IDF (scikit-learn)** Se calculó la frecuencia de términos en cada documento individual y se contrastó con su frecuencia en el corpus completo, de modo de identificar palabras relevantes en contextos específicos. Los vectores resultantes fueron normalizados y visualizados en gráficos de barras y nubes de palabras.

2. **Análisis de sentimientos (Tulio Chilean Spanish BERT)** Se empleó el modelo BERT preentrenado para el español chileno, desarrollado por Ortiz Fuentes (2023), para clasificar las respuestas en positivo, negativo o neutral. Este modelo ha sido entrenado con textos en español de Chile (noticias, web, reclamos y tweets) y libros en español. Se agrupó “muy positivo” con “positivo” y “muy negativo” con “negativo” para simplificar la interpretación.

Instrumento de recolección

El cuestionario incluyó 33 preguntas abiertas y semiabiertas, organizadas en cuatro categorías analíticas: a) antecedentes personales y vínculos familiares; b) percepciones emocionales sobre la Villa Olímpica; c) experiencias y memorias espaciales; y d) prácticas de ciudadanía y regímenes de gusto. Estas categorías respondieron a los objetivos de la tesis doctoral relacionados con la performance ciudadana en la clase media urbana.

Resultados y discusión

Análisis de sentimientos

El resultado principal evidenció una valoración predominantemente positiva de la experiencia residencial en la Villa Olímpica, especialmente en referencias a espacios colectivos (plazas, canchas, áreas verdes). El 68% de las respuestas se clasificó como positivo y el 15% como negativo, mientras que el 17% restante correspondió a neutrales. Los testimonios sobre el arribo inicial se evaluaron mayormente como negativos, con términos como “peladero”, “desolada” y “faltante de servicios”, lo que refleja la percepción de precariedad de los primeros años.

TF-IDF

Previo al análisis automático, se estableció un conjunto de palabras clave asociadas al concepto de modernización —sinónimos y términos afines como “progreso” y “desarrollo” — para explorar las tensiones entre las expectativas estatales de modernidad, las experiencias reales de los vecinos y las representaciones / ansiedades socioeconónmicas. Esta lista de palabras claves se integró al corpus para evaluar la frecuencia, contexto de uso y carga semántica de estos términos y para confirmar o rectificar la tesis del uso de estos térmimos por parte de los vecinos entrevistados como parte de su imaginario. Al contrario de lo esperado, el análisis destacó términos específicos de alto peso en entrevistas individuales, tales como “comité”, “olla común” y “junta de vecinos”, que no figuraban de igual modo en análisis cualitativo manual. Estos hallazgos apuntan a la relevancia de la autogestión barrial como eje discursivo central. La diversidad de vocabulario en distintos grupos generacionales también se hizo evidente: entrevistas de generaciones mayores contenían frecuencias

altas de palabras asociadas a nostalgia y supervivencia, mientras que las de generaciones recientes enfatizan “renovación” y “comunidad”.

Discusión metodológica

Las narrativas orales del corpus presentan una complejidad intrínseca que desafía los supuestos de linealidad y coherencia textual propios del procesamiento automatizado. La presencia de digresiones, silencios prolongados, groserías, muletillas y expresiones emotivas no estructuradas exige algoritmos flexibles capaces de distinguir fragmentos discursivos sin imponer una lógica narrativa artificial. A pesar de estas dificultades, la estrategia de lectura lejana —basada en TF-IDF y análisis de sentimientos— resultó eficaz para revelar patrones globales de contenido, como la preponderancia de términos vinculados a la autogestión barrial y las tensiones generacionales en torno a la modernización. Estos hallazgos cuantitativos sirvieron como marco para la formulación de hipótesis interpretativas, posteriormente sometidas a validación mediante lectura cercana de fragmentos seleccionados.

El proceso de triangulación metodológica evidencia que los resultados obtenidos de TF-IDF y el análisis de sentimientos ganan profundidad cuando se contrastan con la riqueza contextual de los testimonios originales. En particular, la revisión cualitativa ayudó a corregir posibles sesgos de los diccionarios emocionales y a ajustar los umbrales de relevancia para evitar la sobregeneralización de categorías sentimentales. Asimismo, la experiencia demuestra la necesidad de refinar manualmente las listas de palabras vacías y los vocabularios de referencia, atendiendo a modismos y expresiones idiomáticas propias del español chileno.

El empleo de modelos BERT preentrenados para español chileno evidenció ventajas claras en precisión de clasificación de sentimientos frente a modelos genéricos en español. No obstante, persisten retos en la detección de ironías, alusiones culturales y construcciones literarias, subrayando la urgencia de ampliar los datos de entrenamiento con muestras más representativas y anotaciones específicas.

Conclusiones y desafíos futuros

Este estudio demuestra la viabilidad de aplicar PLN a entrevistas de historia oral en español chileno sin sacrificar su riqueza cultural. La integración de TF-IDF y análisis de sentimientos permite identificar patrones discursivos, dimensiones emocionales y marcos temáticos que complementan la investigación cualitativa. Se resalta la necesidad de entrenar y adaptar modelos de PLN a variedades lingüísticas locales para mejorar la precisión analítica.

Desafíos futuros

1. **Ampliación del corpus**: Se propone ampliar el estudio incorporando un mayor número de entrevistas en otros barrios de Santiago y de

Chile, así como testimonios de residentes de diversas generaciones, nacionalidades y procedencias socioeconómicas.

2. **Modelos lingüísticos regionales**: Se plantea desarrollar y entrenar recursos de PLN específicamente en español chileno y sus variantes (jergas juveniles, modismos de zonas periféricas) para mejorar la precisión en el reconocimiento de términos y estructuras sintácticas propias del habla chilena.
   Es necesario ampliar el entrenamiento de modelos de español Chileno con otros tipos de corpus sonoros más allá de los surgidos de escritura formal o informal, pues estos no son representativos del habla cotidiana y vernácula.
3. **Repositorios abiertos**: Se sugiere crear una plataforma colaborativa y de acceso abierto donde se almacenen las transcripciones anotadas, modelos entrenados y metadatos asociados a cada entrevista. Este repositorio posibilitará la reproducibilidad de los análisis, el desarrollo de proyectos derivados por otros investigadores y la integración de estas fuentes en iniciativas de preservación digital de la memoria oral.
4. **Ética y co-construcción**: Este desafío se centra en diseñar protocolos éticos participativos que involucren a las comunidades entrevistadas en cada etapa de la investigación.

Gráficos

a) Análisis de Sentimiento

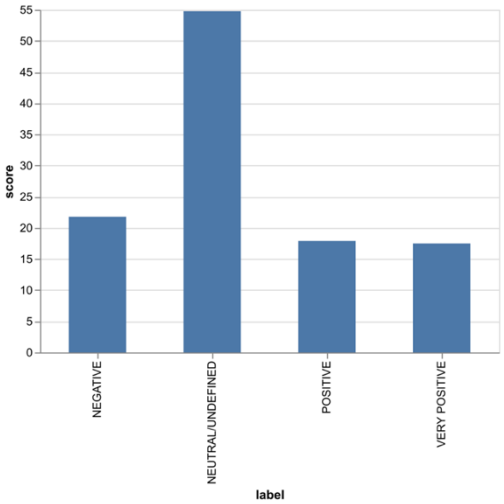


Figura 25. Total Sentimientos.

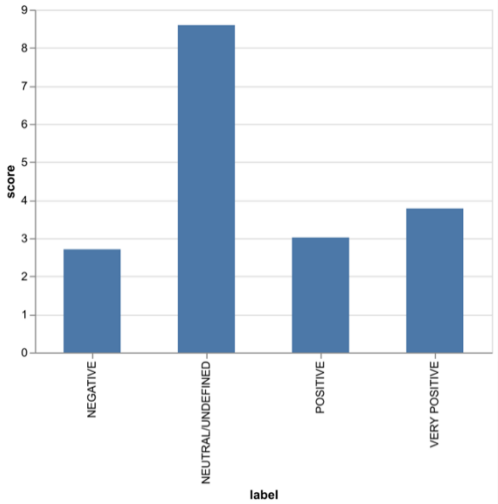


Figura 26. Total Lugar favorito.

Ejemplos puntuales

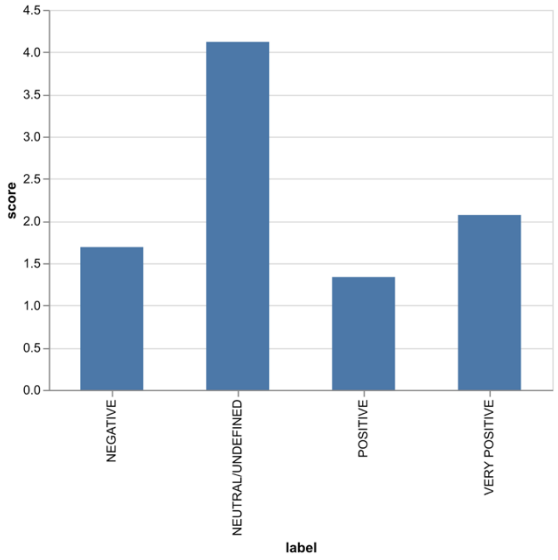


Figura 27. Respuestas pregunta 26

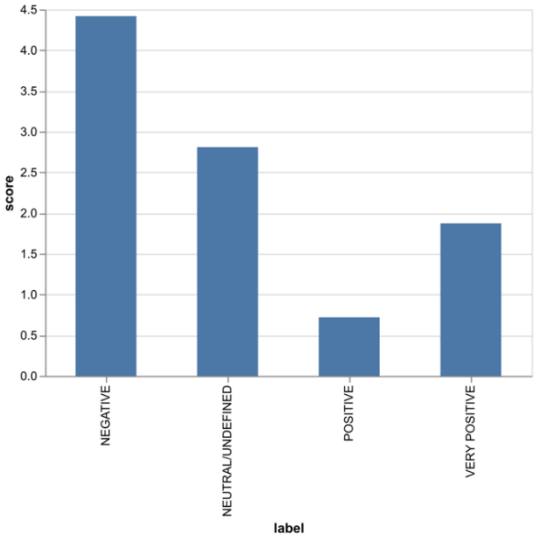
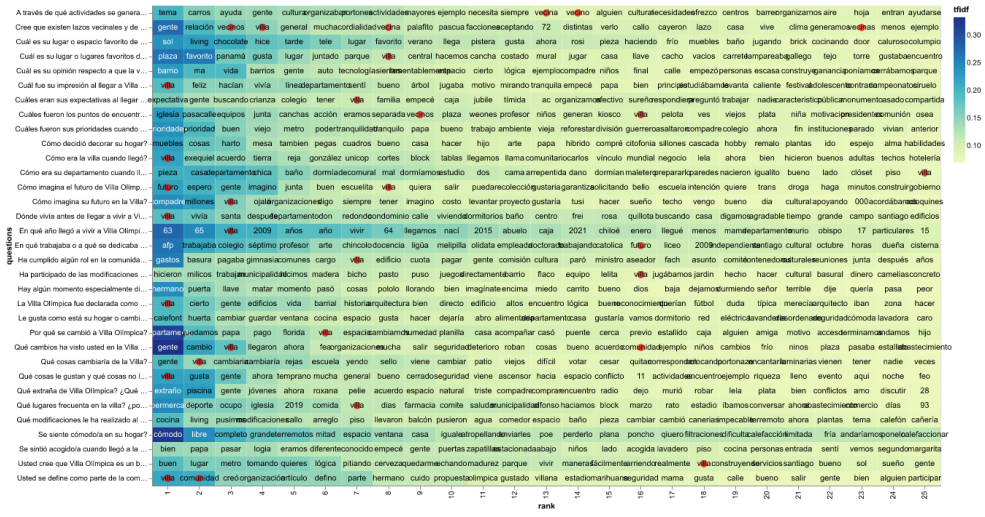


Figura 28. Respuestas pregunta 5.

b) TF-IDF



Heat map de las 33 preguntas y respuestas de cada vecino/a.

Bibliografía

ALBRECHT, J., RAMACHANDRAN, S., & WINKLER, C. (2021), *Blueprints for Text Analytics Using Python: Machine Learning-Based Solutions for Common Real-World (NLP) Applications*, Sebastopol, O'Reilly Media.

ASTORGA, I. (2020), «ignacio-ave/beto-sentiment-analysis-spanish», *Hugging Face*, recurso en línea. https://huggingface.co/ignacio-ave/beto-sentiment-analysis-spanish

RESPONSIBLE DATASETS IN CONTEXT (s. f.), «Home – Responsible Datasets in Context», recurso en línea. https://www.responsible-datasets-in-context.com/ [Acceso: 16/06/2025].

LIU, B. (2010), «Sentiment analysis and subjectivity», en *Handbook of Natural Language Processing*, 2.ª ed., New York.

MAQBOOL, F. (2022, 28 de octubre), «Speech recognition using Python: A beginner's guide», *Stacktrace.Wiki* (blog), recurso en línea. https://www.stacktrace.wiki/speech-recognition-using-python-a-beginners-guide/

MORETTI, F. (2013), *Distant Reading*, Verso. https://www.versobooks.com/en-gb/products/2309-distant-reading

ORTIZ FUENTES, J. (2023, 10 de julio), «dccuchile/tulio-chilean-spanish-bert», *Hugging Face*, recurso en línea. https://huggingface.co/dccuchile/tulio-chilean-spanish-bert

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., *et al.* (2011), «Scikit-learn: Machine learning in Python», *Journal of Machine Learning Research*, 12.

WALSH, M. (2020), «TF-IDF — Introduction to Cultural Analytics & Python», recurso en línea. https://melaniewalsh.github.io/Intro-Cultural-Analytics/05-Text-Analysis/01-TF-IDF.html

Herramientas digitales transmedia y enseñanza ELE: cómo insertar las leyendas en el aula

Carmen Cazorla Vivas
Universidad Complutense de Madrid – IUSMP

Palabras clave: Enseñanza ELE; Literatura; Leyendas; herramientas digitales
Keywords: ELE Teaching; Literature; Legends; Digital Tools

En nuestra propuesta partimos de una primera parte en la que presentamos el proyecto Legendario literario hispánico del siglo XIX: diseño e implantación de una narrativa transmedia (http://descubreleyendas.es/), integrado en la Humanidades digitales. Es el tercer proyecto nacional concedido sobre esta temática y parte de la necesaria conservación de nuestro pasado literario, más concretamente, del patrimonio literario no material concretado en la leyenda literaria, donde se atesoran infinidad de tradiciones, costumbres, y creencias vinculadas a la sabiduría y comunicación popular. A partir de las leyendas, los dos primeros proyectos nacionales se desarrollaron investigando la documentación en la bibliografía literaria y crítica más relevante sobre la leyenda literaria del siglo XIX español y la recopilación de un valioso legendario literario extraído de publicaciones en prensa, libros de viajes, diccionarios geográficos, y crónicas de ciudad y provincia. A partir del proyecto actual se pretende crear diferentes narrativas transmedia (videojuego, podcast, serie para web y otros elementos) para, partiendo de ellos, aplicarlo a diferentes áreas abarcadas por los miembros del proyecto (literatura, creación de videojuegos, enseñanza ELE, rutas culturales...).

Como resultados efectivos y globales de los dos primeros proyectos, podemos hablar, en primer lugar, del propio portal que se ha creado: se trata de una plataforma geolocalizada de leyendas literarias, editadas y anotadas, tras un riguroso trabajo filológico de preservación y facilitación de los textos, con posibilidad de recuperación a través de tesauro de voces, y disponible en la plataforma virtual online de libre acceso (http://descubreleyendas.es/), y que sigue ampliándose, tanto con la introducción de nuevas leyendas como en la inclusión de más metadatos e informaciones sobre cada una de la narraciones; en segundo lugar, el videojuego ‘Caminos Legendarios’, creado a partir de la leyenda de Jimena de Ávila. Como resultado más particular de mi aportación, podemos hablar de dos publicaciones (Cazorla, 2019, y Cazorla y Mainer, 2022) en las que se trabaja con esta leyenda de Jimena de Ávila aplicada a la enseñanza ELE.

En una segunda parte, presentaremos cómo estamos aplicando actualmente el Portal 'Descubre Leyendas' en la integración de este tipo de lecturas en el aula de ELE.

En la actualidad de la enseñanza ELE, desde el enfoque comunicativo, y especialmente desde la publicación del Marco Común Europeo de Referencia (MCER) para las lenguas en el año 2000 (con la adaptación del Marco complementario en 2020) la presencia de la cultura y, específicamente de la literatura, en la enseñanza/aprendizaje de una lengua extranjera se procura trabajar de una manera más integradora en el aula. Casi todos los materiales contienen en sus unidades tareas cada vez más variadas para trabajar con textos literarios. Hoy en día disponemos de herramientas muy variadas con las que se pueden desarrollar y proponer actividades motivadoras, que relacionan la literatura, el arte, la historia, las costumbres o la mitología con aspectos más lingüísticos que permiten trabajar con diferentes destrezas y competencias.

La aproximación a la literatura tiene que ver, en primer lugar, con la comprensión lectora, considerada una de las destrezas que el alumno debe dominar. Pero, además, la literatura, y particularmente las leyendas, nos acercan aspectos socioculturales y nos permiten conocer costumbres, aspectos sociales, personajes históricos relevantes… Por tanto, nos permiten trabajar con los aspectos puramente lingüísticos (como la gramática o el léxico), pero también la aproximación al contexto social y cultural que aporta cada lengua. Y dependiendo de las actividades que se planteen, se puede trabajar también la comprensión auditiva, la expresión oral, la expresión escrita, y la comprensión y expresión audiovisual, tan relacionadas con el proyecto que presentamos y que ha sido, precisamente, una de las más importantes novedades que introdujo el Volumen Complementario del MCER (2021).

Partiendo de las características que favorecen el empleo de textos literarios en el aula de ELE de manera integrada en una programación didáctica, llegaremos a una parte más didáctica para mostrar cómo se podrían utilizar leyendas del siglo XIX en la educación actual empleando herramientas digitales, teniendo en cuenta que estas narraciones facilitan al alumno muestras textuales reales y ricas.

Así, nuestra comunicación pretender mostrar cómo la literatura en general, y las leyendas en particular, pueden ser un vehículo didáctico adecuado para el aula de ELE y eficaz tanto para trabajar la competencia lingüística, el conocimiento de la cultura tradicional o para fomentar la imaginación y creatividad de los alumnos. Y trataremos de demostrar esto con algunas propuestas didácticas concretas que estamos desarrollando en el marco del mencionado proyecto. De esta manera, enmarcamos nuestra comunicación en la posibilidad de emplear herramientas digitales para generar transferencias de conocimiento (literatura, historia, enseñanza, videojuegos, entre otros).

Bibliografía

ALONSO, E., & RODRIGO, V. (2021), *Estrategias de lectura. Leer para comprender*, Madrid, Edelsa-Anaya.

CAZORLA VIVAS, C. (2019), «Las leyendas como recurso didáctico en la enseñanza del español como lengua extranjera (ELE)», en P. VEGA & B. MAINER (coords.), *Lecturas del pasado. Poética y usos culturales de la leyenda literaria*, Madrid, Iberoamericana/Vervuert, pp. 155–173.

CAZORLA VIVAS, C., & MAINER BLANCO, B. (2022), *Las leyendas del siglo XIX desde el siglo XXI: enseñanza ELE y videojuegos*, Valencia, Tirant lo Blanc.

CAZORLA VIVAS, C., & MAINER BLANCO, B. (2025), «Las leyendas como base para un estudio transversal en videojuegos y enseñanza ELE», póster, *II Workshop CLARIAH-CM. Humanidades digitales y ciencia abierta en la era de la inteligencia artificial* (UC3M/CLARIAH-CM, 29–30 mayo 2025). Disponible en: https://zenodo.org/communities/nodoclariahcm/records?q=Cazorla&l=list&p=1&s=10&sort=bestmatch

CONSEJO DE EUROPA (2000/2021), *Marco Común Europeo de Referencia de las Lenguas (MCER). Volumen complementario*.

DIAS LAURIDO, E., & OLIVEIRA DA SILVA MENEZES, N. (2017), «Las leyendas en las clases de español como lengua extranjera: una propuesta a partir del enfoque por tareas», *Revista Comunicación*, 26(1), pp. 57–67.

KOSZLA-SZYMAŃSKA, M. (2014), «Las leyendas como material de apoyo en la enseñanza de la lengua española», en M.ª P. CELMA, M.ª J. GÓMEZ DEL CASTILLO, & S. HEIKEL (eds.), *Actas del XLVIII Congreso de la AEPE. El español en la era digital*, Valladolid, Campus Encuadernaciones. Recurso en línea: https://www.cvc.cervantes.es-ensenanza-biblioteca_ele-aepe-pdf-congreso_48-congreso_48_29.pdf

MENDOZA FILLOLA, A. (2004), «Los materiales literarios en la enseñanza de ELE: funciones y proyección comunicativa», *RedEle*, 1. Recurso en línea: http://www.educacion.es-redele-revista1-mendoza.shtml

TRIGO IBÁÑEZ, E., & ROMERO OLIVA, M. F. (2017), «Geografía literaria en el aula de ELE: una aproximación al Cádiz de 1812 desde la ruta literaria de las libertades», *Studia Iberystyczne*, 16, pp. 283–303.

VEGA RODRÍGUEZ, P., & MAINER BLANCO, B. (eds.) (2019), *Lecturas del pasado. Poética y usos culturales de la leyenda literaria*, Madrid, Iberoamericana/Vervuert.

Hacia un ecosistema FAIR de recursos lingüísticos en gallego: corpus oral *Nos_TranscriSpeech-GL*

Adina Ioana Vladu
Daniel Fernández López
Elisa Fernández Rei
Instituto da Lingua Galega –
Universidade de Santiago de Compostela

Palabras clave: Recursos lingüísticos; corpus oral; gallego; FAIR; tecnologías lingüísticas
Keywords: Linguistic resources; oral corpus; Galician; FAIR; linguistic technologies

Introducción y problema de investigación

La escasez de recursos lingüísticos digitales de calidad y con licencias abiertas es una de las principales limitaciones para el desarrollo de tecnologías lingüísticas en lenguas minorizadas (Ramírez Sánchez y García Mateo, 2022). La creación de estos recursos es costosa y requiere tiempo, coordinación interdisciplinar y una gestión cuidadosa de los aspectos éticos, legales y lingüísticos. En este contexto, nuestro trabajo aborda la creación de *Nos_TranscriSpeech-GL* (Vladu et al., 2023; Vladu et al., 2024), un corpus de audio y texto en lengua gallega elaborado a partir de datos procedentes de fuentes ya existentes, de acuerdo con los principios FAIR (*Findable*, *Accessible*, *Interoperable*, *Reusable*) (Wilkinson et al., 2016).

Partimos de la idea de que es posible generar recursos lingüísticos útiles tanto para la investigación en diferentes ámbitos como para el desarrollo de aplicaciones tecnológicas a partir de datos ya existentes, siempre que se gestione adecuadamente su tratamiento y se diseñe una infraestructura de publicación ajustada a los estándares FAIR (PARTHENOS et al., 2018; De Jong et al., 2018; Filiposka et al., 2024). Este enfoque permite no solo aprovechar el tiempo y esfuerzo invertidos en la consecución de datos para anteriores investigaciones, sino también visibilizar estos datos y proyectos previos, incidiendo en su conservación, su utilidad e, incluso, en su revitalización o nueva proyección/transferencia. En este sentido, nuestro trabajo ofrece una metodología para convertir datos cerrados, fragmentados e infraexplotados en recursos accesibles, visibles y estructurados.

En este marco, la propuesta forma parte de un esfuerzo más amplio de investigación y desarrollo de recursos abiertos, interoperables y auditables (de-

Dios-Flores et al., 2022; Vladu et al., 2022; Külebi et al., 2024), que puedan ser utilizados por distintos perfiles profesionales (lingüistas, ingenieros, docentes, científicos de datos, etc.), en múltiples áreas de conocimiento, y contribuir a la inclusión de lenguas minorizadas en el ámbito de la inteligencia artificial, el reconocimiento y la síntesis del habla y el análisis textual automático.

Metodología

Selección y preparación de los datos

La primera fase del proyecto consistió en la identificación y curación de un conjunto de fuentes orales disponibles y con potencial para ser reaprovechadas con fines lingüísticos y tecnológicos. Estos datos provenían del archivo del Instituto da Lingua Galega (ILG) de la Universidade de Santiago de Compostela (USC) y eran el resultado de su labor de preservación, investigación y difusión del patrimonio lingüístico gallego a lo largo de varias décadas: conferencias, coloquios, actividades de divulgación, fuentes públicas, etc. Aunque respetan los principios éticos de obtención de datos de voz, al haberse generado en actos públicos y con consentimiento implícito para su difusión, estos datos no estaban siendo utilizados debido a diversas limitaciones, como la variación de formato, calidad del audio y lengua empleada, y la falta de metadatos coherentes, de transcripción y de estructura que facilitase su reutilización.

Los criterios de selección de los materiales fueron los siguientes:

- Lengua dominante en la comunicación.
- Calidad mínima del audio.
- Diversidad de estilos, registros y perfiles de hablantes.
- Representatividad de la lengua oral gallega en contextos formales y semiformales.

Los archivos sonoros seleccionados fueron procesados para asegurar un formato homogéneo (*.wav*, 44.1 kHz 16-bit). El corpus resultante se segmentó en cuatro subcorpus, según el género de la intervención: conferencias, entrevistas, discursos y debates. La diversidad temática (lingüística, literatura, filosofía, historia, ciencia, política, cultura popular, etc.) contribuye a la riqueza léxica y estilística del recurso.

Cada archivo fue etiquetado con un código identificador único, que combina un marcador del género de la intervención, un identificador (ID) que representa al hablante y el número de la intervención, permitiendo así una estructuración coherente del corpus. Varios audios correspondientes a un mismo hablante se etiquetaron con números consecutivos. En el caso de que un mismo audio estuviera dividido en varias partes, cada parte se identificó con un número adicional separado del ID del hablante por un guion bajo (p. ex. conf_Alberti1_1.wav).

Transcripción y anotación

Para la transcripción de los archivos sonoros se siguió una guía de anotación detallada, disponible para su consulta en Zenodo. Las principales pautas fueron:

- Transcripción literal ortográfica, respetando la normativa ortográfica oficial del gallego.
- Marcado de las variantes dialectales, formas no normativas y formas léxicas no gallegas.
- Anotación de elementos paralingüísticos y de fenómenos de la oralidad (risa, pausas, interrupciones, fragmentos truncados, texto cantado, etc.).
- Marcado explícito de los turnos de habla, solapamientos, intervenciones del público y otras situaciones comunicativas complejas.
- Normalización de la representación de numerales y siglas (desarrollados y etiquetados).

La transcripción, segmentación y alineación temporal texto-audio se llevaron a cabo manualmente con la herramienta de código abierto Transcriber[1]. Los resultados se entregaron en los formatos *.trs* y *.stm*, junto con ficheros de metadatos sobre los hablantes (sexo, identidad), cuando estos eran conocidos, y de cada archivo (título, duración, fuente, fecha).

Versiones y usos

A partir del corpus completo de audio y texto se generaron dos versiones diferenciadas, con objetivos y niveles de procesamiento distintos:

1. **Versión anotada**. Destinada a la investigación lingüística, puede utilizarse para estudios fonológicos de prosodia, gramaticales de morfología y sintaxis oral, dialectológicos (variación), de pragmática, de análisis del discurso, de didáctica de la lengua, etc. Contiene 53 horas de audio alineadas con transcripciones anotadas en formato *.trs* y *.stm*, acompañadas de metadatos.

2. **Versión sin anotación**. Optimizada para el entrenamiento de modelos de aprendizaje automático, como reconocimiento automático del habla (ASR, por sus siglas en inglés) en gallego, diarización e identificación de hablantes, etc. Esta versión excluye anotaciones lingüísticas y contiene únicamente los audios segmentados en fragmentos de hasta 30 segundos y sus respectivas transcripciones alineadas. A mayores, para asegurar una mayor precisión en la alineación habla-texto, se revisaron las marcas temporales de una parte de los segmentos. Esta versión del corpus cuenta con 50 horas de audio divididas en tres particiones estandarizadas para el entrenamiento de modelos de ASR:

[1] https://trans.sourceforge.net/en/presentation.php

train (80%), *dev* (10%) y *test* (10%). Cada una de estas partes mantiene un equilibrio en la distribución de variables como género del hablante y género textual, sin solapamiento de hablantes entre los conjuntos.

Principios FAIR

Para garantizar la alineación con los principios FAIR, se prestó especial atención a los siguientes aspectos:

- ***Findable*** (Localizable). El corpus está indexado en dos repositorios públicos: Zenodo[2] y Hugging Face[3]. Está pendiente, asimismo, su inclusión en los repositorios de CLARIN y DARIAH[4]. Por otro lado, cada archivo está identificado con una nomenclatura única y coherente, que permite su localización fácil e inequívoca.
- ***Accessible*** (Accesible). Tanto el corpus anotado como la versión sin anotar están disponibles en línea, bajo una licencia permisiva (Creative Commons Attribution 4.0 International).
- ***Interoperable*** (Interoperable). Los formatos de los ficheros siguen estándares reconocidos (*.wav*, *.stm*, *.trs*, etc.), lo que permite la fácil integración de los datos con otras herramientas y plataformas. Además, la inclusión del conjunto de datos en la plataforma Hugging Face permite integrarlo fácilmente en flujos de trabajo de aprendizaje automático, facilitando el entrenamiento y la evaluación de modelos avanzados de procesamiento del lenguaje natural.
- ***Reusable*** (Reutilizable). Además de la licencia abierta que permite la reutilización de los datos, las versiones del corpus se disponibilizan acompañadas de una guía de anotación detallada, una descripción metodológica y documentación sobre el proceso de curación de los datos.

Resultados

El resultado de este trabajo es un corpus de aproximadamente 53 horas de audio en gallego basado en datos reaprovechados de conferencias, entrevistas, debates, discursos, tertulias y otras intervenciones orales que cuenta, por tanto, con una amplia diversidad de estilos discursivos, hablantes y dominios temáticos. Como se comentó en la sección anterior, ofrece dos formatos complementarios y distintas vías de explotación desde diferentes campos de conocimiento.

Ambas versiones están publicadas en repositorios abiertos con DOI, descripciones, guías, metadatos completos y licencia permisiva, lo que garantiza su fácil localización, acceso, interoperabilidad y reutilización, en línea con los

[2] https://doi.org/10.5281/zenodo.7717140.
[3] https://huggingface.co/datasets/proxectonos/Nos_Transcrispeech-GL.
[4] SSH Open Marketplace y Virtual Language Observatory.

principios FAIR. Para garantizar la trazabilidad de los datos, es necesaria una solicitud de acceso y un uso ético que incluya la correspondiente citación.

Así, con la realización de este trabajo se resuelven varios problemas presentes en los datos de origen: 1) a través de la licencia se explicita de manera transparente su uso ético y 2) la estructuración y metadatación corrigen la fragmentación, mejoran su visibilidad y permiten su interoperabilidad. Además, la detallada descripción metodológica facilita la reproducibilidad del proceso.

Nos_TranscriSpeech-GL ha sido ya integrado en el Corpus Oral Informatizado da Lingua Galega (CORILGA) del ILG (USC) de cara a su utilización para la investigación lingüística en diversas disciplinas (desde fonética y gramática hasta análisis del discurso, incluyendo metodologías de variación y cambio lingüísticos). Asimismo, está previsto el uso de *Nos_TranscriSpeech-GL* para entrenar modelos de reconocimiento automático del habla (ASR) en gallego dentro de las líneas de actuación del Proxecto Nós.

Conclusiones

Este trabajo demuestra que la creación de recursos lingüísticos digitales en lenguas minorizadas a partir de datos reaprovechados es no solo viable, sino también replicable y sostenible, siempre que se parta de una infraestructura sólida, se apliquen los principios FAIR y se cuente con equipos interdisciplinares capaces de articular saberes técnicos y lingüísticos.

A través del desarrollo de *Nos_TranscriSpeech-GL*, hemos comprobado que es posible generar corpus multifuncionales a partir de datos ya existentes —procedentes de proyectos anteriores, a veces antiguos o infraexplotados— y darles una nueva vida mediante una curación técnica y documental cuidadosa. Esta estrategia permite ampliar la disponibilidad de recursos en lenguas con escasa representación digital, al tiempo que preserva, visibiliza y reactiva materiales fruto de décadas de actividad investigadora.

La metodología aplicada garantiza no solo la coherencia interna del recurso, sino también su capacidad de ser reproducido, adaptado y ampliado. El diseño del corpus favorece la recuperación automática de información lingüística y metalingüística y lo convierte en un recurso útil para estudios fonológicos, morfosintácticos o del discurso, a la vez que para el entrenamiento de modelos de reconocimiento automático del habla y otras tareas computacionales.

Además, la publicación en formatos abiertos, con licencia permisiva y en plataformas públicas como Zenodo o Hugging Face refuerza el compromiso con una ciencia más abierta, transparente e inclusiva. La adopción de los principios FAIR no se limita al plano técnico: implica una ética del cuidado con los datos y los posibles usos futuros de los recursos, con implicaciones profundas en términos de accesibilidad, equidad, sostenibilidad y gestión y aprovechamiento de recursos públicos.

En definitiva, este trabajo ejemplifica cómo la tecnología puede ser herramienta, objetivo y desencadenante en los procesos de investigación:

herramienta para resolver problemas (como la reutilización de datos), objetivo final en forma de nuevos recursos o aplicaciones, y punto de partida para nuevas preguntas, desafíos y líneas de trabajo futuras en la intersección entre lingüística, ciencia abierta y procesamiento del lenguaje natural.

Bibliografía

DE-DIOS-FLORES, I., MAGARIÑOS, C., VLADU, A. I., ORTEGA, J. E., PICHEL, J. R., GARCÍA, M., GAMALLO, P., FERNÁNDEZ REI, E., BUGARÍN-DIZ, A., GONZÁLEZ GONZÁLEZ, M., BARRO, S., & REGUEIRA, X. L. (2022, 20 de junio), «The Nós Project: Opening routes for the Galician language in the field of language technologies», en *Proceedings of the Workshop Towards Digital Language Equality within LREC 2022*, pp. 52–61. http://www.lrec-conf.org/proceedings/lrec2022/workshops/TDLE/pdf/2022.tdle-1.6.pdf

DE JONG, F., MAEGAARD, B., DE SMEDT, K., FIŠER, D., & VAN UYTVANCK, D. (2018), «CLARIN: Towards FAIR and responsible data science using language resources», en *Proceedings of LREC 2018*, pp. 3259–3264. https://aclanthology.org/L18-1515

FILIPOSKA, S., MISHEV, A., ANASTASOPOULOU, A., FRONTINI, F., PEDONESE, G., & VAN DER LEK, I. (2024, 20 de septiembre), *FAIR-by-Design training for the CLARIN community*, Zenodo. https://doi.org/10.5281/zenodo.13816183

HOLLANDER, H., MORSELLI, F., UITERWAAL, F., ADMIRAAL, F., TRIPPEL, T., & DI GIORGIO, S. (2018), *PARTHENOS guidelines to FAIRify data management and make data reusable*, Zenodo. https://doi.org/10.5281/zenodo.2668479

KÜLEBI, B., HERNÁEZ, I., FERNÁNDEZ REI, E., MONTOYO, A., SOLITO, S., ARMENTANO-OLLER, C., HERNANDO, J., NAVAS, E., MAGARIÑOS, C., VLADU, A., SARATXAGA, I., SÁNCHEZ, J., GARCÍA ROMILLO, V., HERRANZ, A., SOUGANIDIS, C., GARCÍA, N., MOSCOSO SÁNCHEZ, A., REGUEIRA, X. L., DUBERT, F., & GUTIÉRREZ, Y. (2024), «Speech technologies in the ILENIA Project: Generating resources to develop voice applications in the official languages of Spain», en *Proceedings of IberSPEECH 2024*, pp. 289–292. https://doi.org/10.21437/IberSPEECH.2024-60

RAMÍREZ SÁNCHEZ, J. M., & GARCÍA MATEO, C. (2022), *Report on the Galician language (D1.15)*, European Language Equality. https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_15__Language_Report_Galician_.pdf

VLADU, A. I., DE-DIOS-FLORES, I., MAGARIÑOS, C., ORTEGA, J. E., PICHEL, J. R., GARCÍA, M., GAMALLO, P., FERNÁNDEZ REI, E., BUGARÍN-DIZ, A., GONZÁLEZ GONZÁLEZ, M., BARRO, S., & REGUEIRA, X. L. (2022),

«Proxecto Nós: Artificial intelligence at the service of the Galician language», en *SEPLN-PD 2022: Projects and Demonstrations*, pp. 26–30. https://ceur-ws.org/Vol-3224/paper07.pdf

VLADU, A. I., VÁZQUEZ ABUÍN, M., FERNÁNDEZ REI, E., GARCÍA DÍAZ, N., VIDAL MIGÜÉNS, A., & MAGARIÑOS, C. (2023), *Nos_TranscriSpeech-GL: Galician ASR corpus* (v1.0.0) [Conjunto de datos], Zenodo. https://doi.org/10.5281/zenodo.7717140

VLADU, A. I., VÁZQUEZ ABUÍN, M., GONZÁLEZ CORBELLE, J., FERNÁNDEZ REI, E., GARCÍA DÍAZ, N., VIDAL MIGÜÉNS, A., & MAGARIÑOS, C. (2024), *Nos_Transcrispeech-GL* [Conjunto de datos], Hugging Face. https://huggingface.co/datasets/proxectonos/Nos_Transcrispeech-GL

WILKINSON, M. D., DUMONTIER, M., AALBERSBERG, I. J., APPLETON, G., AXTON, M., BAAK, A., BLOMBERG, N., BOITEN, J.-W., DA SILVA SANTOS, L. B., BOURNE, P. E., *et al.* (2016), «The FAIR guiding principles for scientific data management and stewardship», *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Lingüística de corpus e inteligencia artificial: desambiguación semántica de *pobo* (‘pueblo’) en gallego

Vítor Míguez Rego
Instituto da Lingua Galega –
Universidade de Santiago de Compostela

Palabras clave: Grandes modelos de lenguaje; acuerdo entre anotadores; corpus
Keywords: Large language models; inter-annotator agreement; corpus

En las últimas décadas, la lingüística ha vivido una serie de transformaciones metodológicas que se han entendido como un «giro cuantitativo» (Kortmann, 2021). Esto ha supuesto no solo que cada vez más estudios usen herramientas estadísticas para explorar datos lingüísticos y probar hipótesis, sino también una creciente sofisticación de los métodos estadísticos empleados, que ha llevado a que se generalice el uso de técnicas predictivas multifactoriales, como los modelos de regresión y los basados en árboles de decisión. Este hecho es fácilmente perceptible en una disciplina como la lingüística de corpus, en la cual los aspectos cuantitativos han ocupado un lugar central desde su nacimiento (Stefanowitsch, 2020).

De manera paralela, la anotación automática de los textos que integran los corpus ha ido incrementando su precisión, pero este avance se ha dado principalmente en la anotación morfosintáctica de las clases de palabras, de manera que, tras la extracción de los datos del corpus, sigue siendo necesario anotar manualmente muchos otros aspectos que interesan a los lingüistas. Sin embargo, los avances recientes en inteligencia artificial y, particularmente, el desarrollo de grandes modelos de lenguaje (LLM, por sus siglas en inglés) ofrecen la oportunidad de agilizar las labores de anotación de datos de corpus. El éxito de este tipo de aproximaciones no solo reduciría el tiempo requerido para llevar a cabo un estudio lingüístico basado en corpus, sino que aumentaría la fiabilidad de los resultados, al reducirse el papel humano en la anotación y, en consecuencia, la propensión a errores e inconsistencias.

Los LLM son una herramienta de desarrollo muy reciente, pero han atraído mucha atención en muy variados ámbitos de conocimiento por su aplicabilidad a todo tipo de escenarios. En el caso de la lingüística de corpus, se ha puesto el foco en cómo emplear los LLM como «copilotos del lingüista» (Torrent et al., 2024). Así, por ejemplo, en el ámbito de la gramática, Morin y Larsson (2025) proponen un método para identificar las diferentes variantes de la construcción evaluativa con el verbo *consider* en inglés. Su objetivo es que el investigador no

tenga que dedicar tiempo a seleccionar los casos que efectivamente corresponden a la construcción objeto de estudio, pues, en una búsqueda que devuelve 18 millones de *tokens*, como es el caso, esta perspectiva es poco realista. Según sus resultados, los LLM superan el 90% de precisión en la identificación de la construcción, lo cual valida su método de automatización de anotación gramatical.

En esta contribución, nos proponemos explorar el uso de LLM con el fin de filtrar los usos de la palabra gallega *pobo* con el valor de entidad de población en ejemplos de corpus. Mariño Paz (2024) analiza información correspondiente a este lema nominal en las principales fases de la lengua gallega, con especial atención a la aparición y extensión de la innovación semántica consistente en emplear *pobo* para designar una entidad de población. Los datos analizados por el autor muestran que este uso penetró en el gallego en la época medieval a través de traducciones de textos en castellano y se generalizó a partir del siglo XIX. En la actualidad, el uso de *pobo* con el sentido de entidad de población es considerado un castellanismo. Esto ha llevado a que la innovación semántica sea más común en textos que no suelen estar sujetos a procesos de revisión lingüística, como los periodísticos o procedentes de fuentes orales.

Esta investigación se divide en dos fases. La primera se centra en la obtención y anotación de los datos y en las posibilidades de automatizar este proceso mediante el uso de LLM, mientras que la segunda tiene por objetivo llevar a cabo el análisis de los datos anotados y comparar el uso de *pobo* con el valor anteriormente citado con sus formas estándar correspondientes, *aldea* y *vila*. En esta propuesta, presentamos el método y los resultados de la primera fase.

Los datos fueron extraídos del *Corpus de Referencia do Galego Actual* (CORGA) en su versión más reciente (4.1) mediante la búsqueda de la cadena ortográfica <pobo|pobos>, que devuelve todos los casos de *pobo*, *pobos*, *Pobo* y *Pobos* presentes en el corpus. Se seleccionaron aleatoriamente 400 ejemplos del conjunto de datos: 300 para servir como conjunto de prueba tanto para anotadores humanos como para LLM, y 100 adicionales para el conjunto de preentrenamiento utilizado en una condición experimental.

Estos datos fueron anotados manualmente por tres lingüistas nativos de gallego y automáticamente con dos versiones del LLM Claude 4 (Anthropic, 2025): Sonnet y Opus. El primero es considerado el modelo más eficiente y de uso diario, mientras que el segundo es más potente y adecuado para tareas complejas. Además, para la anotación automática se emplearon dos estrategias: una condición básica sin ejemplos previos y otra con preentrenamiento, donde el modelo examina 100 ejemplos anotados antes de proceder con la tarea principal.

La tarea consiste en asignar a cada ejemplo de la muestra de 300 casos una de dos clasificaciones: «mantener» (positiva) o «descartar» (negativa). Ya que el objetivo es usar la anotación con LLM como un primer filtro, de manera que luego los casos positivos sean revisados manualmente, se otorga prioridad

en la descripción de la tarea a la cobertura (*recall*) sobre la precisión, con el objetivo de evitar los falsos negativos, que acarrearían la pérdida irremediable de datos.

Un aspecto metodológico crucial de este estudio es la selección de métricas de evaluación. Mientras que trabajos previos en lingüística de corpus han empleado principalmente F1 y el coeficiente de correlación de Matthews (MCC) para validar el rendimiento de los LLM, estas métricas otorgan igual peso a la precisión y a la cobertura. Sin embargo, en tareas de filtrado de corpus, los falsos negativos son mucho más costosos que los falsos positivos, ya que los casos descartados erróneamente se pierden permanentemente del análisis, comprometiendo la representatividad de la muestra. Por ello, este estudio propone el uso de F2, una métrica que da más peso a la cobertura, como medida principal de evaluación en trabajos basados en corpus que emplean LLM como filtros de primera pasada.

Para medir el acuerdo entre anotadores, se empleó la κ de Fleiss (Fleiss, 1971). El acuerdo entre los tres anotadores humanos de nuestra muestra se sitúa en 0,656, lo que indica un acuerdo sustancial y sugiere una interpretación consistente de la tarea entre todos los anotadores. Para medir el rendimiento de los LLM, se estableció el consenso humano mediante voto por mayoría en los casos de desacuerdo, que sirvió como estándar de referencia. El acuerdo unánime entre los tres anotadores se produjo en el 87,3 % de los casos (262/300), mientras que en los restantes solo hubo acuerdo entre dos de ellos.

A partir del consenso humano, se calcularon, para cada LLM, las medidas de precisión, cobertura, F1 y F2. Los resultados, ordenados por el valor de F2, son los siguientes:

Modelo	**Precisión**	**Cobertura**	**F1**	**F2**
Opus (preentrenamiento)	0,581	0,973	0,727	0,857
Opus (básico)	0,646	0,838	0,729	0,791
Sonnet (básico)	1,000	0,622	0,767	0,673
Sonnet (preentrenamiento)	1,000	0,595	0,746	0,647

Según F2, el modelo Opus con preentrenamiento sería el más adecuado para esta tarea. De acuerdo con las métricas, este modelo identifica el 97,3% de los verdaderos positivos (cobertura), es decir, los falsos negativos se reducen a menos del 3% de la muestra. Por otra parte, solo el 58,1% de los casos identificados como positivos lo son realmente (precisión), es decir, los falsos positivos superan el 40%. Aunque esta cifra es elevada, resulta asumible en el contexto de una tarea donde es preferible maximizar la cobertura y reducir al mínimo la omisión de casos relevantes.

La comparación entre F1/MCC y F2 revela diferencias cruciales en la selección del modelo. Basándose en F1 o MCC, se elegiría Sonnet sobre Opus, a pesar de que el primero tiene un rendimiento deficiente en términos de

cobertura. F2, por el contrario, identifica exitosamente el modelo que mejor maneja las prioridades de la tarea, demostrando la importancia de seleccionar métricas de evaluación que se alineen con los objetivos específicos del trabajo de investigación.

En cuanto al impacto del preentrenamiento, Opus mostró mejor rendimiento después de examinar el conjunto de ejemplos de preentrenamiento, mientras que esta estrategia tuvo un efecto negativo en Sonnet, sugiriendo que los modelos más computacionalmente costosos pueden comprender mejor las prioridades sensibles de la tarea.

El análisis de eficiencia revela implicaciones prácticas significativas. Con el uso de la estrategia de filtrado basada en el mejor modelo (Opus con preentrenamiento), se pasaría de revisar manualmente 300 ejemplos para obtener 37 positivos a revisar solo 62 para obtener 36, lo que representa una reducción del 79,3% en la carga de trabajo con una pérdida mínima de información (menos del 3% de los casos relevantes).

Estos resultados demuestran la viabilidad de emplear LLM como filtros de primera pasada en tareas de desambiguación semántica en lingüística de corpus, particularmente para lenguas de recursos limitados como el gallego. El estudio contribuye al campo emergente de la lingüística de corpus asistida por LLM al establecer un marco metodológico riguroso que prioriza la cobertura sobre la precisión mediante el uso de F2 como métrica de evaluación principal, y al demostrar que se pueden obtener ganancias significativas de eficiencia sin sacrificar el rigor científico cuando se mantiene la supervisión humana en el flujo de trabajo.

Bibliografía

ANTHROPIC (2025), *Claude* (ver. 4), recurso en línea. https://claude.ai/

CENTRO RAMÓN PIÑEIRO PARA A INVESTIGACIÓN EN HUMANIDADES (2024), *Corpus de Referencia do Galego Actual (CORGA)* (ver. 4.1), recurso en línea. https://corpus.cirp.gal/corga/

FLEISS, J. L. (1971), «Measuring nominal scale agreement among many raters», *Psychological Bulletin*, 76(5), pp. 378–382. https://doi.org/10.1037/h0031619

KORTMANN, B. (2021), «Reflecting on the quantitative turn in linguistics», *Linguistics*, 59(5), pp. 1207–1226. https://doi.org/10.1515/ling-2019-0046

LANDIS, J. R., & KOCH, G. G. (1977), «The measurement of observer agreement for categorical data», *Biometrics*, 33(1), pp. 159–174. https://www.jstor.org/stable/2529310

MARIÑO PAZ, R. (2024), «O uso de pobo coa acepción de “entidade de poboación” na historia da lingua galega», *Revista de Filología Románica*, 41, pp. 59–74. https://revistas.ucm.es/index.php/RFRM/article/view/91445

MORIN, C., & LARSSON, M. M. (2025), «Large corpora and large language models: a replicable method for automating grammatical annotation», *Linguistics Vanguard*. https://doi.org/10.1515/lingvan-2024-0228

STEFANOWITSCH, A. (2020), *Corpus Linguistics: A Guide to the Methodology*, Language Science Press. https://doi.org/10.5281/zenodo.3735822

TORRENT, T. T., HOFFMANN, T., ALMEIDA, A. L., & TURNER, M. (2024), *Copilots for Linguists: AI, Constructions, and Frames*, Cambridge University Press. https://doi.org/10.1017/9781009439190

Las redes sociales como herramientas para la enseñanza y difusión de la lengua gallega entre la juventud: iniciativas protagonizadas por docentes de Secundaria y Primaria

Carme Silva Domínguez
Universidade de Santiago de Compostela – Instituto da Lingua Galega

Palabras clave: Gallego en la juventud; redes sociales; docencia en las redes
Keywords: Galician in youth; social media; teaching on social media

A partir de la publicación de los resultados de la última encuesta a la población gallega acerca de su conocimiento y uso de la lengua propia (IGE 2024), se ha producido una llamada de atención acerca del notable retroceso del gallego como lengua habitual entre la juventud. Paralelamente, durante los últimos años se ha notado un aumento de la presencia de creadores y creadoras de contenido digital que utilizan las redes sociales. La creciente y diversa colectividad que utiliza la lengua gallega en redes como Instagram, Tiktok y Youtube ha contado con el respaldo político proporcionado por la creación del concurso *Youtubeiras* (https://youtubeiras.gal/), que acumula ya ocho ediciones desde su primera convocatoria en el año 2017. Muchas de las personas que en la actualidad crean de manera estable contenido en lengua gallega han comenzado en este certamen.

Por otro lado, si bien la orientación de la política lingüística institucional (esto es, "desde arriba", en el sentido laboviano del término), todavía no está clara a este respecto, sí parece haber una cierta voluntad por parte de los poderes públicos orientada a transitar hacia el ámbito de las redes sociales como espacio privilegiado de impulso del gallego en la juventud, asumiendo la influencia que las y los creadores de contenido alcanzan en todos los ámbitos de la vida de este sector de la población. Como muestra de esta orientación aducimos la reciente afirmación del Secretario Xeral de Lingua de la Xunta de Galicia, Valentín García (https://www.xunta.gal/es/notas-de-prensa/-/nova/003744/valentin-garcia-destaca-relevancia-los-proyectos-que-adaptan-uso-del-gallego), quien confiesa creer en "la importancia de proyectos que trabajen en la normalización lingüística y adapten el uso de la lengua a las nuevas formas de comunicación". Además, la actual ley marco que regula el proceso de enseñanza y aprendizaje, la Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de

Educación, conocida como LOMLOE, ofrece un contexto adecuado para el impulso al aprendizaje a través de las redes sociales, al insistir en la contextualización de todo tipo de saberes en las experiencias vitales del alumnado, así como en la aproximación a sus hábitos e intereses.

Asumiendo este marco general y también el carácter transicional del momento social que se vive en la Comunidad Autónoma de Galicia, la comunicación que proponemos afronta el estudio de varias iniciativas ya en marcha, que pretenden asentar la lengua gallega en las redes sociales. Aunque tanto en Instagram como en Tiktok o Youtube son importantes cuantitativamente los perfiles que tienen como tema central el aprendizaje de la lengua gallega y que se orientan a colectivos diversos, hemos decidido trabajar únicamente con aquellos que son responsabilidad de docentes (o grupos de docentes) en ejercicio en la Comunidad Autónoma de Galicia. Nos interesa conocer diversos aspectos relativos a su incursión en las redes sociales: su perspectiva acerca de la utilidad de estas como herramienta de difusión de contenido lingüístico, la posible complementariedad de su trabajo como creadores y creadoras de contenido y su principal ocupación docente, los aspectos básicos de su ideología lingüística, la estrategia que emplean para la toma de decisiones acerca de la variedad de lengua presente en sus vídeos, entre otras cuestiones.

Se ha diseñado una aproximación cualitativa, que parte, por un lado, del análisis de los materiales disponibles en los perfiles de estas personas creadoras y, por otro, de una entrevista a los y las responsables de las cuentas que se estudian, centrada tanto en el proceso de creación en sí mismo (guionizado, subtitulado, selección de variantes de la lengua, etc.), como en los objetivos e ideología lingüística desde los que este proceso tiene lugar. En el estudio de los materiales creados para las diferentes redes resulta también valioso y necesario enfrentarse a las interacciones que estos provocan, tanto en términos cualitativos (número de *likes*, perfil predominante de las personas que interactúan con este perfil de creadores y creadoras), como cuantitativos (tipología de los discursos expuestos por los y las usuarias).

El trabajo ha permitido aproximarse a la ideología lingüística de docentes en ejercicio que utilizan las redes como herramientas para la ampliación del conocimiento y uso de la lengua gallega. Como cabía esperar, las cuentas analizadas tienen una orientación fundamentalmente normativa en cuanto a la variedad utilizada, lo cual es perfectamente lógico, dada la complementariedad de su labor en redes con respecto a la ocupación profesional principal de estas personas. Se ha desvelado una tipología de intervenciones que va desde aquellas más amplias y ambiciosas, en las que están implicados ya equipos de docentes y no personas a título individual, a aquellas otras más reducidas, presentadas públicamente como una especie de *hobby* y orientadas en muchos casos a géneros específicos (culinaria, estilo de vida, etc.).

Mediante esta aportación analítica continuamos en la estela de trabajos anteriores sobre el universo de las redes sociales en lengua gallega (Regueira

2020, Regueira & Basanta 2017, Amarelo 2025, Regueira & Silva-Domínguez 2025) y que sin duda será complementada en breve con otros trabajos de interés.

Bibliografía

AMARELO, D. (2025), «Memesphere as a Counterpublic. Fake Instagram Accounts of Galician City Councils and the Contestation of Institutional Understandings of Community», en D. AMARELO & L. LESTA (eds.), *Beyond sentidiño. New Diasporic Reflections on Galician Culture*, London, Routledge, pp. 213–227.

BASANTA LLANES, N., & REGUEIRA, X. L. (2017), «"Este mapa tal vez ajudava a romper algum preconceito": o valor múltiple da autenticidade no seo do capitalismo global», *III Simposi Internacional EDiSo: Desigualtats i nous discursos socials*, Barcelona, 28–30 de junio de 2017.

INSTITUTO GALEGO DE ESTATÍSTICA (IGE) (2024), *Enquisa estrutural a fogares. Coñecemento e uso do galego. Resumo de resultados* (11/10/2024), recurso en línea. https://www.ige.gal/web/mostrar_actividade_estatistica.jsp?codigo=0206004 [Acceso: 13/10/2025].

MINISTERIO DE EDUCACIÓN (ESPAÑA) (2020), *Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación (LOMLOE)*, *Boletín Oficial del Estado*, recurso en línea. https://www.boe.es/buscar/doc.php?id=BOE-A-2020-17264

REGUEIRA, X. L. (2020), «Contacto, estándar e ideoloxías: a lingua galega na esfera pública», en L. de C. MOUTINHO, R. L. COIMBRA, & A. GÓMEZ BAUTISTA (eds.), *Línguas minoritarias e variação linguística*, Aveiro, Universidade de Aveiro, pp. 27–51.

REGUEIRA, X. L., & SILVA-DOMÍNGUEZ, C. (2025), «As redes sociais nunha lingua oficial minorizada. Ideoloxías e impacto social das redes orais en galego», comunicación presentada en el *I Congreso Internacional Vulnerom. Vulnerabilidades sociolingüísticas en el espacio romance*, Universidad de Oviedo, 19–20 de junio de 2025.

ESPRINT: un corpus digital para el análisis interaccional del conflicto verbal en contextos íntimos

Marta Albelda Marco
Maria Estellés Arguedas
Universitat de València

Palabras clave: Corpus; conflicto comunicativo; conversación espontánea; metodología lingüística
Keywords: Corpus; communicative conflict; spontaneous conversation; linguistic methodology

La escasez de datos accesibles para el estudio lingüístico del conflicto interpersonal representa un importante desafío para la investigación en lingüística y para el análisis del discurso. En esta comunicación se presenta el corpus ESPRINT, un corpus digital concebida para abordar esta carencia mediante la recopilación sistemática de conversaciones verbales espontáneas en contextos de conflicto familiar y de pareja. Este corpus pretende cubrir un hueco metodológico y bibliográfico en el registro y género de la conversación informal, que complementa a la ya hoy en día extendida tradición de corpus de conversación coloquial “armónica”.

Desde una perspectiva metodológica y ética, se expondrán los principales retos afrontados en la construcción del corpus: desde la gestión legal de datos sensibles (en cumplimiento con la LOPD 3/2018), hasta cuestiones técnicas de grabación, anonimización, y transcripción parcial de los materiales. Los desafíos que implican el diseño de un corpus de estas características afectan a sus distintas fases de creación (grabación, transcripción, almacenamiento, accesibilidad).

Junto con las dimensiones legales y técnicas, se han afrontado cuestiones metodológicas referentes a la coherencia y representatividad de las muestras: asegurar un número suficiente de hablantes para no sesgar los resultados a nivel idiolectal y conceptual y obtener muestras extensas de conversación con el fin de disponer de visiones holísticas de la génesis de las confrontaciones, así como contar con un número suficiente de datos para poder recuperar diversos patrones frecuentes de comunicación y contar con un corpus de control de estos hablantes más allá del habla conflictiva. También, metodológicamente se ha desarrollado una definición operativa del conflicto comunicativo para poder identificar muestras del objeto de estudio.

El corpus, que actualmente contiene unas 750 horas de grabaciones procedentes de 8 parejas españolas —con aproximadamente 50 horas

transcritas—, se ha desarrollado en colaboración con profesionales del ámbito de la psicología, terapia familiar.

Asimismo, se presentarán algunos resultados preliminares que ilustran las posibilidades del corpus para el análisis computacional e interaccional del conflicto, entre las que se incluye el estudio de patrones lingüísticos, dinámicas de turno y recursos retóricos en situaciones de fuerte carga emocional. La presentación de este corpus pretende servir como reflexión sobre el valor de las infraestructuras digitales especializadas para el avance de las Humanidades Digitales aplicadas al discurso interpersonal y la comunicación en contextos sensibles.

Bibliografía

ALBELDA, M. (2024), «Factores comunicativos agravadores o/y moderadores del conflicto de pareja», *Cultura, Lenguaje y Representación*, 35, pp. 159–180.

ALLWOOD, J. (2007), «Cooperation, competition, conflict and communication», *Gothenburg Papers in Theoretical Linguistics*, 94, pp. 1–13.

BAÑÓN HERNÁNDEZ, A. (2010), «Comunicación destructiva y agresividad verbal en la Red. Notas en torno al discurso sobre las enfermedades raras», *Discurso & Sociedad*, 4(4), pp. 649–673.

BLUM-KULKA, S. (1997), *Dinner-Talk: Cultural Patterns of Sociability and Socialization in Family Discourse*, Mahwah (NJ), Lawrence Erlbaum.

BOCHNER, A. (1984), «The function of human communication in interpersonal bonding», en C. ARNOLD & J. W. BOWERS (eds.), *Handbook of Rhetorical and Communication Theory*, Boston, Allyn and Bacon, pp. 544–621.

BOU-FRANCH, P. (2021), «Evaluation, conflict and prescriptive metapragmatic comments: (Re)constructing transmedia stories», en M. JOHANSSON, S.-K. TANSKANEN, & J. CHOVANEC (eds.), *Analyzing Digital Discourses*, Cham, Palgrave Macmillan, pp. 187–210. https://doi.org/10.1007/978-3-030-84602-2_8

BOXER, D., & PLACENCIA, M. E. (eds.) (2018), *Closeness and Conflict: The Discourse of Domestic Discord across English and Spanish-Speaking Communities* (monográfico), *Journal of Language Aggression and Conflict*, 7(1).

BOXER, D., & RADICE, J. (2018), «Bickering: A conflict speech behavior of close social distance», *Journal of Language Aggression and Conflict*, 6(2), pp. 177–202.

BRENES PEÑA, E. (2009), *La agresividad verbal y sus mecanismos de expresión en el español actual* [Tesis doctoral], Universidad de Sevilla.

BRIZ, A. (2023), «Para el análisis de conversaciones en el conflicto de pareja», en C. FUENTES & E. BRENES (eds.), *Macrosintaxis y pragmática: discurso persuasivo*, Pamplona, EUNSA, pp. 259–284.

BRIZ, A. (2024), «Los conflictos en la conversación coloquial entre familiares, amigos o conocidos», *Cultura, Lenguaje y Representación*, 35, pp. 13–35.

CABEDO NEBOT, A. (2024), «Análisis multicapa del habla conversacional conflictiva a partir de la herramienta computacional Oralstats Furious», *Cultura, Lenguaje y Representación*, 35, pp. 75–94.

CARCELÉN GUERRERO, A., & UCLÉS RAMADA, G. (2019), «Diseño y construcción de un corpus oral multidialectal», *Normas: Revista de Estudios Lingüísticos Hispánicos*, 9(1), pp. 17–36.

CARCELÉN, A. (2023), *Bases teórico-metodológicas para la construcción de un corpus multidialectal de conversación coloquial: el corpus Ameresco* [Tesis doctoral], Universitat de València.

CARCELÉN, A. (2024), «Aspectos éticos y legales en la construcción de un corpus oral de conversación conflictiva», ms. inéd.

COLEMAN, P. T. (2006), «Intractable conflict», en M. DEUTSCH, P. T. COLEMAN, & E. C. MARCUS (eds.), *The Handbook of Conflict Resolution: Theory and Practice*, San Francisco, Jossey-Bass/Wiley, pp. 533–559.

CLANCY, B. (2018), «Conflict in corpora: Investigating family conflict sequences using a corpus pragmatic approach», *Journal of Language Aggression and Conflict*, 6(2), pp. 228–247. https://doi.org/10.1075/jlac.6.2.04cla

ESTELLÉS, M. (2023), «Visualizando el conflicto discursivo a través de la expresión fónica: un estudio a partir de dos conversaciones», *Normas*, 13(1), pp. 224–247.

ESTELLÉS, M. (2024), «Dos visiones del conflicto en conversación espontánea: participantes vs. analistas», *Cultura, Lenguaje y Representación*, 35, pp. 53–73.

FUENTES RODRÍGUEZ, C., & ALCAIDE LARA, E. (2008), *(Des)cortesía, agresividad y violencia verbal en la sociedad actual*, La Rábida, Universidad Internacional de Andalucía.

GARCÉS-CONEJOS BLITVICH, P. (2024), *Pragmatics, (Im)politeness, and Intergroup Communication: A Multilayered, Discursive Analysis of Cancel Culture*, Cambridge, Cambridge University Press.

GRIMSHAW, A. D. (ed.) (1990), *Conflict Talk*, Cambridge, Cambridge University Press.

GUTHRIE, D., & NOLLER, P. (1988), «Married couples' perceptions of one another in emotional situations», en P. NOLLER & M. FITZPATRICK (eds.), *Perspectives on Marital Interaction*, Clevedon, Multilingual Matters, pp. 153–181.

KAISER, H. (2018), «Intimacy matters: Uruguayan women's refusal behavior in couples talk», *Journal of Language Aggression and Conflict*, 6(2), pp. 269–296.

LAFOREST, M. (2002), «Scenes of family life: Complaining in everyday conversation», *Journal of Pragmatics*, 34(12), pp. 1595–1620.

LEVINSON, S. C., & ENFIELD, N. J. (eds.) (2006), *Roots of Human Sociality: Culture, Cognition and Interaction*, Oxford, Berg.

OETZEL, J. G., & TING-TOOMEY, S. (eds.) (2006), *The SAGE Handbook of Conflict Communication: Integrating Theory, Research, and Practice*, Thousand Oaks, Sage.

SIFIANOU, M. (2019), «Conflict, disagreement and (im)politeness», en M. EVANS *et al.* (eds.), *The Routledge Handbook of Language in Conflict*, London/New York, Routledge, pp. 176–195.

Hacia un corpus de películas y series de televisión españolas del español coloquial: criterios de selección y metadatos

Ana Llopis Cardona
Universitat de València

Palabras clave: Corpus; películas; series de televisión; español coloquial
Keywords: Corpus; films; television series; colloquial Spanish

En estudios recientes se ha reconocido que la representación de la conversación coloquial es bastante fiable en películas y series de televisión (Jucker 2021; Jucker & Landert 2023; Tonetti & Landert 2023). No obstante, en la actualidad solo existe un corpus de subtítulos de películas y series de televisión en lengua inglesa (*TV Corpus/Movies Corpus*, Davies 2021), y un corpus de transcripciones de series de televisión (*Sydney Corpus of Television Dialogue*, Bednarek 2018). El español no cuenta con un corpus similar ni los corpus generales han incluido este tipo de material (CORDE; CREA; CdH; CdE) o han incorporado un escaso número de archivos (CORPES XXI). Ante este vacío, surge el *Corpus oral de coloquialidad ficticia* (COLFI), compuesto por transcripciones de películas y series de televisión españolas desde los años cincuenta del siglo pasado hasta la actualidad. Este corpus se inició en 2020 con el proyecto “Difusión del cambio lingüístico en el español coloquial durante los últimos cincuenta años” (GV/2020/157) y se ha desarrollado en una primera fase gracias a la Beca Leonardo a Investigadores y Creadores Culturales 2023 de la Fundación BBVA (LEO23-2-10560).

En esta comunicación se exponen aspectos cruciales del proceso de elaboración de este corpus, a saber: los criterios de selección de las obras audiovisuales, las fichas técnicas de metadatos de la obra y de los personajes, así como algunas cuestiones del sistema de transcripción. La selección de las obras fílmicas y televisivas se basó principalmente en cinco criterios: historias coetáneas, con diálogos propios de la inmediatez comunicativa (Oesterreicher 1996) o de la conversación coloquial (Briz 1998), con marcas de variación diafásica y diastrática –al menos en cierta medida–, escritas en su versión original en español europeo y por guionistas españoles (o en colaboración de guionistas españoles).

La aplicación de estos parámetros excluye el cine fantástico y el wéstern, y afina la elección de las comedias y de los dramas. Las historias narradas no solo deben ser coetáneas, sino que sus personajes protagonistas deben mantener una relación de cercanía, sus interacciones deben tener lugar preferentemente

en entornos cotidianos o familiares (sin predominio de un espacio profesional) y deben presentar un tono informal. Un ejemplo representativo son las películas neorrealistas del cine español (*Surcos, Felices Pascuas*), entre las que se encuentran varias películas de Luis García Berlanga coguionizadas con Rafael Azcona (*Se vende un tranvía, Plácido*). En cualquier caso, no cabe duda de que la oralidad fingida (Brumme 2008) o la mímesis conversacional (Del Rey Quesada 2019) difiere de las conversaciones espontáneas por la menor presencia de estrategias de dinámica conversacional (solapamientos, interrupciones, etc.) y estrategias sintácticas relacionadas con la planificación del discurso (López Serena 2022).

Los metadatos de la obra se componen de los datos básicos de las fichas técnicas de cine (año, director/a/es, guionista/s, género, duración), a los que se ha añadido la información de temas tratados en la obra y el modo de acceso. Para la elaboración de la ficha técnica de los personajes, se tomaron como punto de partida los rasgos sociales básicos procedentes de la sociolingüística (sexo, grupo de edad, nivel sociocultural, zona geográfica). Se decidió añadir otros datos, como el tipo de ocupación, la profesión, la relación entre personajes y el/la actor/actriz que representa el personaje. Ciertamente, los guionistas caracterizan los personajes mediante el uso de marcas lingüísticas que son representativas de un grupo etario o social; de esta forma simulan el habla real, a la vez que transmiten estereotipos y los consolidan. Con respecto a los dialectos, se debe considerar la estandarización habitual en el cine y el predominio del lenguaje urbano (cf. Moreno 2022). Las marcas dialectales pueden ser simbólicas y servir a fines estético-humorísticos; así se observa en la caricaturización del pueblerino aragonés representado en personajes interpretados por Paco Martínez Soria.

En cuanto al origen y procesamiento de los archivos, se puede recurrir a distintos materiales en bruto y procedimientos para obtener transcripciones fiables de películas y series de televisión, tales como guiones, subtítulos, subtítulos para sordos, transcripción manual y, recientemente, transcripción automática. Sea cual sea el material de base (guiones o subtítulos), se precisa una revisión para comprobar la correspondencia del texto escrito con respecto al audio. En este sentido, estudios recientes sobre corpus en inglés (Jucker 2021; Bednarek 2018, 2023) insisten en la dudosa fiabilidad de utilizar subtítulos sin revisión. Así, después de explorar las distintas opciones (dificultades de acceso, inconvenientes para su procesamiento y/o coste), se tomó la decisión de transcribir los archivos de forma manual; esta modalidad se combina en la actualidad con la revisión de las transcripciones automáticas obtenidas con el programa Whisper.

Para la transcripción manual, se han empleado todos los signos de transcripción del sistema Val.Es.Co., salvo el signo de sucesión inmediata y el de la aspiración, que no se han marcado. Para el etiquetado, se han utilizado las etiquetas señaladas en los documentos de trabajo del corpus Ameresco, que tienen un formato compatible con XML que garantiza su recuperación posterior por medio de motor de búsqueda. Se incorporaron, además, dos etiquetas

necesarias para marcar información multimodal o contextual (diegética o extradiegética), atributos que aparecen entrecomillados.

En definitiva, COLFI surge como una herramienta de Humanidades digitales, cuyo objeto de estudio corresponde principalmente a la filología (el español coloquial contemporáneo) y secundariamente a las artes visuales (ficción fílmica y televisiva); por tanto, se trata de un recurso vinculado a distintas parcelas del conocimiento –diacronía y sincronía del español coloquial contemporáneo, ficción audiovisual en cine y series televisivas– o a su intersección.

Bibliografía

BEDNAREK, M. (2018), *Guide to the Sydney Corpus of Television Dialogue (SydTV)*, corpus manual, recurso en línea: www.syd-tv.com/publications/.

BRIZ GÓMEZ, A. (1998), *El español coloquial en la conversación. Esbozo de pragmagramática*, Barcelona, Ariel.

BRUMME, J. (2008), *La oralidad fingida: descripción y traducción. Teatro, cómic y medios audiovisuales*, Madrid/Francfort del Meno, Iberoamericana/Vervuert.

DAVIES, M. (2002–2021), *Corpus del español*, recurso en línea: https://www.corpusdelespanol.org/

DAVIES, M. (2019), *TV and Movie Corpora*, recurso en línea: https://www.english-corpora.org/iweb/

DEL REY QUESADA, S. (2019), «Variantes de la oralidad elaborada en la segunda mitad del siglo XIX: dos traducciones coetáneas de *Los cautivos* de Plauto», *Oralia*, 22, pp. 283–326.

JUCKER, A. H., & LANDERT, D. (2023), «The diachrony of im/politeness in American and British movies (1930–2019)», *Journal of Pragmatics*, 209, pp. 123–141.

JUCKER, A. H. (2021), «Features of orality in the language of fiction: A corpus-based investigation», *Language and Literature*, 0(0), pp. 1–20.

LLOPIS CARDONA, A. (ed.) (2024), *Culturas juveniles, contracultura y procesos de difusión en el español coloquial* (monográfico), *Spanish in Context*, 21(1).

LÓPEZ SERENA, A. (2022), «Discursividad y mímesis de la oralidad en *Paquita Salas*: tras las huellas de la improvisación dialógica en la ficción televisiva contemporánea», en M. MÉNDEZ ORENSE & V. PÉREZ BÉJAR (eds.), *Perspectivas integradas para el análisis de la oralidad*, Sevilla, Universidad de Sevilla, pp. 57–84.

MORENO FERNÁNDEZ, F. (2022), «De Hollywood a Netflix. Las variedades del español en el cine y la televisión», en *Instituto Cervantes (ed.), Nuevo mundo. El impacto del audiovisual en la difusión internacional de la lengua y de la cultura en español*, pp. 349–386.

OESTERREICHER, W. (1996), «Lo hablado en lo escrito. Reflexiones metodológicas y aproximación a una tipología», en T. KOTSCHI, W. OESTERREICHER, & K. ZIMMERMANN (eds.), *El español hablado y la cultura oral en España e Hispanoamérica*, Madrid/Frankfurt, Iberoamericana/Vervuert, pp. 317–340.

TONETTI TÜBBEN, I., & LANDERT, D. (2023), «*Uh* and *um* as pragmatic markers in dialogues: A contrastive perspective on the functions of planners in fiction and conversation», *Contrastive Pragmatics*, 4(2), pp. 350–381.

Análisis pragmalingüístico del discurso digital con IA: evidencialidad y vigilancia epistémica en debates sobre teorías de la conspiración

Dorota Kotwica
Universitat de València

Palabras clave: Lectura Activa; Prompting; Chain-of-Thought; Marcación de textos; Desambiguación
Keywords: Active Reading; Prompting; Chain of Thought; Text Markup; Disambiguation

Este trabajo aborda un análisis pragmalingüístico de las categorías evidenciales y epistémicas como estrategias argumentativas en contextos de interacción digital conflictiva, específicamente en debates sobre teorías de la conspiración. Se emplea un enfoque experimental e interdisciplinar, para explorar el alcance, las limitaciones y las ventajas del uso de la inteligencia artificial (IA) para este tipo de análisis.

Partimos de la hipótesis de que, en escenarios altamente polarizados, como los que enfrentan el discurso científico con narrativas conspiranoicas, la evidencialidad se erige como un recurso discursivo clave. Sirve para construir autoridad epistémica, reforzar la credibilidad de los propios argumentos y deslegitimar la postura del oponente (Miecznikowski y Jacquin 2023, Kotwica y Albelda 2024). Además, en estos intercambios polémicos, se identifican indicios lingüísticos de la vigilancia epistémica (Piatti 2024, Sperber et al. 2010). Esta revela una actitud cautelosa de los hablantes ante discursos alternativos, percibidos como potencialmente falaces o manipuladores. Así, los recursos evidenciales no solo gestionan el grado de compromiso con la información, sino que también operan como mecanismos de posicionamiento ideológico y control de la credibilidad en la esfera pública digital.

La evidencialidad, en su definición amplia (Boye, 2012), alude a las fuentes de información en las que el hablante basa sus afirmaciones y engloba sus actitudes hacia el conocimiento (Chafe, 1986). El uso estratégico de la evidencialidad implica la selección de tipos de fuentes (experiencia propia, referencias externas, inferencias, sabiduría popular) que confieren un grado de fiabilidad variable. Si bien se han propuesto jerarquías universales de fiabilidad, estudios recientes indican que esta depende del contexto. La elección de marcas evidenciales que señalen fuentes fiables en un contexto específico puede ser una estrategia para reforzar argumentos o desacreditar los del oponente o para remarcar la autoridad epistémica del hablante (Stivers et al., 2011).

Por su parte, la vigilancia epistémica se refiere a los mecanismos cognitivos que los hablantes activan para evaluar y filtrar información, para reducir el riesgo de desinformación y garantizar la credibilidad del mensaje (Sperber et al., 2010). Puede dirigirse tanto a la fuente como al contenido y manifestarse a través de marcadores discursivos, expresiones de verdad, de (no) saber, sorpresa o incredulidad (Piatti, 2024).

Las teorías de la conspiración se definen como explicaciones alternativas de hechos, eventos sociales o políticos que atribuyen su causa a complots secretos organizados por grupos poderosos que actúan intenciones ocultas (Butter y Knight 2020, Demata et al. 2022, Douglas y Sutton 2023). Se construyen como narrativas que contradicen las versiones oficiales y desconfían de autoridades epistemológicas. Las redes sociales se han convertido en uno de los principales canales para la circulación de ideas conspirativas y son espacios habituales para debatir sobre ellas. En estos entornos, la información se comparte de manera «horizontal» entre usuarios, lo que suele afectar negativamente su precisión y credibilidad. Como resultado, la distinción entre hechos y opiniones tiende a desdibujarse (Demata et al., 2022).

El presente estudio se centra en un debate de YouTube de 2024, de aproximadamente 140 minutos, entre dos figuras científicas/divulgativas y dos difusores de teorías de la conspiración. Este evento tuvo una repercusión mediática considerable, con 6,2 millones de espectadores en directo y 26,341 comentarios. Analizaremos una selección de comentarios prominentes, en especial, en aquellos que reaccionan a las afirmaciones y fuentes presentadas por ambos bandos. A través de una metodología diseñada con IA, observaremos el empleo de recursos evidenciales-epistémicos y las estrategias lingüísticas que revelan la vigilancia epistémica de los espectadores, manifiestan acuerdo o desacuerdo o juzgan la fiabilidad de afirmaciones y fuentes.

Para el análisis, partimos de un marco teórico actualizado sobre las categorías evidenciales y epistémicas, con el objetivo de operacionalizar su identificación y estudio en contextos discursivos conflictivos. De manera experimental, desarrollamos una herramienta de análisis que combina métodos lingüísticos con recursos computacionales, mediante la implementación de *scripts* en Python diseñados con IA. El entorno elegido para su desarrollo y ejecución es Jupyter Notebook, una plataforma de código abierto ampliamente utilizada en la investigación científica y técnica. Este entorno permite la creación de documentos interactivos que integran código, texto explicativo y visualizaciones, lo que facilita no solo el análisis automatizado de datos lingüísticos, sino también la documentación detallada del proceso y la presentación reproducible de resultados, especialmente en entornos como Python.

La principal novedad de este estudio radica en el intento de crear una herramienta computacional orientada al análisis pragmático-discursivo de categorías evidenciales y epistémicas, un campo que tradicionalmente ha estado restringido al plano cualitativo y ha carecido de recursos automatizados

específicos. Nuestro enfoque experimental somete esta metodología a una reflexión crítica: ¿qué resultados ofrece un análisis automatizado?, ¿cuáles son sus limitaciones?, ¿hasta dónde llega su potencial complementario al análisis lingüístico?, y ¿cómo la IA permite a personas sin conocimientos de programación generar programas deterministas para el análisis de datos lingüísticos?

El componente digital es doble en este estudio: el objeto de análisis es el discurso digital (comentarios de YouTube), y la metodología incorpora herramientas digitales para la recopilación, tratamiento y análisis de los datos.

Bibliografía

BOYE, K. (2012), *Epistemic Meaning: A Crosslinguistic and Functional-Cognitive Study*, De Gruyter Mouton.

BUTTER, M., & KNIGHT, P. (eds.) (2020), *Routledge Handbook of Conspiracy Theories*, Routledge.

CHAFE, W. (1986), «Evidentiality in English Conversation and Academic Writing», en W. CHAFE & J. NICHOLS (eds.), *Evidentiality: The Linguistic Coding of Epistemology*, Norwood (NJ), Ablex, pp. 261–273.

DEMATA, M., ZORZI, V., & ZOTTOLA, A. (eds.) (2022), *Conspiracy Theory Discourses*, John Benjamins.

DOUGLAS, K. M., & SUTTON, R. M. (2023), «What Are Conspiracy Theories? A Definitional Approach to Their Correlates, Consequences, and Communication», *Annual Review of Psychology*, 74, pp. 271–298.

KOTWICA, D., & ALBELDA, M. (2024), «Degrees of disagreement and reliability of information sources in pro- and anti-vaccination comments on Facebook», *Journal of Language Aggression and Conflict*, pp. 1–27.

MIECZNIKOWSKI, J., & JACQUIN, J. (2023), «Epistemic and evidential markers in contexts of disagreement», *Journal of Pragmatics*, 213, pp. 4–11.

PIATTI, G. (2024), «Vigilancia epistémica y marcación discursiva en la coconstrucción interaccional», *Plurentes*, 15.

SPERBER, D., CLÉMENT, F., HEINTZ, C., MASCARO, O., MERCIER, H., ORIGGI, G., & WILSON, D. (2010), «Epistemic Vigilance», *Mind & Language*, 25, pp. 359–393.

STIVERS, T., MONDADA, L., & STEENSIG, J. (2011), *The Morality of Knowledge in Conversation*, Cambridge, Cambridge University Press.

Modelo metodológico para el análisis semántico de corpus históricos: integración de OCR, ontologías y SPARQL

Rafael Prezado
Universidade de Évora

Palabras clave: Humanidades Digitales; Análises de Corpus; Ontologías; OCR; SPARQL
Keywords: Digital Humanities; Corpus Analysis; Ontologies; OCR; SPARQL

Este estudio presenta un modelo metodológico para el análisis semántico de corpus textuales históricos utilizando tecnologías de reconocimiento óptico de caracteres (OCR), ontologías y consultas SPARQL. El objetivo es demostrar cómo estos recursos pueden aplicarse de forma integrada para estructurar, analizar e interpretar grandes volúmenes de texto no estructurado, como los que se encuentran en archivos de prensa histórica. A partir de la digitalización y tratamiento computacional de documentos, la propuesta busca ofrecer un enfoque replicable, escalable y adaptable para proyectos de investigación en humanidades digitales, con énfasis en la reutilización de datos y la interoperabilidad entre disciplinas.

Uno de los principales retos en el trabajo con fuentes históricas digitalizadas es su formato desestructurado, que dificulta el análisis sistemático mediante métodos tradicionales. La clasificación manual de datos, común en la historiografía clásica, limita tanto la escalabilidad como la reproducibilidad de los resultados, y a menudo impide una lectura crítica de los discursos dominantes, tal como lo plantea Fairclough en su enfoque sobre el lenguaje como forma de práctica social. Esta propuesta aborda dicho problema mediante una metodología semiautomatizada que combina herramientas digitales y modelos semánticos para transformar texto plano en datos estructurados y reutilizables, permitiendo el desarrollo de análisis tanto cuantitativos como cualitativos.

Metodología

Para el proceso de OCR se utilizó PaddleOCR, en combinación con scripts en Python, con el objetivo de automatizar la conversión de los archivos digitalizados, en formato PDF, PNG o JPG, a texto legible por máquina. En una fase inicial, se realizaron pruebas con Tesseract OCR, una herramienta de código abierto desarrollada por los HP Labs y distribuida como software libre en 2005.

Aunque ambas herramientas cuentan con arquitecturas sólidas, se observó que PaddleOCR ofrecía mayor precisión, especialmente en páginas de prensa degradadas o con diseños irregulares. Por ello, se adoptó en la fase principal del proyecto, aprovechando su arquitectura ligera y modular (PP-OCR), que combina detección, rectificación geométrica y reconocimiento de texto optimizados para contextos exigentes.

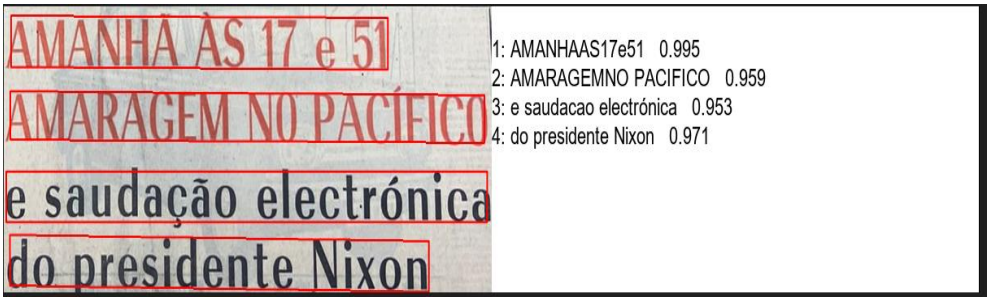


Figura 1. OCR de un titular histórico con PaddleOCR. Las cajas rojas marcan el texto detectado, con resultados y valores de confianza. Se observa alta precisión.

Una vez extraído el contenido textual, se procedió a una fase de análisis semi-automatizado orientada a la identificación de entidades clave, tales como nombres propios, lugares, eventos, instituciones y conceptos temáticos. Con este fin, se desarrolló una ontología propia, basada en el modelo CIDOC Conceptual Reference Model (CIDOC CRM), que permite estructurar lógicamente las relaciones entre entidades, en conformidad con las normas internacionales aplicadas a la gestión del patrimonio cultural y documental.

El modelo ontológico, relativo al caso de estudio presentado a continuación, fue diseñado para representar de forma explícita los vínculos entre:

- periódicos y acontecimientos históricos
- actores políticos y científicos
- elementos visuales, como caricaturas, fotografías e infografias
- localizaciones físicas y contextos ideológicos.

Estas relaciones fueron codificadas en forma de tripletas RDF, lo que permitió explorar el corpus mediante consultas semánticas en SPARQL. El entorno de análisis seleccionado fue Apache Jena Fuseki, donde se almacenaron las bases de datos RDF y se realizaron las consultas.

Cabe destacar que la ontología desarrollada no fue utilizada como sistema de anotación directa sobre los textos del corpus. En lugar de aplicar marcas semánticas a los documentos originales, se optó por una estrategia de carácter conceptual –en la que las entidades y relaciones se modelaron a partir de la información extraída–. Esta decisión metodológica permitió estructurar lógicamente los datos sin modificar los textos fuente, privilegiando la interoperabilidad y la eficiencia de las consultas SPARQL.

Aplicación: un estudio de caso como validación

Como validación del modelo, se aplicó el flujo de trabajo sobre un conjunto de periódicos portugueses del siglo XX, centrado en un evento clave de la carrera espacial: la llegada del Apolo 11 a la Luna (1969). Este corpus fue seleccionado por su relevancia simbólica dentro de los discursos sobre modernidad, progreso científico y propaganda política, en el contexto del régimen autoritario del Estado Novo en Portugal.

La información extraída permitió analizar la cobertura mediática y los discursos editoriales. Desde un enfoque cuantitativo, se generaron tablas que registran la frecuencia de palabras clave, la distribución temática por secciones y la presencia en portada. Cada periódico analizado (O Século, Diário de Lisboa, A Flama y Avante!) fue sistematizado según ubicación documental, tipo de contenido y contexto temporal.

Desde el punto de vista cualitativo, se identificaron patrones discursivos como la exaltación de la ciencia occidental, el silenciamiento de actores soviéticos, o la apropiación ideológica del evento por parte del régimen. La vinculación de estos patrones al modelo CIDOC CRM permitió no solo describir, sino estratificar las relaciones semánticas entre actores, eventos y medios.

Modelado ontológico en CIDOC CRM

En el entorno Protégé, se modelaron múltiples instancias siguiendo el esquema CIDOC CRM, y también se crearon nuevas entidades y propiedades adaptadas a las necesidades específicas del análisis mediático. Por ejemplo:

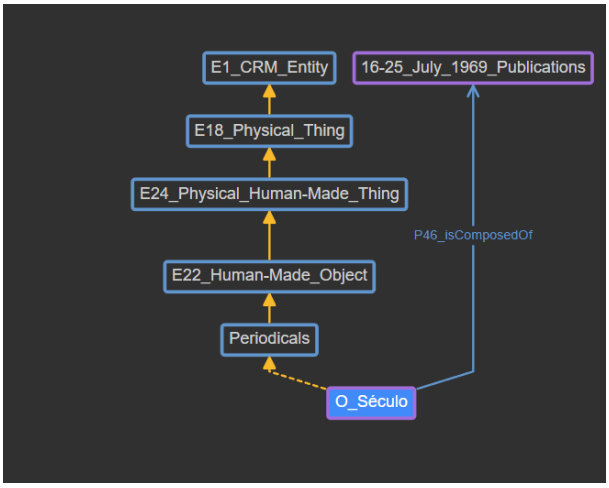


Figura 2. Representación ontológica del periódico *O Século* en Protégé

La figura 1 muestra la instancia del periódico *O Século* modelada en Protégé como parte de una subclase creada específicamente, Periodicals, derivada de E22_Human-Made_Object dentro del modelo CIDOC CRM.

Esta instancia se vincula a las publicaciones del periodo 16–25 de julio de 1969 mediante la propiedad P46_isComposedOf, permitiendo representar de forma estructurada la relación entre el objeto físico (el periódico) y sus ediciones

específicas. La extensión del modelo mediante la subclase Periodicals facilita un tratamiento más preciso de fuentes hemerográficas en contextos históricos y semánticos.

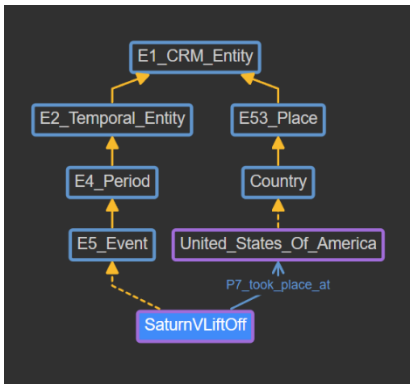


Figura 3. Instancia de la descolagem del cohete Saturn V modelada en Protégé

La figura representa el evento del despegue del Saturn V (*SaturnVLiftOff*) modelado como una instancia de E5_Event en el entorno Protégé, siguiendo el esquema del CIDOC CRM.

Este evento se clasifica dentro de la jerarquía temporal (E2_Temporal_Entity → E4_Period → E5_Event) y se vincula mediante la propiedad P7_took_place_at a la instancia United_States_Of_America, definida como un E53_Place y subclase de Country. Esta estructura permite situar el evento en un marco espacio-temporal específico y coherente con los estándares del modelado semántico.

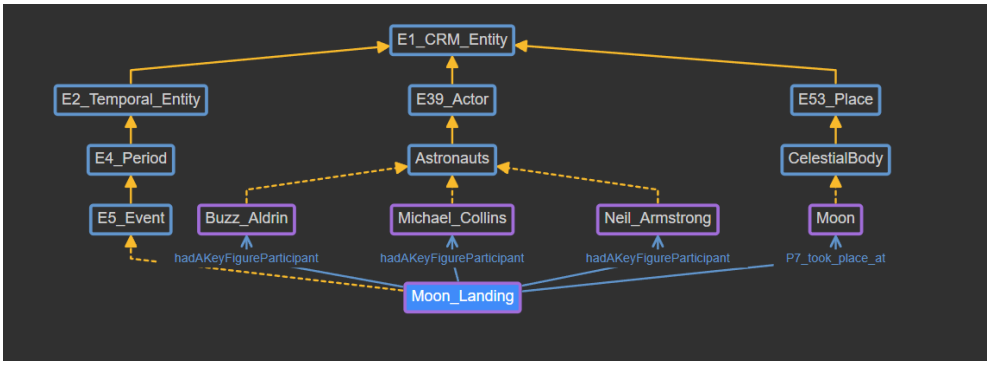


Figura 4. Instancia de la alunizaje representada en Protégé

La figura representa el evento Moon_Landing como una instancia de E5_Event en Protégé, siguiendo el modelo CIDOC CRM. El evento se vincula al lugar Moon (una instancia de E53_Place y subclase de CelestialBody) mediante la propiedad P7_took_place_at.

Los astronautas Buzz Aldrin, Michael Collins y Neil Armstrong son modelados como instancias de E39_Actor, agrupados bajo Astronauts y relacionados con el evento mediante la propiedad hadAKeyFigureParticipant, que destaca su papel central en la misión y es derivada de p11_hadParticipant.

Estos nodos semánticos permiten consultas detalladas en SPARQL. Ejemplos reales ejecutados incluyen:

- ¿Cómo se enmarcaron los discursos sobre ciencia y tecnología en un régimen autoritario?

Estas preguntas no solo demuestran la capacidad analítica del modelo, sino su potencial para generar nuevas líneas de investigación a partir de datos existentes.

Resultados y visualización

Los resultados se visualizaron en múltiples niveles. A través de Apache Jena Fuseki, se organizaron respuestas a consultas específicas, generando visualizaciones sobre:

- El volumen de referencias por periódico y fecha.

	journal	name	totalReferences
1	<http://example.org/space_race/O_Século>	O Século	"177"^^<http://www.w3.org/2001/XMLSchema#integer>
2	<http://example.org/space_race/Diário_Popular>	Diário Popular	"129"^^<http://www.w3.org/2001/XMLSchema#integer>
3	<http://example.org/space_race/Diário_de_Lisboa>	Diário de Lisboa	"108"^^<http://www.w3.org/2001/XMLSchema#integer>
4	<http://example.org/space_race/A_Capital>	A Capital	"91"^^<http://www.w3.org/2001/XMLSchema#integer>
5	<http://example.org/space_race/Diário_do_Sul>	Diário do Sul	"13"^^<http://www.w3.org/2001/XMLSchema#integer>
6	<http://example.org/space_race/Flama>	Flama	"10"^^<http://www.w3.org/2001/XMLSchema#integer>
7	<http://example.org/space_race/Avante>	Avante	"0"^^<http://www.w3.org/2001/XMLSchema#integer>

Figura 5. Organización por total de referencias – Apache Jena Fuseki

La figura muestra una visualización generada en Apache Jena Fuseki a partir de consultas SPARQL, indicando el número total de referencias encontradas en cada periódico del corpus. Esta organización permite comparar de forma cuantitativa la cobertura mediática de los eventos analizados, facilitando la identificación de patrones discursivos según el medio y el contexto temporal:

- Tipos de contenido más representados (artículos, caricaturas, entrevistas).

Por ejemplo, se generó una consulta que determina:

name	firstPageReferences
Diário de Lisboa	"17"^^<http://www.w3.org/2001/XMLSchema#integer>

Figura 6. Consulta específica sobre *Diário de Lisboa* – Apache Jena Fuseki

La figura muestra el resultado de una consulta SPARQL. Se hizo en Apache Jena Fuseki. Responde a una pregunta: "¿Cuántas referencias hay en la primera página del Diário de Lisboa entre el 16 y el 25 de julio de 1969?". Tal consulta explora el cuerpo de datos de forma directa. Obtiene números exactos sobre la visibilidad y lo importante que son los eventos analizados para la publicación.

Discusión y limitaciones

El modelo propuesto presenta ventajas significativas:

- Permite una reutilización de datos estructurados.
- Fomenta la interoperabilidad entre proyectos de HD.
- Facilita la expansión progresiva del corpus (escalabilidad).
- Establece un marco reproducible para otros estudios en español, portugués o inglés.

Sin embargo, también se detectaron limitaciones importantes:

- La calidad del OCR varía según el estado del soporte físico (papel deteriorado, fuentes tipográficas antiguas).
- La extracción semántica requiere validaciones manuales periódicas para asegurar precisión.
- La inclusión de elementos visuales (como caricaturas) requiere procesos específicos de anotación, aún en fase experimental.

Conclusiones

Este trabajo presenta una metodología sólida para el análisis semántico de corpus históricos mediante herramientas computacionales interoperables. A través de OCR (Tesseract y PaddleOCR), ontologías basadas en CIDOC CRM y consultas SPARQL en Apache Jena, se desarrolló un modelo capaz de estructurar y analizar rigurosamente fuentes de prensa histórica.

Aunque aún en fase teórica y exploratoria, ha demostrado eficacia en organizar y explorar datos heterogéneos y no estructurados. Su aplicación al caso de la cobertura de la misión Apolo 11 en periódicos portugueses durante el Estado Novo validó su potencial para identificar patrones discursivos y representar relaciones semánticas estructuradas.

Esta propuesta impulsa el avance metodológico en Humanidades Digitales, ofreciendo un flujo de trabajo replicable, escalable y alineado con estándares internacionales. Además, su carácter modular permite adaptarlo a diversos contextos y facilita su integración en infraestructuras digitales existentes, fomentando investigaciones más abiertas, reproducibles y sostenibles.

Bibliografía

BEKIARI, C., BRUSEKER, G., CANNING, E., DOERR, M., MICHON, P., ORE, C.-E., STEAD, S., & VELIOS, A. (2024), *Definition of the CIDOC Conceptual Reference Model*, recurso en línea.

FAIRCLOUGH, N. (2009), *Discourse and Social Change*, Cambridge, Polity Press.

SMITH, R. W. (2007), «An overview of the Tesseract OCR engine», en *IEEE International Conference on Document Analysis and Recognition (ICDAR 2007)*.

Transcripción de textos históricos en español con diferentes fuentes y formatos: análisis y comparación de resultados con tres enfoques basados en IA

Ana García-Serrano
Jaione Macicior
Marina Míguez
Yanco Torterolo
UNED

Palabras clave: Transcripción monomodal y multimodal; Transkribus; Tesseract; Granite; Inteligencia Artificial (IA)
Keywords: Monomodal and multimodal transcription; Transkribus; Tesseract; Granite; Artificial Intelligence (AI)

En este trabajo se describen y analizan conjuntamente los experimentos y resultados obtenidos en el marco de la tarea compartida PastReader – *Transcribig texts from the past*[1] que aportaba un corpus de miles de ficheros con textos históricos en español cedidos por la BNE, para que los participantes desarrollasen algoritmos, modelos o aplicaciones para transcripción automática de los textos.

En una tarea compartida, la colección de documentos o corpus inicial se crea y revisa por los organizadores. Una parte de este corpus, que se denomina *gold standard*, se utiliza para que los algoritmos basados en IA aprendan modelos especializados en el dominio concreto de la tarea (*training corpus* o de entrenamiento*)*. Otra parte se usa para analizar y comparar, de acuerdo con las métricas establecidas previamente, los resultados aportados por cada participante (corpus de prueba/test*)*. En ocasiones, una pequeña parte del corpus se utiliza para validación durante el desarrollo del modelo o aplicación (*development corpus*). Estos corpus también pueden denominarse *ground truth* porque contienen los resultados "verdaderos" para cada ejemplo del corpus en la tarea.

Se ha participado con dos enfoques monomodales y uno multimodal. Para el primero se ha utilizado la aplicación web TransKribus[2] que contiene modelos basados en IA, una buena opción para usuarios sin experiencia en tecnología que desean transcribir. El segundo y tercer enfoque se basan en dos modelos del

[1] https://sites.google.com/view/pastreader2025
[2] https://www.transkribus.org/

lenguaje, Tesseract[3] (Google) y Granite[4] (IBM) ambos en abierto (Macicior, 2025), (Torterolo and Miguez, 2025). La conclusión mas importante es que todas nuestras ejecuciones (run) participantes tuvieron una calidad aceptable en comparación con el *baseline* (resultados de base para la comparación) aportado por los organizadores. Destacó el buen rendimiento del modelo Granite3.2-vision:2b, principalmente en métricas semánticas y a nivel de palabra, aunque fue menor del esperado por las limitaciones de hardware y de computación que obligaron al uso de una variante pequeña del modelo.

Los resultados obtenidos son también un reflejo de la experiencia previa de los autores: organización de la tarea compartida FinCausal 2025 (Moreno et al., 2025), participación con mejor puntuación en la tarea compartida *FinancES 2023* (Porta et al., 2023) o en la tarea compartida multimodal CLEF (Martínez et al., 2005), el trabajo sobre transcripción de textos históricos (Sánchez-Salido et al., 2023) y sobre procesamiento del lenguaje natural (García-Serrano y Menta, 2022) con desarrollo de *benchmarks* (Lastra et al., 2021).

1. Descripción de las aproximaciones y resultados

El corpus proporcionado en la tarea *PastReader* 2025 (Montejo et al, 2025) contiene, casi 9000 documentos (imágenes de páginas en formato pdf) para el corpus de entrenamiento, 500 en el corpus de validación/desarrollo y 3000 en el corpus de prueba/test. Además, se dispone de las transcripciones generadas automáticamente y revisadas manualmente de los corpus de entrenamiento y de validación. Las páginas son de documentos históricos procedentes de diversos archivos y repositorios, desde revistas culturales y de noticias generales (*Juventud, El Español*), hasta revistas satíricas y humorísticas (*El Duende Satírico del día*), crónicas de espiritualismo (*La luz del porvenir*), publicaciones seriadas de narrativa y poesía (*Revista Nueva, La patria de Cervantes*) o incluso periódicos sobre moda y costumbres femeninas (*Periódico de las damas*), así como publicaciones científicas sobre temas médicos (*Revista frenopática española*). Estas páginas difieren en sus objetivos, temas y formatos y el corpus es muy heterogéneo e incluye variaciones en las fuentes, diferente calidad de impresión y color de fondo y páginas con notas marginales manuscritas o elementos visuales incrustados, como sellos o ilustraciones. Esta diversidad refleja fielmente las condiciones reales de la digitalización, pero también representa un desafío para los sistemas de transcripción basados en IA, que suelen estar entrenados con corpus estandarizados.

En Transkribus todos los documentos del corpus de prueba se procesaron directamente utilizando el modelo *Coloso Español*[5], modelo basado en IA que se entrenó con la colaboración de varios investigadores utilizando sus modelos de español para textos históricos de diferentes períodos, desde manuscritos

[3] https://sourceforge.net/projects/tesseract-ocr.mirror/
[4] https://www.ibm.com/es-es/granite
[5] https://readcoop.eu/model/coloso-espanol

medievales hasta documentos del siglo XX. Para nuestro proceso, fue necesario utilizar los créditos de pago de Transkribus y exportar los resultados en dos formatos, con y sin saltos de línea (GRESEL1_run1 y GRESEL2_run2, respectivamente), para finalmente configurar los archivos TXT en el formato requerido por las tareas compartidas de PastReader.

Con el modelo monomodal Tesseract, instalado en un equipo Windows, se realizaron dos ejecuciones experimentales para evaluar el rendimiento en diferentes configuraciones. En la primera ejecución (GRESEL1_run2), se ejecutó el modelo predeterminado de Tesseract directamente sobre el corpus de entrenamiento sin ninguna adaptación adicional y con parámetros estándar. En la segunda ejecución (GRESEL1_run3), se ajustó el modelo de la versión 5.5.0 de Tesseract de GitHub[6] de forma iterativa sobre la base de diferente número de páginas (100, 1000, 2000 y todas) del corpus de entrenamiento. Los modelos resultantes se evaluaron con el corpus de validación para seleccionar el mejor, aunque como los documentos históricos se desvían de los patrones para los que está optimizado el modelo LSTM[7] de Tesseract, su eficacia fue peor que la esperada.

La capacidad para comprender imágenes y responder preguntas sobre ellas de los modelos multimodales, convierte al modelo Granite3.2-vision:2b en la mejor opción para transcribir documentos. Aunque el refinamiento o ajuste fino para OCR es muy recomendable, debido a limitaciones de tiempo y a las restricciones debidas al uso de un ordenador con poca memoria, esta opción queda pendiente para un futuro próximo.

La aproximación realizada con el modelo multimodal se basa en el uso de *prompts* (instrucciones o indicaciones al sistema)[8]. Cada *prompt* requiere, de forma similar a la interacción con un modelo de *chat*, (a) una parte para indicar al sistema que no invente información ni modifique el texto, mientras intenta obtener el texto original sin añadir nada, (b) una imagen, (c) un mensaje del usuario o *prompt* del usuario ("Por favor, realice el OCR en este documento en español"), y (d) lo que esperamos del modelo: el texto resultado del OCR.

Un reto importante fue el uso de memoria RAM para cargar los archivos de imagen del corpus. El ordenador utilizado, un PC para juegos con 32 GB de RAM, era insuficiente para procesar todas las imágenes del corpus a resolución completa. Hubo que reducir la resolución a 414 x 585 píxeles.

Además, las imágenes del corpus tienen relaciones de aspecto ligeramente diferentes, por lo que hubo que adaptarlas para que todas tuvieran las mismas dimensiones, conservando su relación de aspecto original. En segundo lugar, se

[6] https://github.com/tesseract-ocr/tesstrain.git

[7] Un modelo LSTM (Memoria a Corto y Largo Plazo) es un tipo de red neuronal recurrente (RNN) diseñada para
manejar secuencias de datos y aprender patrones a largo plazo.

[8] https://colab.research.google.com/github/huggingface/cookbook/blob/main/notebooks/en/fine_tuning_granite_vi sion_sft_trl.ipynb

realizó un proceso de cuantificación[9] para reducir el consumo de memoria durante los procesos y el refinamiento. La configuración utilizada de los hiperparámetros se basó principalmente en estimaciones predeterminadas o iniciales, con pruebas mínimas.

El corpus de entrenamiento y de validación fueron divididos aleatoriamente en dos partes: 90% para entrenamiento y 10% para pruebas. Tras la obtención del modelo ajustado como se ha indicado, fue necesaria la adaptación del corpus de prueba al formato de chat, y tras su procesado los resultados formaron GRESEL2_run1.

2. Análisis de los resultados y conclusiones

La figura 2 muestra un comportamiento aceptable de todas nuestras ejecuciones. La mayoría de ellas superan el run *baseline* (resultado base para comparación) en las diferentes métricas indicando un rendimiento mínimo bueno.

La *precisión de los resultados* se evaluó con métricas tanto a nivel de carácter como son la Distancia de Levenshtein y NED (Distancia de Edición Normalizada), que contabilizan el número de operaciones de edición necesarias, como con WER (Tasa de Error de Palabras), que cuantifica las discordancias léxicas. En estas métricas los valores más bajos indican un mejor rendimiento. La calidad semántica se midió utilizando BLEU (Estudio de Evaluación Bilingüe), que evalúa la superposición de n-gramas, y la familia de métricas ROUGE: ROUGE-1 se centra en el *recall* a nivel de término, ROUGE-2 captura las relaciones contextuales de corto alcance y ROUGE-L/ROUGE-LSum evalúan la coherencia general del discurso. En estas métricas, los valores más altos indican mejor conservación del significado lingüístico, incluso cuando existen diferencias a nivel de carácter.

Dado que los tres enfoques son muy diferentes, se analizan uno a uno en lo que sigue.

Los resultados de ejecuciones (runes) con Transkribus muestran que cuando se preservan las líneas de corte (GRESEL1_run1), se obtiene mejor rendimiento en métricas de precisión como Levenshtein, WER y NED, que cuando se unen las líneas de corte (GRE-SEL2_run2). Es una consecuencia al preservar el formato de las transcripciones del corpus *groundtruth* (las buenas según los organizadores).

Transkribus uniendo las líneas obtuvo un mejor rendimiento en las métricas semánticas, BLUE y ROUGE, incluso alcanzando el segundo puesto en BLUE.

Los resultados de las ejecuciones de Tesseract coinciden con los experimentos realizados con el conjunto de datos de entrenamiento. La versión sin refinamiento (GRESEL1_run2) obtuvo mejores resultados en seis de las ocho métricas, y la versión con refinamiento (GRESEL1_run3) solo obtuvo una puntuación superior en la Distancia de Lavenshtein (segundo puesto) y NED. Estos resultados sugieren que el proceso de refinamiento estaba sobreajustado a

[9] https://github.com/artidoro/qlora

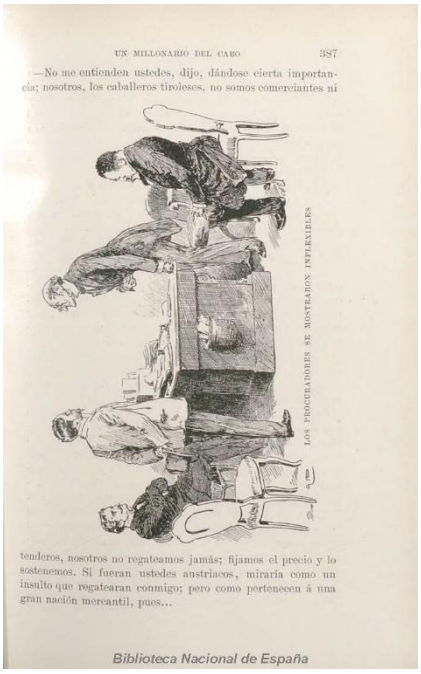
características específicas del conjunto de entrenamiento o que no se generaliza adecuadamente a consecuencia de la variabilidad del corpus de entrenamiento. No obstante, en las métricas semánticas como ROUGE-L y BLEU estuvo muy cerca de las del *baseline*.

La ejecución con el modelo Granite3.2-vision:2b, fue la mejor en general ya que logró un segundo puesto en cinco de las ocho métricas de evaluación, incluyendo WER y todas las variantes de ROUGE. De hecho, las diferencias entre Granite y el sistema con mejor rendimiento OCRTITS (Narbona and Ros, 2025) en estas métricas fueron mínimas. También ocupó el tercer puesto en BLEU, solo por detrás de la ejecución de Transkribus que unía saltos de línea (GRESEL2_run2) y del run ganador. Sin embargo, Granite obtuvo un rendimiento inferior en precisión a nivel de carácter, ocupando el cuarto lugar tanto en Distancia de Levenshtein como en NED. Estos resultados sugieren que el modelo muestra una buena comprensión semántica y reconocimiento a nivel de palabra con el *ground truth*, pero tiene margen de mejora en precisión a nivel de caracteres. Dado el tamaño compacto del modelo y las limitaciones del hardware mencionadas anteriormente, el rendimiento alcanzado es notable.

Agradecimientos

Este trabajo se enmarca en el proyecto coordinado (UNED y UAM) financiado por el Ministerio de Ciencia e Innovación y la AEI, GRESEL (*AI generation results enriched with simplified explanations base don linguistic features*, PID2023-151280OB) y el contrato PTA2023-023812-I de la convocatoria MICIU/AEI/10.13039/501100011033.

Figuras, tablas, mapas o gráficas



UN MILLONARIO DEL CABO
387
—No me entienden ustedes, dijo, dándose cierta importancia; nosotros, los caballeros tiroleses, no somos comerciantes ni tenderos, nosotros no regateamos jamás; fijamos el precio y !o sostenemos. Si fueran ustedes austriacos, miraría como un insulto que regatearan conmigo; pero como pertenecen á una gran nación mercantil, pues...

Figura 1. Ejemplo de página (a) con su transcripción alineada (b).

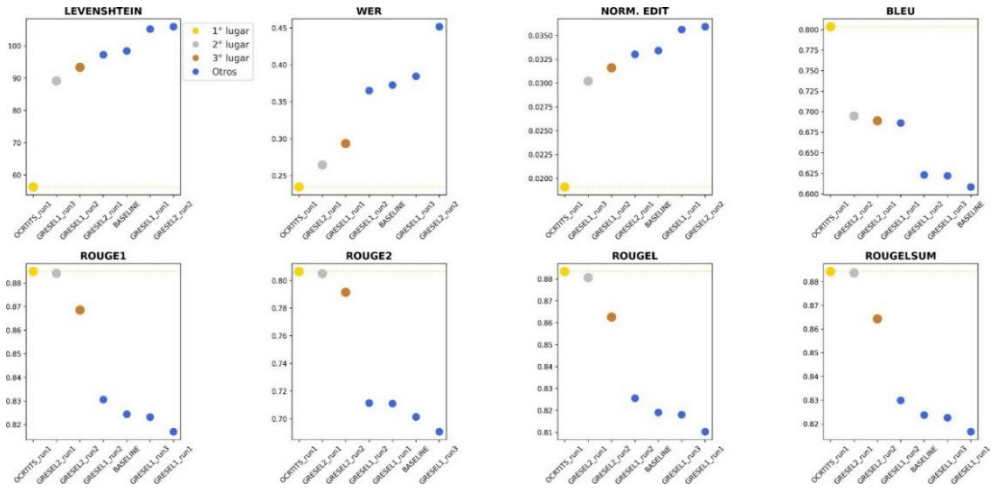


Figura 2. Scattered plots para cada métrica ordenados por las mejores puntuaciones.

Bibliografía

GARCIA-SERRANO, A., & MENTA GARUZ, A. (2022), «La inteligencia artificial en las humanidades digitales: dos experiencias con corpus digitales», *Revista de Humanidades Digitales*, 7, pp. 19–39. https://doi.org/10.5944/rhd.vol.7.2022.30928

LASTRA-DÍAZ, J. J., GOIKOETXEA, J., HADJ TAIEB, M. A., GARCIA-SERRANO, A., BEN AOUICHA, M., AGIRRE, E., & SÁNCHEZ, D. (2021), «A large reproducible benchmark of ontology-based methods and word embeddings for word similarity», *Information Systems*, 96, art. 101636. https://doi.org/10.1016/j.is.2020.101636

MACICIOR-MITXELENA, J. (2025), «Transcribing history with Tesseract: A monomodal OCR approach in the PastReader 2025 Shared Task», en *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with SEPLN 2025*, CEUR Workshop Proceedings, Zaragoza. (Aceptado para publicación; vol. y URL por determinar).

MARTÍNEZ-FERNÁNDEZ, J. L., ROMÁN, J. V., GARCIA-SERRANO, A. M., & GONZÁLEZ-CRISTÓBAL, J. C. (2006), «Combining textual and visual features for image retrieval», en C. PETERS *et al.* (eds.), *Accessing Multilingual Information Repositories*, Berlin/Heidelberg, Springer, pp. 680–691. https://doi.org/10.1007/11878773_75

MONTEJO-RÁEZ, A., SÁNCHEZ-NOGALES, E., EXPÓSITO-ÁLVAREZ, G., UREÑA-LÓPEZ, L. A., MARTÍN-VALDIVIA, M. T., COLLADO-MONTAÑEZ, J., CABRERA-DE CASTRO, I., CANTERO-ROMERO, M. V., & ORTUÑO-CASANOVA, R. (2025), «Overview of PastReader shared task in IberLEF 2025: Transcribing texts from the past», *Procesamiento del Lenguaje Natural*, 75.

MORENO-SANDOVAL, A., PORTA, J., CARBAJO-CORONADO, B., TORTEROLO, Y., & SAMY, D. (2025), «The financial document causality

detection shared task (FinCausal 2025)», en C.-C. CHEN *et al.* (eds.), *Proceedings of the Joint Workshop of FinNLP 9, FNP 6, and LLMFinLegal 1*, Abu Dhabi, ACL, pp. 214–221. https://aclanthology.org/2025.finnlp-1.21/

NARBONA, A., & ROS, S. (2025), «Beyond traditional OCR: Exploring the efficiency of LLMs in document processing», en *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with SEPLN 2025*, CEUR Workshop Proceedings, Zaragoza. (Aceptado para publicación; vol. y URL por determinar).

PORTA-ZAMORANO, J., TORTEROLO-ORTA, Y., & MORENO-SANDOVAL, A. (2023), «LLI-UAM Team at FinancES 2023: Noise, data augmentation and hallucinations», en *Proceedings of IberLEF 2023, co-located with SEPLN 2023*, CEUR Workshop Proceedings, 3496, Jaén, pp. 1–11. https://ceur-ws.org/Vol-3496/finances-paper1.pdf

SANCHEZ-SALIDO, E., MENTA, A., & GARCIA-SERRANO, A. (2023), «Seeking information in Spanish historical newspapers: The case of *Diario de Madrid* (18th and 19th centuries)», *Digital Humanities Quarterly*.

TORTEROLO-ORTA, Y. A., & MIGUEZ-LAMANUZZI, M. (2025), «Fine-tuning a compact multimodal model on consumer-grade hardware at PastReader 2025», en *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2025), co-located with SEPLN 2025*, CEUR Workshop Proceedings, Zaragoza. (Aceptado para publicación; vol. y URL por determinar).

Análisis de documentos novohispanos por medio del Reconocimiento y Desambiguación de Entidades Nombradas (NERD)

Diego Jiménez-Badillo
Instituto Nacional de Antropología e Historia

Palabras clave: Reconocimiento de Entidades Nombradas; Relaciones Geográficas del siglo XVI; Aprendizaje de máquina aplicado a fuentes históricas; Humanidades geoespaciales.
Keywords: Named Entity Recognition; 16th-Century Geographic Relations; Machine Learning Applied to Historical Sources; Geospatial Humanities.

Problema de Investigación

En esta ponencia describo una investigación orientada a facilitar el análisis de las *Relaciones Geográficas de Nueva España*, un conjunto de reportes redactados entre 1577 y 1585 en varias provincias de México y Guatemala por orden del rey Felipe II.

El corpus está constituido por más de 1000 folios manuscritos en cuya redacción participaron informantes indígenas y oficiales españoles (Figura 1). Su extensión es de casi un millón de palabras, que han sido transcritas por varios editores modernos. Estos reportes son considerados fuentes imprescindibles sobre la historia, la geografía y la cultura del siglo XVI (Acuña 1982-1988; De la Garza, Del Paso y Troncoso 1905-1912). Otra parte importante del corpus son 78 mapas pictográficos que complementan las descripciones textuales (Figuras 2 y 3), algunos fueron pintados por artistas nativos siguiendo las convenciones indígenas de representación de lugares y pasajes y otros muestran clara influencia europea (Mundy 1996; Robertson 1972).

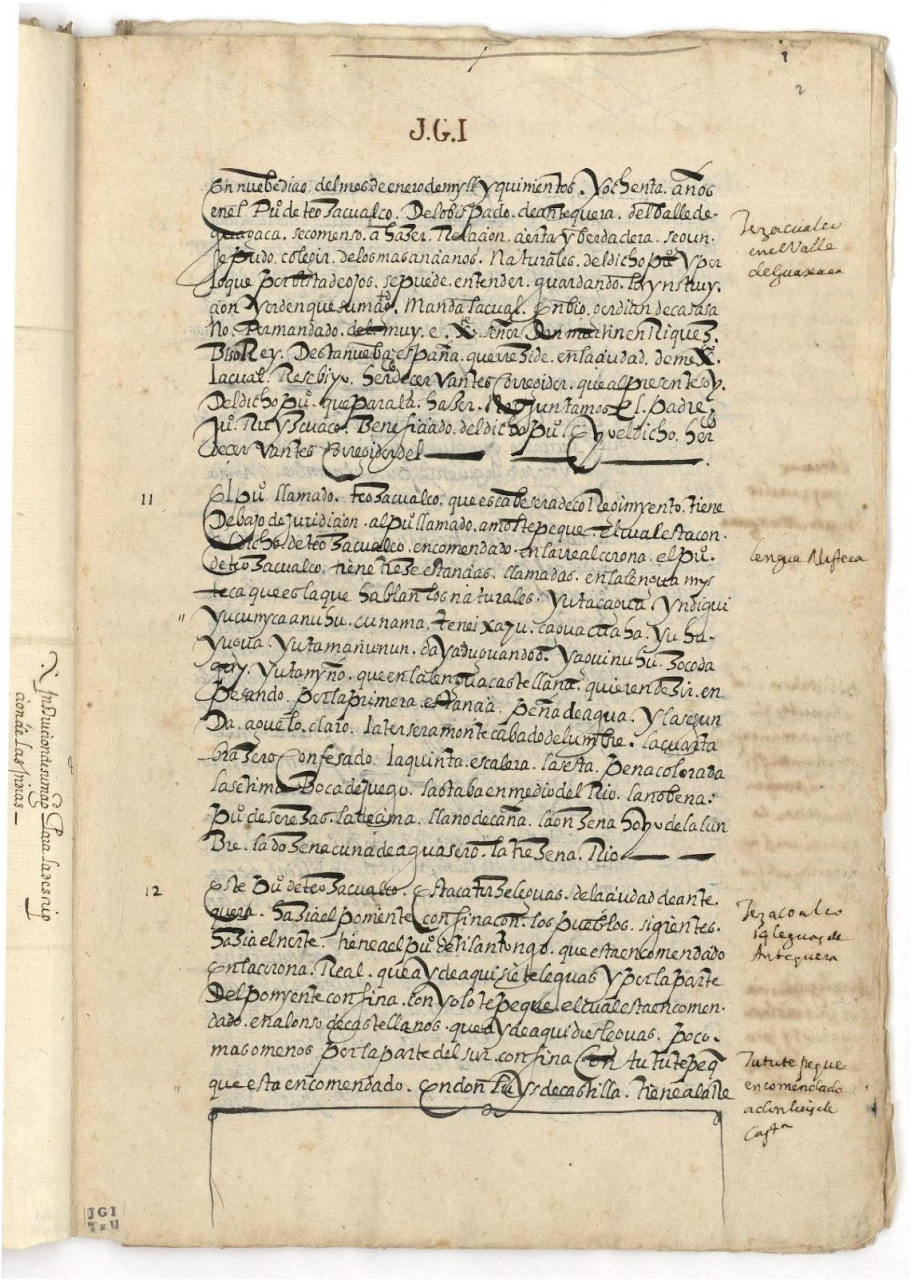


Figura 1. Un ejemplo del tipo de folios manuscritos que componen el corpus de las *Relaciones Geográficas de Nueva España (1577-1585)*. Este corresponde al inicio de la *Relación de Teozacualco y Amoltepec.* Reproducido por cortesía de la Benson Latin American Collection. The General Libraries, The University of Texas at Austin (Ms_txu-oclc-34022876-JGI-XXV-3).

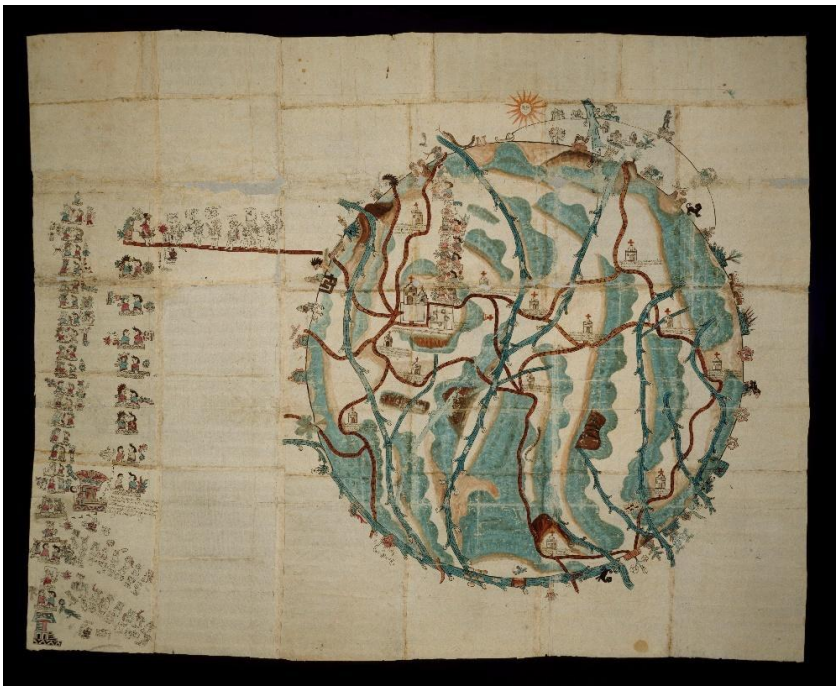


Figura 2. El mapa de Teozacualco, uno de las dos pinturas de tradición indígena que acompañan la *Relación de Teozacualco y Amotepec.* Reproducido por cortesía de la Benson Latin American Collection. The General Libraries, The University of Texas at Austin (JGI-XXIV-5).

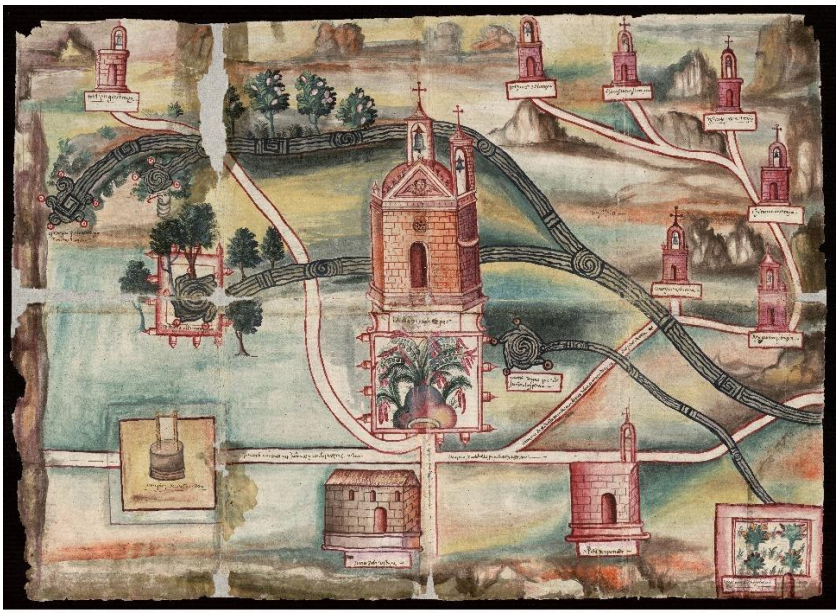


Figura 3. El mapa de Huaxtepeque, un ejemplo de pintura con influencia europea que acompaña la *Relación Geográfica de Huaxtepeque*. Reproducido por cortesía de la Benson Latin American Collection. The General Libraries, The University of Texas at Austin (JGI-XXIV-2).

Los reportes son producto de una encuesta de cincuenta preguntas redactadas por Juan de Velasco, cosmógrafo del Consejo de Indias, con el

propósito de conocer la organización territorial y el modo de vida de los habitantes de Nueva España. Gracias a la riqueza y variedad de las preguntas, el corpus constituye una de las fuentes más importantes para analizar la historia, la geografía, las jurisdicciones administrativas, la cultura, la religión, la economía, la interacción social y los procesos de transculturación que afectaron a las comunidades nativas y a los colonizadores.

Se conservan 168 reportes -en el Archivo de Indias de Sevilla, la Biblioteca Benson de la Universidad de Texas y la Biblioteca Real de Madrid- provenientes de igual número de cabeceras principales de Nueva España; en ellos se encuentras descripciones de 248 pueblos subordinados (i.e. pueblos sujetos), así como de 414 pueblos secundarios y aldeas. Los datos corresponden a más de la mitad de las subdivisiones políticas existentes en Nuevas España durante el tercer cuarto del siglo XVI y por ello son una valiosa fuente de historia geográfica (Cline 1972).

El volumen de datos contenidos en dichos documentos obliga a los historiados modernos -acostumbrados a leer los textos transcritos en forma impresa- a invertir una gran cantidad de tiempo para encontrar la información deseada y para detectar patrones en el contenido de los documentos. Para mejorar esta situación, un grupo de investigadores de la Universidad de Lancaster, el Instituto Nacional de Antropología e Historia de México y el Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa de la Universidad de Lisboa emprendió el proyecto *Digging into Early Colonial México.*

Decidimos combinar técnicas de procesamiento de lenguaje natural (NLP), aprendizaje de máquina (*machine learning*) y ciencias de la información geográfica para crear herramientas en línea con las cuales potenciar el aprovechamiento de estos documentos.

La ponencia se dedica a describir brevemente el protocolo de investigación y mostrar en detalle el funcionamiento de dos de las herramientas implementadas. Ambas pueden utilizarse para identificar la información deseada, extraerla, y responder preguntas de investigación significativas, tales como: ¿cuál era la organización territorial en diversas regiones de la Nueva España en la época de los Relaciones Geográficas? ¿Cómo cambiaron los patrones de asentamiento antes y después de la conquista española? ¿Cómo era el aprovechamiento de recursos naturales; ¿Cómo era la movilidad terrestre y acuática en diversas zonas? etc.

2. Metodología

Una de las estrategias del proyecto fue el perfeccionamiento de la metodología denominada Análisis Geográfico de Textos (*Geographical Text Analysis*), propuesta hace algunos años por Ian Gregory, Patricia Murrieta-Flores (dos participantes de este proyecto), en colaboración con otros investigadores (Murrieta-Flores et al. 2015). Dicha metodología incluye el desarrollo de un

modelo de Reconocimiento y Desambiguación de Entidades Nombradas (*Named Entity Recognition y Desambiguation*, NERD) para clasificar y etiquetar palabras de acuerdo a una serie de categorías analíticas relevantes para la investigación histórica. Dicho modelo ha sido adaptado específicamente a las características de las *Relaciones Geográficas*, caracterizadas por estar escritas en castellano premoderno y por contener una gran cantidad de términos en lenguas indígenas, lo que supone un reto significativo que ha sido superado gracias a la experiencia de los miembros del proyecto de la Universidad de Lisboa, en especial Bruno Martins y Luis Miguel Paiva Sampaio Santos. A continuación, resumo las principales fases del proyecto (descripciones más detalladas pueden consultarse en Jiménez-Badillo et al. 2020, 2021, y Murrieta-Flores et al. 2022).

1. Preprocesamiento del Corpus
    - Digitalización y limpieza de los textos por medio de técnicas de Reconocimiento Óptico de Caracteres (OCR).
    - Conversión de los textos a formato legible por ordenador.
2. Reconocimiento de Entidades Nombradas (NER)
    - Identificación automática de entidades (categorías de interés histórico como personas, instituciones, plantas, animales, datos culturales, rituales y religiosos, recursos naturales, etc.). La identificación se concentró especialmente en el reconocimiento de topónimos, pues dichas entidades son clave para georreferenciar el resto de datos históricos, geográficos y culturales.
    - Uso de técnicas de PLN adaptándolas a los textos históricos y multilingües (principalmente castellano premoderno).

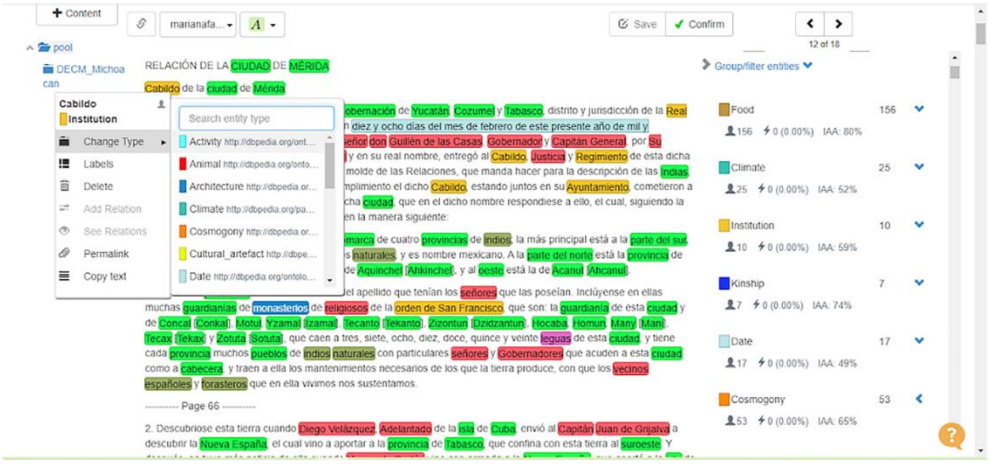


Figura 4. Vista de las anotaciones obtenidas durante la fase 4 del proyecto. Los colores corresponden a las distintas categorías analíticas que consideramos relevantes para la investigación histórica.

3. Desambiguación de Topónimos
    - Resolución de ambigüedades lingüísticas y geográficas.
    - Georreferenciación (asignación de coordenadas).

- Creación de un directorio de topónimos de Nueva España, siglo XVI.

4. Anotación del Corpus
   - Análisis sintáctico (para distinguir partes del discurso).
   - Anotaciones temáticas personalizadas, por ejemplo, deidades, productos, etc. (Figura 4).
   - Uso de esquemas XML para estructurar la información.
5. Análisis de Colocación Geográfica
   - Adaptación del análisis de colocación de la Lingüística de Corpus para ligar automáticamente los topónimos con el resto de las entidades reconocidas.
   - Identificación de topónimos cercanos a palabras clave (por ejemplo, “maíz” con otras entidades, por ejemplo, lugares relacionados con agricultura).
6. Implementación de las herramientas en línea para extraer información
   - Creación de la herramienta GTA.
   - Creación de la plataforma en línea “Relaciones Geográficas de la Nueva España (1577-1585)”

3. Resultados obtenidos

El proyecto ha producido diversos resultados descritos detalladamente en varias publicaciones (Jiménez-Badillo et al. 2020, 2021, y Murrieta-Flores et al. 2022, 2025). En esta ponencia, me concentro en dos que no han recibido la difusión debida. La primera herramienta es una plataforma web que intenta reunir por primera vez toda la información existente del corpus de las Relaciones Geográficas. Está dirigida a aquellos usuarios que deseen consultar el corpus sin incurrir en procedimientos de búsqueda complejos y que además deseen obtener información en distintos medios, es decir imágenes en alta resolución de los folios manuscritos, las distintas transcripciones que se han publicado a lo largo del tiempo, los mapas que muestran la localización de todos los topónimos mencionados y los mapas pictográficos del corpus. (Figuras 5, 6, y 7).

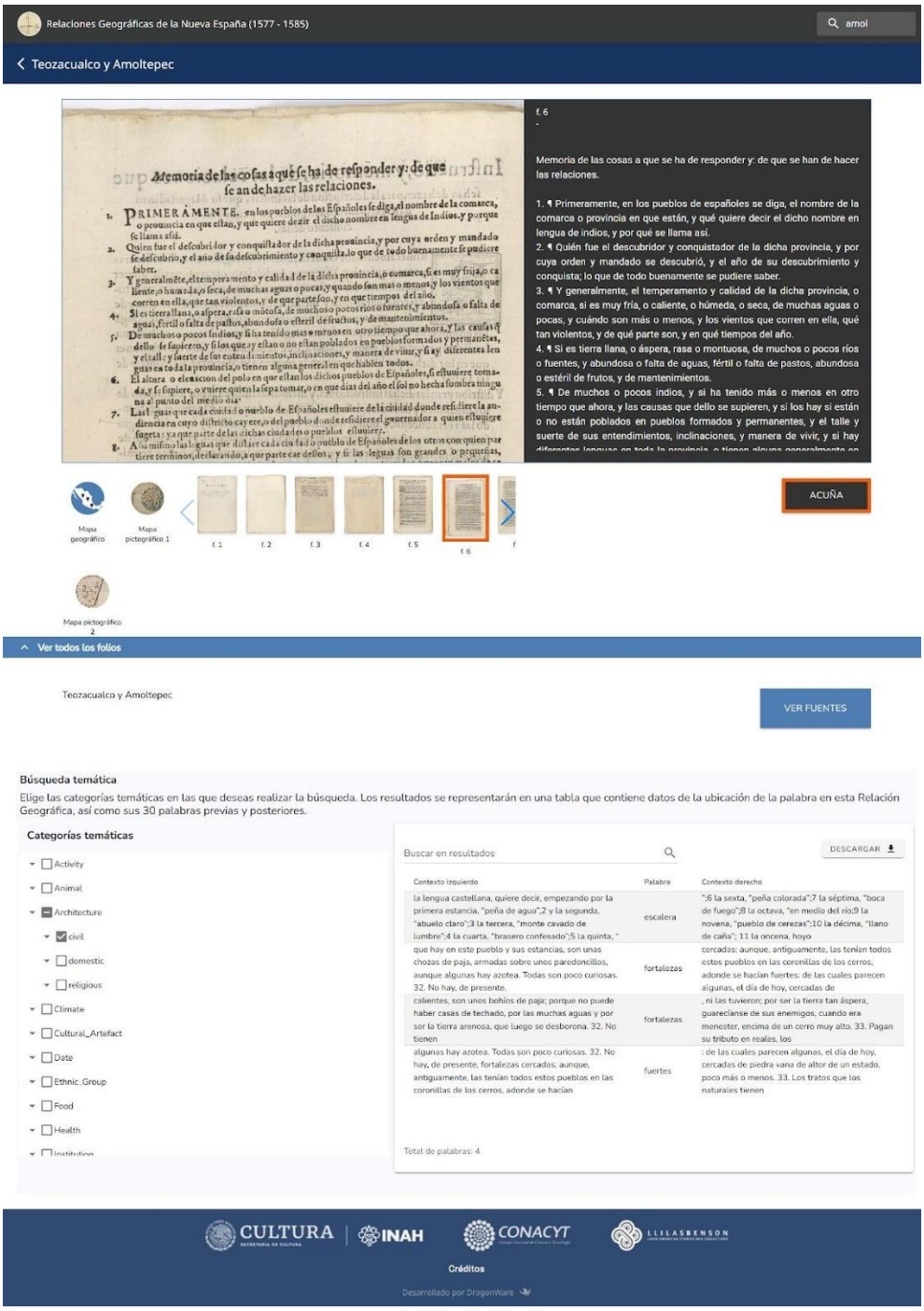


Figura 5. Una página de la plataforma en línea para consultar y descargar información de las *Relaciones Geográficas de Nueva España (1577-1585)*. En la parte superior pueden verse uno de folios de la *Relación de Teozacualco y Amoltepec* y su transcripción. Abajo a la izquierda, se encuentra la interfaz para seleccionar las categorías de datos que pueden descargarse y a la derecha, una tabla (descargable en formato Excel) muestra las instancias de datos pertenecientes a la categoría analítica seleccionada, así como fragmentos de los textos que conforman el contexto de cada palabra clave.

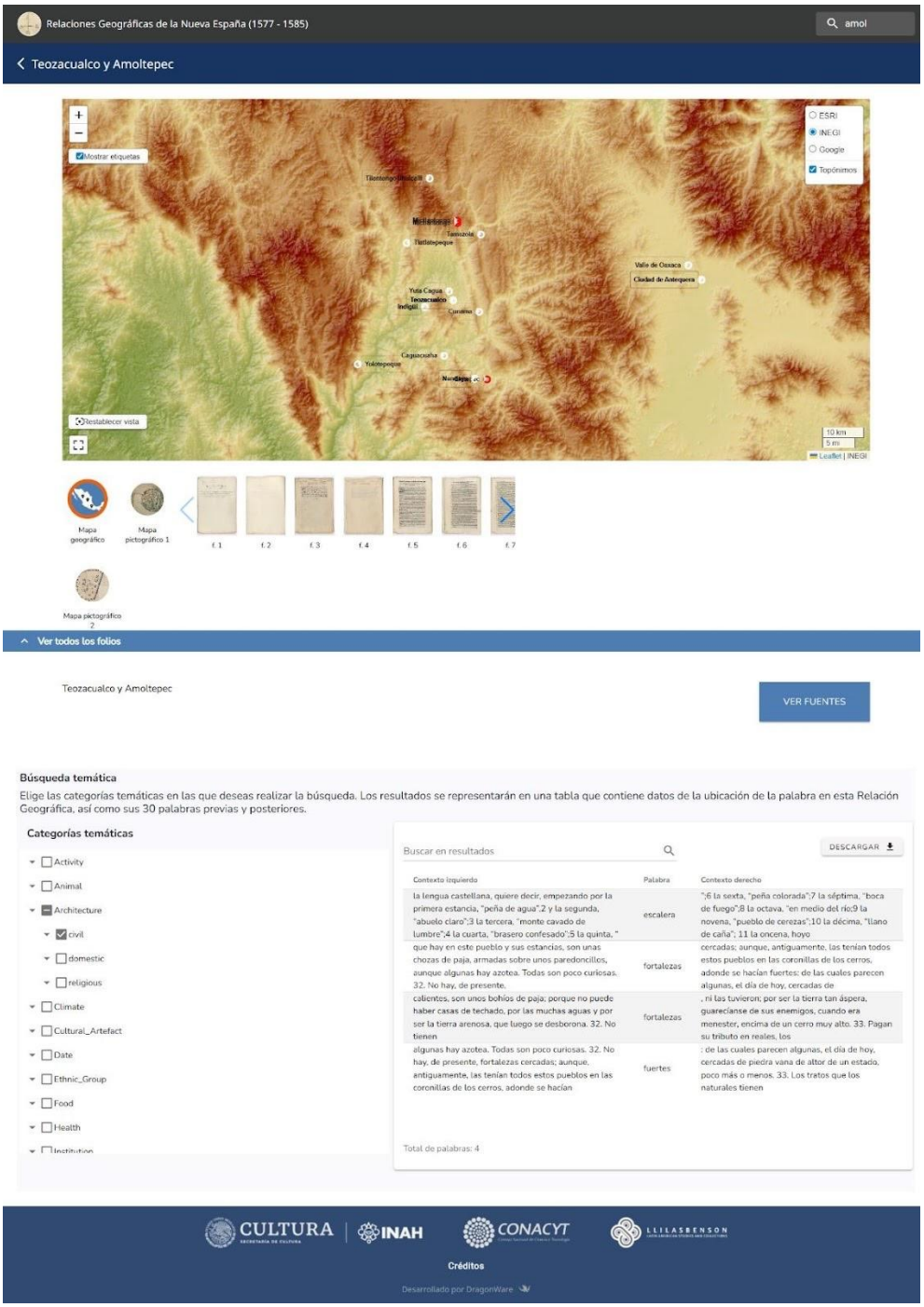


Figura 6. Otra página de la plataforma en línea para consultar y descargar información de las *Relaciones Geográficas de Nueva España (1577-1585)*. En la parte superior puede verse un mapa con la geolocalización de los topónimos mencionados en el documento.

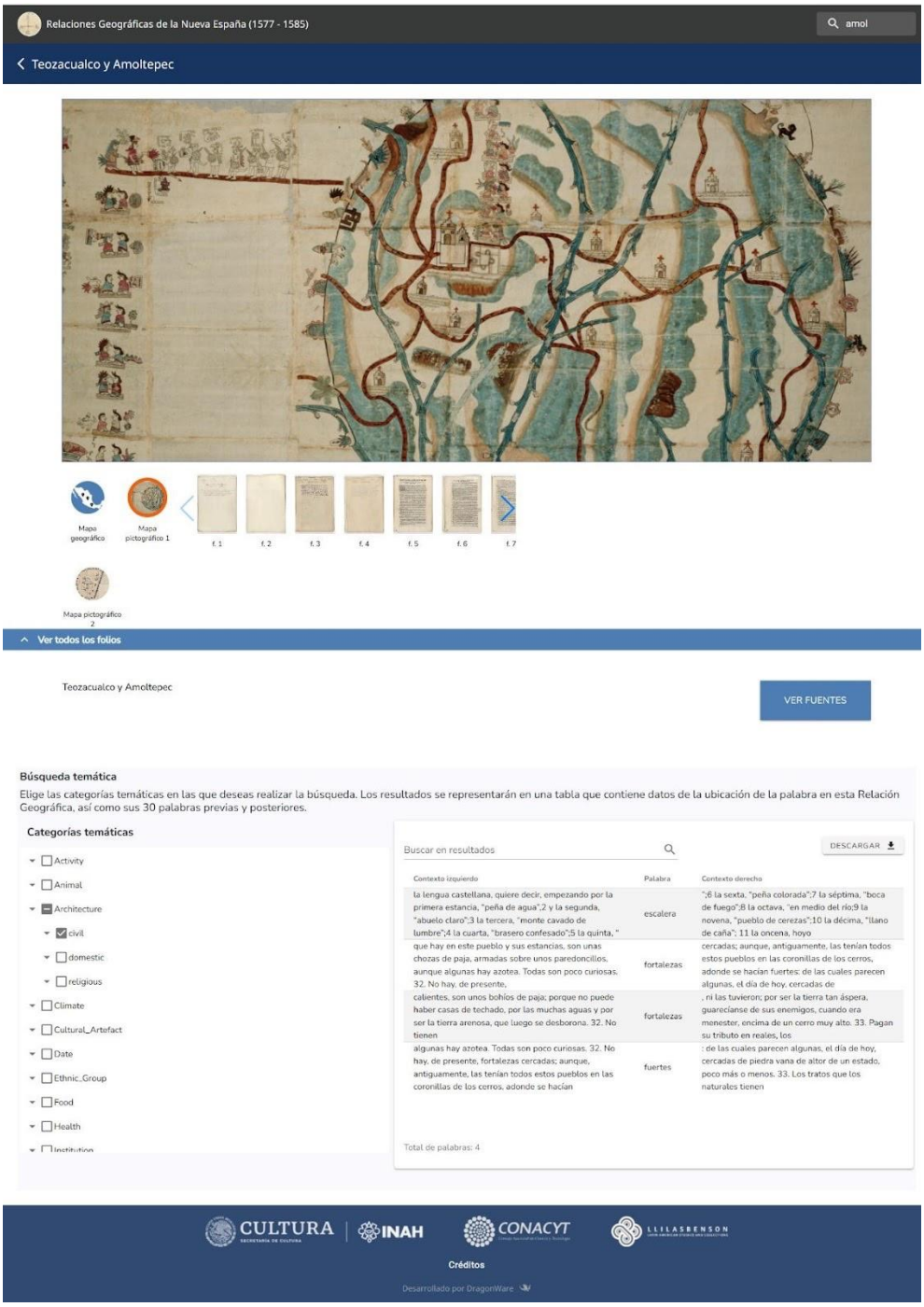


Figura 7. Una tercera opción es desplegar el mapa pictográfico asociado al documento consultado, en este caso la *Relación de Teozacualco y Amoltepec*.

En dicha plataforma pueden descargarse tablas de datos sobre las siguientes categorías de interés histórico:

- Actividades de tipo político, civil, bélico, ritual, doméstico, agrícola, económico y relacionadas a la minería.
- Animales (domesticados y silvestres como reptiles, mamíferos, anfibios, pájaros, acuáticos, etc.).
- Arquitectura doméstica, civil y religiosa.

- Datos climáticos
- Artefactos culturales: vestimentas, objectos rituales, domésticos, armas.
- Fechas
- Grupos étnicos
- Comida
- Datos sanitarios (enfermedades y remedios)
- Instituciones
- Lenguajes
- Movilidad terrestre y acuática
- Recursos naturales
- Datos de localización espacial: topónimos, jurisdicciones civiles, eclesiásticas, rasgos geográficos del paisaje, etc.
- Personajes mencionados
- Categorías de grupos sociales

La segunda herramienta que presentaré en la ponencia es una herramienta de software llamada GTA que permite realizar consultas personalizadas de acuerdo a intereses de cada investigador para extraer las palabras o frases más significativas del corpus y analizar su contexto. Con ese recurso pueden investigarse temas como la correlación geográfica de distintos hechos históricos, el análisis de datos de carácter social, cultural o histórico asociados a nombres de lugar geo-localizados, la delimitación de jurisdicciones territoriales antiguas sobre la base de información espacial contenida en los documentos, o bien, el descubrimiento de patrones de interacción entre personas, actividades, instituciones y lugares. La búsqueda se realiza introduciendo una palabra clave de interés de acuerdo con un umbral de proximidad predefinido, generalmente definido en términos del número máximo de palabras entre el término de búsqueda y su contexto (Figura 8).

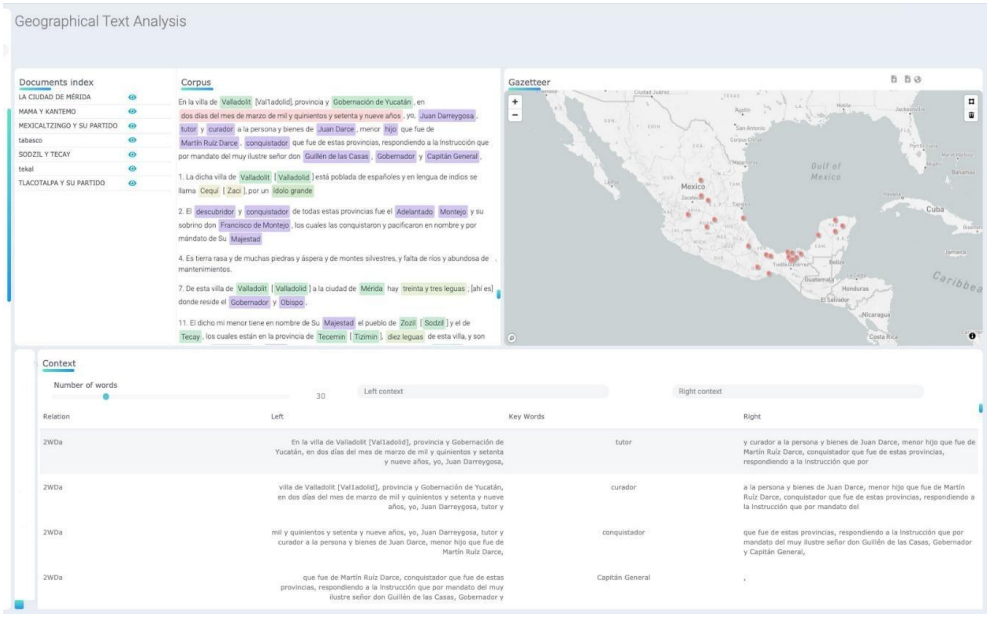


Figura 8. Interfaz de la herramienta GTA durante un análisis de palabras clave en contexto. Pueden observarse cuatro ventanas: En *Documents Index* (arriba a la izquierda), el usuario elige los textos que desea analizar, éstos pueden corresponder a un solo documento, a un grupo, o a todo el corpus. En la ventana *Corpus* (arriba, al centro) la herramienta despliega el contenido histórico del texto resaltando los términos significativos que resultan de la búsqueda. En el ejemplo ilustrado, la extracción estuvo basada en la combinación de dos criterios: “persona = género masculino” y “persona = profesión”. La ventana inferior (*Context*) muestra los resultados de la consulta en cuatro columnas (nombre del documento; contexto a la izquierda; palabra clave; contexto a la derecha; y posición exacta de la palabra clave en el texto), mientras que el mapa (pantalla *Gazetteeer*) señala la localización geográfica de los datos.

De esta manera, si un investigador está interesado en investigar aspectos económicos relacionados con la agricultura, podría recuperar, por ejemplo, todas las frases en el contexto de palabras como maíz, siembra, subsistencia, mantenimiento, o comida, que aparecerían vinculadas a un mapa que muestra las ubicaciones geográficas donde aparecen dichos términos. Al facilitar la búsqueda de palabras en contexto, un historiador puede extraer en pocos segundos los datos, en lugar de las semanas que tardaría en acceder a la información sin la herramienta GTA.

En conclusión, durante la presentación espero mostrar cómo la adaptación de técnicas de PLN a textos históricos multilingües y escritos en castellano premoderno contribuye a mejorar la práctica de la investigación histórica, especialmente la del período novohispano de México.

Bibliografía

ACUÑA, R. (ed.) (1982–1988), *Relaciones Geográficas del siglo XVI*, 10 vols., México, Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México.

CLINE, H. F. (1972), «The Relaciones Geográficas of the Spanish Indies, 1577–1648», en H. F. CLINE (ed.), *Handbook of Middle American Indians*, vol. 12: *Guide to Ethnohistorical Sources, Part One*, Austin, University of Texas Press, pp. 183–242.

DE LA GARZA, M. (ed.) (1983), *Relaciones histórico-geográficas de la gobernación de Yucatán (Mérida, Valladolid y Tabasco)*, 1.ª ed., 2 vols., México, Universidad Nacional Autónoma de México.

DEL PASO Y TRONCOSO, F. (ed.) (1905–1912), *Papeles de Nueva España. Segunda serie: Geografía y Estadística*, vols. 1, 3–7, Madrid, Sucesores de Rivadeneyra.

JIMÉNEZ-BADILLO, D., MURRIETA-FLORES, P., MARTINS, B., GREGORY, I., FAVILA VÁZQUEZ, M., & LICERAS-GARRIDO, R. (2020), «Developing geographically oriented NLP approaches to sixteenth-century historical documents: Digging into Early Colonial Mexico», *Digital Humanities Quarterly*, 14(4) (n.º monográfico editado por H. Alpert-Abrams & C. McCarl). https://dhq.digitalhumanities.org/vol/14/4/000490/000490.html

JIMÉNEZ-BADILLO, D., MURRIETA-FLORES, P., MARTINS, B., GREGORY, I., FAVILA VÁZQUEZ, M., LICERAS-GARRIDO, R., BELLAMY, K., WON, M., & ESTIMA, J. (2021), «Análisis histórico-geográfico de documentos novohispanos del siglo XVI mediante técnicas de lingüística computacional y análisis espacial», en D. JIMÉNEZ-BADILLO (ed.), *Patrimonio digital. Métodos computacionales y medios interactivos para estudiar y divulgar el patrimonio cultural*, México, Instituto Nacional de Antropología e Historia, pp. 35–64. ISBN impreso: 978-607-539-602-6; ISBN PDF: 978-607-539-597-5. https://mediateca.inah.gob.mx/repositorio/islandora/object/libro%3A853

MUNDY, B. (1996), *The Mapping of New Spain: Indigenous Cartography and the Maps of the Relaciones Geográficas*, Chicago, University of Chicago Press.

MURRIETA-FLORES, P., BARON, A., GREGORY, I., HARDIE, A., & RAYSON, P. (2015), «Automatically analyzing large texts in a GIS environment: The Registrar General's Reports and cholera in the 19th century», *Transactions in GIS*, 19(2), pp. 296–320. https://doi.org/10.1111/tgis.12106

MURRIETA-FLORES, P., JIMÉNEZ-BADILLO, D., & MARTINS, B. (2022), «Digital resources: Artificial intelligence, computational approaches, and geographical text analysis to investigate Early Colonial Mexico», *Oxford Research Encyclopedia of Latin American History*. https://doi.org/10.1093/acrefore/9780199366439.013.977

MURRIETA-FLORES, P., JIMÉNEZ-BADILLO, D., & FAVILA-VÁZQUEZ, M. (2025), «Placing New Spain through Early Modern Big Data: Developing a

Geographical Text Analysis approach to the *Relaciones Geográficas* de Nueva España», en M. COOLEY & W. HUIYI (eds.), *Knowing the Empire in Early Modern China and Spain*, Ann Arbor, Lever Press.

ROBERTSON, D. (1972), «The *pinturas* (maps) of the *Relaciones Geográficas* with a catalogue», en H. F. CLINE (ed.), *Handbook of Middle American Indians*, vol. 12: *Guide to Ethnohistorical Sources, Part One*, Austin, University of Texas Press, pp. 243–278.

La memoria individual como herramienta de documentación etnográfica para la reconstrucción digital: el caso del penal de Valdenoceda

María del Pilar Bermúdez Martínez
Roberto Martín González
Mario Alaguero Rodríguez
Universidad de Burgos

Palabras clave: Memoria; digitalización; reconstrucción digital
Keywords: Memory; digitization; digital reconstruction

Introducción

Las políticas de protección del patrimonio se han consolidado tanto en el ámbito académico como profesional, respaldadas por un marco legal centrado en su conservación y valorización como bien cultural y social. Sin embargo, persisten desigualdades en la atención prestada a distintos tipos de patrimonio, especialmente aquellos menos visibles o vinculados a políticas culturales dominantes (Alaguero, 2023). En particular, ciertas edificaciones del último siglo se encuentran en riesgo de desaparecer por demoliciones, abandono o crecimiento urbano, lo que deja fuera del foco patrimonial a construcciones recientes frente a las más antiguas o monumentales (Calduch Cervera, 2009).

Esta pérdida afecta directamente a la memoria colectiva, al eliminar espacios cuya historia aún no ha sido documentada. Conservar su recuerdo es esencial para evitar interpretaciones incompletas del pasado. Benito del Pozo (2012) destaca el valor de las fuentes orales cuando los documentos escritos son escasos, aunque advierte sobre su subjetividad.

En este sentido, la memoria individual contiene información irreemplazable que debe registrarse de forma ordenada para preservar tanto el patrimonio material como el inmaterial contemporáneo. Testimonios, ilustraciones, biografías y relatos personales pueden rescatar información sobre espacios olvidados o en ruinas. Combinados con tecnologías digitales, estos recursos ofrecen nuevas posibilidades de conservación y difusión.

Este trabajo se centra en el registro de la memoria individual como herramienta para la reconstrucción virtual del Penal de Valdenoceda, edificio vinculado al patrimonio industrial e histórico de Burgos. Los testimonios orales, las biografías, las ilustraciones y otros relatos personales permiten recuperar información valiosa sobre espacios en ruinas, olvidados o próximos a desaparecer. Su diversidad de usos y estado de deterioro lo convierten en un

caso representativo de los retos y oportunidades que plantea el uso de la memoria como fuente documental. Estos recursos, combinados con el uso de nuevas tecnologías y procesos de digitalización, constituyen herramientas eficaces para, a partir de esa conservación, poder trazar vías de difusión del patrimonio.

El proyecto ha sido desarrollado en el Centro de Innovación y Tecnología en Videojuegos y Comunicación Audiovisual de la Universidad de Burgos, con financiación del Ministerio de Educación, Formación Profesional y Deportes.

Caso de estudio

El Penal de Valdenoceda se ubica en la comarca burgalesa de Las Merindades, en la pedanía de Valdenoceda de Valdivielso, a unos 70 km al norte de Burgos. Rodeado por el río Ebro, la Sierra de Tesla y el Alto de la Mazorra, su historia comienza en el siglo XIX como fábrica de harinas. A finales de ese siglo se convirtió en la primera fábrica de seda artificial de España, gracias a una central hidroeléctrica. En 1928, la producción se trasladó a Burgos y, durante la Guerra Civil, el recinto fue transformado en prisión. Esta función se oficializó el 15 de noviembre de 1938 mediante orden del ministro franquista Tomás Domínguez Arévalo (Boletín Oficial del Estado, 1938).

Cerrado en 1943, el complejo fue reutilizado como granja, almacén y residencia, hasta quedar abandonado. Hoy, solo permanecen en pie cuatro edificios, algunos en ruina. En 2021, fue incluido en la Lista Roja del Patrimonio de Hispania Nostra por su valor histórico y riesgo de desaparición (Avintia, 2022).

La elección del Penal de Valdenoceda como caso de estudio responde a su condición de patrimonio reciente en peligro, así como a su carácter representativo de los usos múltiples que puede asumir un edificio a lo largo del tiempo. Este constituye un ejemplo del potencial que ofrece el patrimonio etnográfico y la memoria como fuentes fundamentales para la conservación y reconstrucción del pasado. En un contexto donde la arquitectura carcelaria de la posguerra ha sido, en muchos casos, olvidada o borrada, este trabajo contribuye a preservar un espacio significativo tanto por su valor material como por la carga simbólica y testimonial que encierra. En este aspecto, aún hay personas que conservan la memoria de lo ocurrido, y es fundamental aprovechar ese legado vivo para documentarlo antes de que desaparezca.

Metodología

El registro de la memoria individual protagonista de este estudio ha formado parte de un proyecto más amplio: la reconstrucción virtual del Penal de Valdenoceda. Esta fase de registro del patrimonio inmaterial se ha incluido dentro de la metodología que se ha seguido, la Matriz Extendida de Cinco Puntos (Demetrescu & Ferdani, 2021).

En primer lugar, se ha llevado a cabo la recopilación de datos. Paralelamente a este estudio, se ha recopilado la información contenida en las

fuentes documentales tradicionales. Para ello, se han utilizado múltiples recursos como documentación histórica, fotografías antiguas, inventarios y documentación administrativa.

También se ha realizado la búsqueda de testimonios registrados documentalmente. Se han recogido testimonios de personas que conocieron el penal, tanto en formato de entrevista como mediante documentación gráfica y otros tipos de relatos. Para este estudio, se han revisado de manera sistemática diferentes recursos documentales no convencionales en el ámbito de la reconstrucción digital.

Una vez reunidos los datos, se ha procedido a la gestión y análisis de la información. Se han creado modelos simplificados (proxies) de los restos existentes, que se han enriquecido semánticamente con información y se han vinculado a una base de datos gráfica que permite comprender las relaciones cronológicas y estructurales entre los distintos elementos.

A continuación, se ha registrado la memoria. Se ha identificado qué información complementaria podría necesitarse más allá de la recogida en las fases anteriores. Para ello, se han seleccionado personas con una vinculación directa o indirecta con el penal, y se han realizado entrevistas en profundidad orientadas a registrar la información que complementara las demás fuentes documentales consultadas. Esta fase se ha realizado tras el blocking para poner el foco en los vacíos de información susceptibles de ser completados.

En la etapa de interpretación y reconstrucción virtual, se han elaborado hipótesis reconstructivas a partir de la comparación entre los modelos digitales y las fuentes documentales. Cada hipótesis se ha visualizado mediante modelos 3D, los cuales se han analizado y validado con base en los datos disponibles y el nivel de certeza de cada componente.

Finalmente, se ha desarrollado un modelo de representación. Una vez evaluadas las hipótesis, se han generado modelos más detallados con materiales y texturas que reflejan con precisión el aspecto original de las estructuras reconstruidas.

El proyecto ha culminado con la publicación y difusión de los resultados mediante herramientas digitales interactivas o publicaciones científicas. Este paso ha incluido la documentación completa del proceso, garantizando la transparencia y trazabilidad de los datos utilizados.

De otro modo, y sirviendo a modo de combinación, se han seleccionado determinadas metodologías como conservación y recopilación de la Memoria Histórica.

Resultados

Para llevar a cabo la reconstrucción del Penal de Valdenoceda, y como se menciona anteriormente, se recurrió a diversas fuentes, tanto gráficas como testimoniales. En primer lugar, se recopilaron testimonios y biografías de antiguos presos del penal. Varios de ellos fueron localizados en blogs de

memoria histórica como Exhumación de Valdenoceda o Las Merindades en la Memoria, donde se encontraron relatos y nombres que resultaron fundamentales para la investigación.

Entre los testimonios más relevantes destacan los de Ernesto Sempere, Marcos Ana y José Robledano, todos ellos reclusos del penal. Las memorias y dibujos de Sempere, los testimonios de Ana y las ilustraciones de Robledano —ubicadas en la Biblioteca Nacional— que fueron esenciales para identificar espacios y corregir interpretaciones erróneas previas.

Con el objetivo de contrastar y ampliar estos datos, se llevaron a cabo entrevistas con figuras clave como Ángel Arce, exalcalde de Valdenoceda, quien proporcionó información sobre la ubicación de diversas edificaciones. Asimismo, José María González, presidente de la Asociación de Familiares de Represaliados, aportó datos sobre las modificaciones que las instalaciones han experimentado con el paso del tiempo. Ambos ofrecieron información crucial sobre la estructura original del penal, sus transformaciones y su estado actual. Posteriormente, estos aportes se plasmaron en un plano general del recinto.

Además, la información recopilada se organizó en formatos tanto físicos como digitales, fue archivada y digitalizada. Los resultados de los testimonios y los datos históricos fueron validados por el historiador Juan José Martín García, de la Universidad de Burgos.

Esta documentación se incorporó a un repositorio y se utilizó para ubicar las distintas instalaciones dentro del recinto. A partir de esta base, y con el apoyo de fotografías y ortofotos actuales, fue posible tomar medidas precisas que facilitaron el posterior modelado.

Una vez reunida toda la información, se procedió a una catalogación basada en niveles de fiabilidad histórica. Se empleó un sistema de colores: el verde se identifican las instalaciones respaldadas por fuentes confirmadas; el amarillo, aquellas con menor respaldo documental. En el caso de las fachadas, que se organizaron a parte, el verde representaba alta fiabilidad; el rosa, un nivel medio; y el rojo, aquellas zonas con mayor grado de incertidumbre o transformaciones a lo largo del tiempo o sobre las que existía poca información.

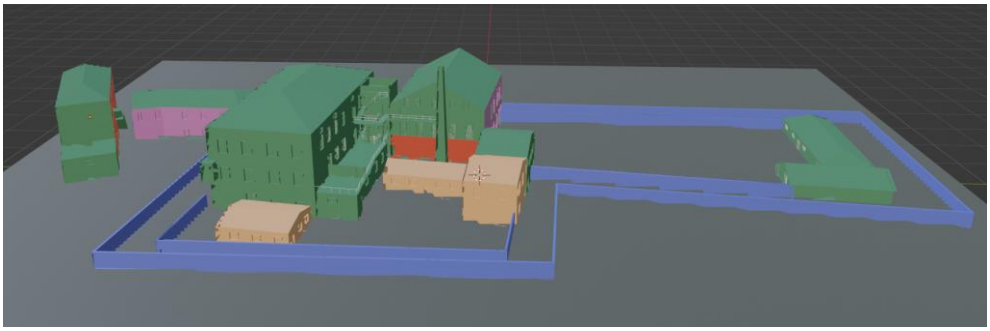


Ilustración 1. Resultados de fiabilidad.

Finalmente, se realizó un modelado 3D manual del penal, utilizando herramientas como Google Earth para obtener medidas aproximadas y diseñar

un plano general de los edificios. El texturizado del modelo se llevó a cabo mediante la técnica Image-Based Modeling and Rendering (IBMR).

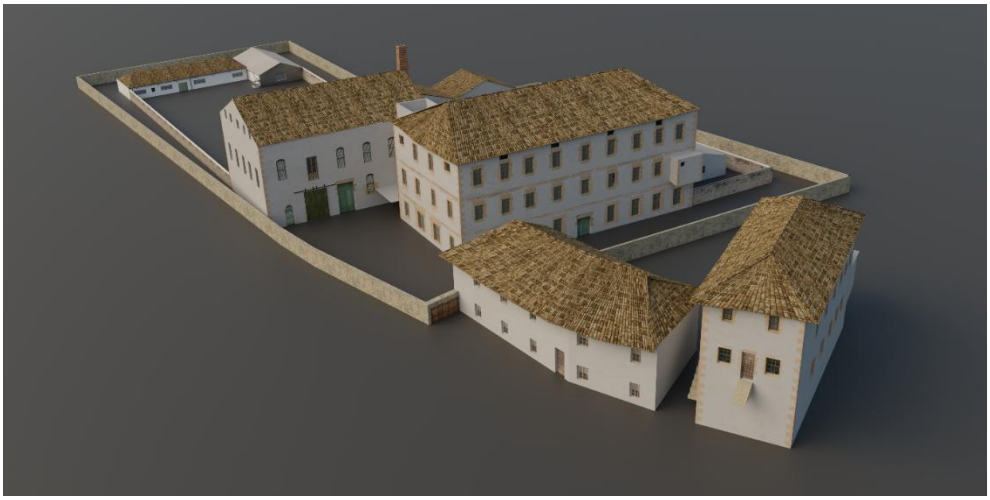


Ilustración 2. Resultado final.

Bibliografía

ALAGUERO RODRÍGUEZ, M. (2023), «Los niveles de detalle en la reconstrucción virtual de núcleos urbanos históricos: estudio del caso de Vitoria-Gasteiz en el siglo XII», recurso en línea.

AVINTIA (2022, 12 de diciembre), «Penal de Valdenoceda – Lista Roja», *Lista Roja del Patrimonio* (Hispania Nostra). https://listaroja.hispanianostra.org/ficha/penal-de-valdenoceda/

BENITO DEL POZO, P. (2012), «Herramientas para intervenir en el patrimonio: el Inventario de Patrimonio Industrial de Castilla y León», *Polígonos. Revista de Geografía*, 19, pp. 181–192. https://doi.org/10.18002/pol.v0i19.76

BOLETÍN OFICIAL DEL ESTADO (1938), [Conjunto de datos]. https://www.boe.es/datos/pdfs/BOE//1938/138/A02391-02391.pdf

CALDUCH CERVERA, J. (2009), «El declive de la arquitectura moderna: deterioro, obsolescencia, ruina», *Palapa*, IV(11), pp. 29–43.

CAZZARO, I. (2023), *Digital 3D Reconstruction as a Research Environment in Art and Architecture History: Uncertainty Classification and Visualisation* [Tesis doctoral], Alma Mater Studiorum – Università di Bologna. https://amsdottorato.unibo.it/id/eprint/10817/1/irene_cazzaro_digital_3d_reconstructions.pdf

DEMETRESCU, E., & FERDANI, D. (2021), «From field archaeology to virtual reconstruction: A five steps method using the extended matrix», *Applied Sciences*, 11(11), art. 5206. https://doi.org/10.3390/app11115206

RÍOS, L., MARTÍNEZ, B., GARCÍA-RUBIO, A., & JIMÉNEZ, J. (2008), «Muertes en cautiverio en el primer Franquismo: exhumación del cementerio del

penal de Valdenoceda (1938–1943)», *Complutum*, 19(2), pp. 139–160. https://www.researchgate.net/publication/277274749

YOFUIPRESADEFRANCO (2016, 12 de junio), «Los lugares del horror; el penal de Valdenoceda y la cárcel de Burgos», *Yo Fui Presa de Franco* (blog). https://yofuipresadefranco.wordpress.com/2016/05/24/los-lugares-del-horror-el-penal-de-valdenoceda-y-la-carcel-de-burgos/

Ni son todas las que están, ni están todas las que son: Estructuras, márgenes y mujeres artistas en el sistema expositivo andaluz

Bárbara Romero-Ferrón
Provenance Lab – Leuphana University;
iArtHisLab – Universidad de Málaga
Nuria Rodríguez-Ortega
iArtHis_Lab – Universidad de Málaga

Palabras clave: Exposiciones; análisis de redes; estudios de género; teorías de grafo; fenómenos culturales
Keywords: Exhibitions; network analysis; gender studies; graph theories; cultural phenomena

Introducción

Durante siglos, la historia del arte ha sido narrada desde una perspectiva que relegó a las mujeres a un papel secundario, convirtiendo su presencia en un eco intermitente dentro del relato institucional. Aunque en las últimas décadas museos e instituciones culturales han impulsado exposiciones monográficas y revisiones críticas — como *Invitadas* del Museo del Prado, la primera exposición dedicada por la institución, da las mujeres artistas — sigue siendo necesario preguntarse si las mujeres artistas han sido realmente integradas en las narrativas curatoriales o si continúan ocupando una periferia estructural y, en su caso, cómo se configura dicha periferia.

Las exposiciones temporales han sido el principal instrumento para subsanar omisiones históricas, pero también han dado lugar a nuevas formas de segregación simbólica. Exposiciones temáticas centradas en mujeres o artistas LGTBIQ+, aunque bienintencionadas, pueden aislar estas prácticas de los discursos centrales. Así, el problema no radica únicamente en cuántas mujeres son expuestas, sino en cómo se integran en el tejido del sistema artístico. Las exposiciones no son eventos aislados, sino nodos dentro de un ecosistema relacional: entre artistas, curadores, instituciones y discursos. Cada muestra es un acto de resignificación que puede abrir genealogías alternativas o reforzar lógicas dominantes.

Además de reflejar los intereses culturales del presente, las exposiciones temporales ofrecen un prisma a través del cual observar las transformaciones estructurales de la cultura contemporánea. Lejos de limitarse a un evento efímero, cada exposición implica decisiones curatoriales, institucionales y discursivas que revelan cómo se configuran las relaciones de poder, visibilidad y exclusión. En este sentido, el ecosistema expositivo es una estructura compleja cuya evolución está ligada a las lógicas de legitimación artística, las políticas de memoria y las tensiones entre centro y periferia. Desde el siglo XIX, las grandes exposiciones de bellas artes en España han mostrado una escasa representación de mujeres artistas, mientras que muchas de ellas se vieron forzadas a crear sus propios circuitos de exhibición, como asociaciones, salones independientes y espacios alternativos (Rodrigo Villena, 2017; Aguilera Sastre, 2011). Con el tiempo, este tipo de lógicas de funcionamiento periféricas no han desaparecido, sino que se han ido consolidando y evolucionando para cohabitar el ecosistema cultural.

En las últimas décadas, los enfoques curatoriales han comenzado a incorporar discursos de restitución y equidad, incluyendo perspectivas feministas, decoloniales y *queer*. Sin embargo, estos cambios no se producen de forma homogénea ni garantizan una transformación estructural real. Muchas veces, las medidas de inclusión se implementan de forma superficial o como estrategias simbólicas de reparación. Con todo, la pregunta central sigue vigente: ¿Estas iniciativas integran genuinamente a las mujeres artistas en el ecosistema cultural o simplemente les asignan espacios segregados dentro de él?

Desde el punto de vista metodológico, gran parte de los estudios sobre representación de género en el ámbito expositivo han adoptado un enfoque cuantitativo basado en recuentos: cuántas mujeres han sido expuestas frente a cuántos hombres[1], lo que se aviene bien con la *Ley Orgánica 3/2007* de Igualdad, cuyo artículo 26 insta a las instituciones públicas de cultura a promover la “presencia equilibrada de mujeres y hombres en la oferta artística y cultural” (Informes MAV #5, #12, #19, #21, #24; Cabré, *et al*., 2016; Nualart, 2019). Sin embargo, si bien estas cifras son necesarias, resultan insuficientes para captar las lógicas estructurales del sistema. Por ejemplo, estos análisis no exploran si las mujeres artistas logran mantener una presencia sostenida en el tiempo ni si acceden a los espacios de mayor centralidad institucional. Tampoco abordan la manera en que las conexiones entre artistas, curadores e instituciones afectan la inclusión o exclusión dentro del sistema (Rodríguez Ortega *et al.*, 2024).

[1] La Asociación de Mujeres en las Artes Visuales (https://mav.org.es/category/informes/) publica anualmente informes sobre la presencia de mujeres en exposiciones en España. Estos recuentos, aunque necesarios, calculan el porcentaje de artistas mujeres, sin embargo, no hacen una valoración comprensiva de la posición de dichas mujeres en el sistema expositivo y cultural español.

Corpus y metodología

Este estudio aborda el ecosistema expositivo en su condición de c red interconectada. En el marco de proyectos de investigación sucesivos (Exhibitium (2015-2016), Andalex (2017-2021), Complexhibit (2021-)) hemos construido un corpus que analiza 15.677 exposiciones celebradas en Andalucía entre 1990 y 2023, abarcando tanto capitales de provincia como pequeñas localidades[2]. Esta elección territorial permite ampliar el enfoque más allá de los grandes centros, como Sevilla o Málaga, que suelen concentrar los estudios, y explorar la dinámica cultural en espacios periféricos. En este estudio hemos optado por trabajar con el corpus de Andalucía debido a la forma sistemática y con un nivel de granularidad uniforme con la que se han recopilado los datos. En el caso concreto del subcorpus expositivo andaluz, el 75 % de las exposiciones están celebradas en las capitales de provincia y el 25 % en pequeñas ciudades y pueblos. Para cada exposición se extrajeron variables como: título, fechas, instituciones organizadoras, sedes, comisarios/as, artistas participantes y entidades colaboradoras.

En lo que respecta al atributo “género”, elemento central de nuestro análisis, la estructura de metadatos se basa actualmente en el binarismo tradicional (hombre, mujer), si bien incluye la categoría “otro” para los casos no binarios. La asignación de género se ha efectuado tomando como referencia el sexo biológico de los artistas, dado que aún no contamos con fuentes sistemáticas y fiables que permitan un abordaje alternativo. No obstante, entre los objetivos de desarrollo futuro se encuentra implementar mecanismos de corrección y ajuste, por ejemplo, la consulta directa con los artistas o la investigación en fuentes primarias y secundarias donde consten sus declaraciones personales. Dado que aún no disponemos de toda la información necesaria, en esta fase del proyecto (y para la presente presentación) hemos optado por centrarnos únicamente en la categoría “mujer”, entendida, como se ha señalado, en términos de sexo biológico.

Un estudio precedente abordó la presencia de las artistas mujeres en el ecosistema expositivo andaluz desde un enfoque puramente estadístico (Rodríguez Ortega, 2020). Sin embargo, entendiendo el sistema expositivo como un entramado complejo que opera bajo las lógicas de un sistema libre de escala[3], consideramos que este análisis requiere ser complementado con el examen de las dinámicas de la periferia a través de las teorías de grafos y el análisis de redes. Dado que nuestra propuesta metodológica se centra en los y las artistas,

[2] Estos datos han sido recogidos en varios proyectos del grupo de investigación iArtHisLab en la Universidad de Málaga, IP: Dr. Nuria Rodríguez Ortega. Desde 2015 a la actualidad hemos trabajado en el registro sistemático de exposiciones artísticas. Estos proyectos han sido financiados por varias instituciones públicas y privadas a nivel nacional y autonómico. Para más información sobre el proyecto actual: https://complexhibit.eu/

[3] Un sistema libre de escala presenta una distribución de grados que sigue una ley de potencias, con unos pocos nodos muy conectados (*hubs*) y muchos nodos con pocas conexiones.

optamos por construir redes en las que los nodos representan los principales actores del sistema: instituciones, comisarios y artistas. El análisis de esta red nos permite abordar tres dimensiones clave: la persistencia de las mujeres artistas en el sistema (es decir, su participación recurrente a lo largo del tiempo), su distribución entre distintos tipos de instituciones (grandes museos, centros independientes, espacios municipales, etc.) y su posición estructural medida mediante algoritmos de centralidad (como grado ponderado, cercanía, intermediación o PageRank) y la detección de comunidades. Este enfoque no solo permite observar qué artistas están presentes o su presencia porcentual en términos absolutos, sino también cómo circulan dentro del sistema y con quién establecen conexiones. Las artistas mejor conectadas no siempre son las más frecuentemente expuestas, y la centralidad en la red ofrece una medida alternativa del reconocimiento institucional.

Resultados preliminares

Los resultados muestran que el género no altera sustancialmente las lógicas estructurales del sistema expositivo. Aunque se incrementa la participación femenina en términos porcentuales, como evidencia el último informe MAV de 2024, que cuantifica una presencia porcentual del 40 %, la posición estructural de muchas artistas sigue siendo periférica. Los hombres presentan una centralidad de grado promedio mayor (0.00238 frente a 0.00196), con mayor variabilidad y valores extremos más altos. Esto indica que, en esta red, los artistas hombres tienden a estar más conectados que las mujeres.

Ahora bien, ¿con quién se conectan las mujeres? El 58 % de sus conexiones son con artistas hombres, frente al 42 % con otras mujeres (Fig.1). Aunque la diferencia no es abismal, sugiere que, cuando las artistas son incluidas en exposiciones, tienden a compartir espacio expositivo con hombres más que con otras mujeres. Sin embargo, este patrón se matiza: cuando analizamos a las mujeres más conectadas, observamos que el 55,06 % tiene más vínculos con hombres que con mujeres, lo que indica que, aunque persisten ciertas políticas expositivas de agrupamiento femenino, aquellas artistas con mayor centralidad siguen participando de redes dominadas por hombres.

Este patrón se refuerza al cruzarlo con el número de exposiciones: las artistas con mayor presencia expositiva también tienden a tener más conexiones con hombres. Es decir, la visibilidad y la centralidad siguen dependiendo de la integración en redes mixtas donde los vínculos con hombres predominan. Si observamos el género de los comisarios, los patrones se acentúan: en las exposiciones comisariadas por mujeres, el 30 % de los artistas son mujeres, mientras que en las organizadas por hombres, solo el 24 % de los artistas son mujeres. Cuando miramos el poder estructural de los comisarios en la red, aparece de nuevo la imperante lógica de género donde las artistas con más conexiones en la red son expuestas por comisarios.

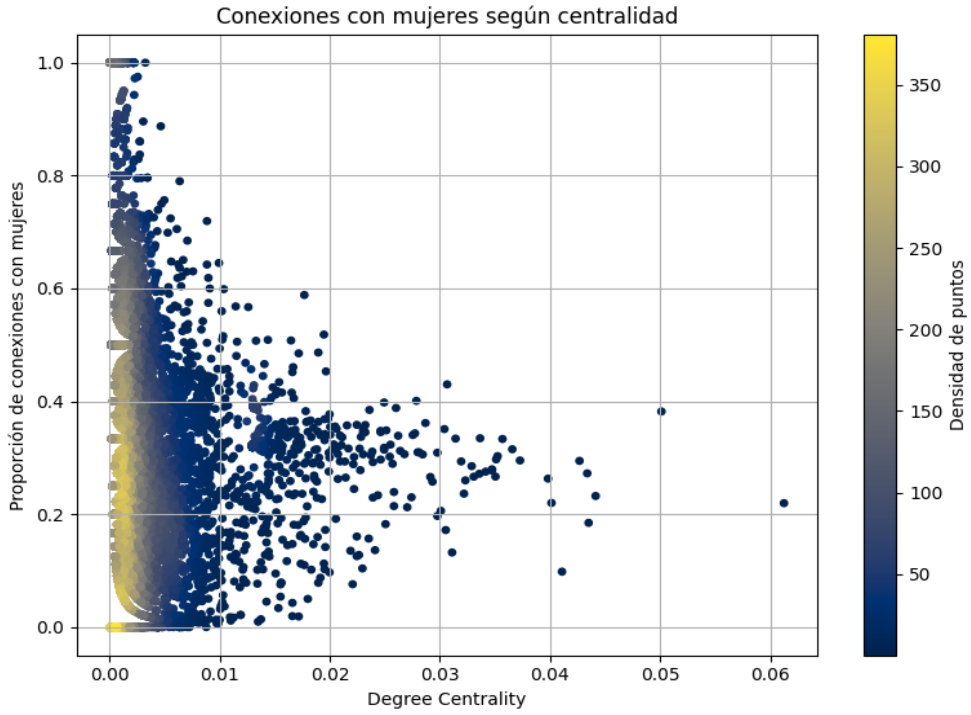


Figura 1. Relación entre la centralidad de grado y la proporción de conexiones con mujeres en la red.

Si eliminamos a los artistas hombres y analizamos solo las dinámicas entre mujeres, el patrón estructural se reproduce. Al aplicar un algoritmo de detección de comunidades, emerge una gran comunidad de más de 800 artistas (un núcleo que tiende a exponerse conjuntamente), mientras que el resto de nodos femeninos se dispersa en pequeños clústeres periféricos. Esto evidencia que, incluso dentro del sistema expositivo femenino, la lógica centro–periferia persiste.

Al entender el ecosistema expositivo como un conjunto de subsistemas que comparten dinámicas estructurales similares, y con el objetivo de analizar específicamente el subsistema de mujeres, se ha aplicado el algoritmo de k-core decomposition. Este método permite identificar jerárquicamente los subgrafos más densamente conectados de una red, asignando a cada nodo su núcleo estructural. En lugar de aplicarlo a la red completa, se ha focalizado en la red de mujeres para poder explorar y resaltar estructuras internas. El resultado revela una red claramente estratificada: un núcleo compacto y cohesionado de artistas que han sido expuestas recurrentemente juntas, frente a una mayoría que ocupa posiciones periféricas o escasamente conectadas (Fig.2). El núcleo estructural más denso está compuesto por 81 mujeres con un valor de k=80, lo que las sitúa como las artistas más integradas dentro del sistema expositivo femenino. Este análisis demuestra que, incluso dentro de la subred de mujeres artistas, se reproduce una lógica centro-periferia. Un grupo reducido altamente interconectado contrasta con una periferia amplia y dispersa, donde se concentra el 60 % de las artistas, todas ellas con un k=0.

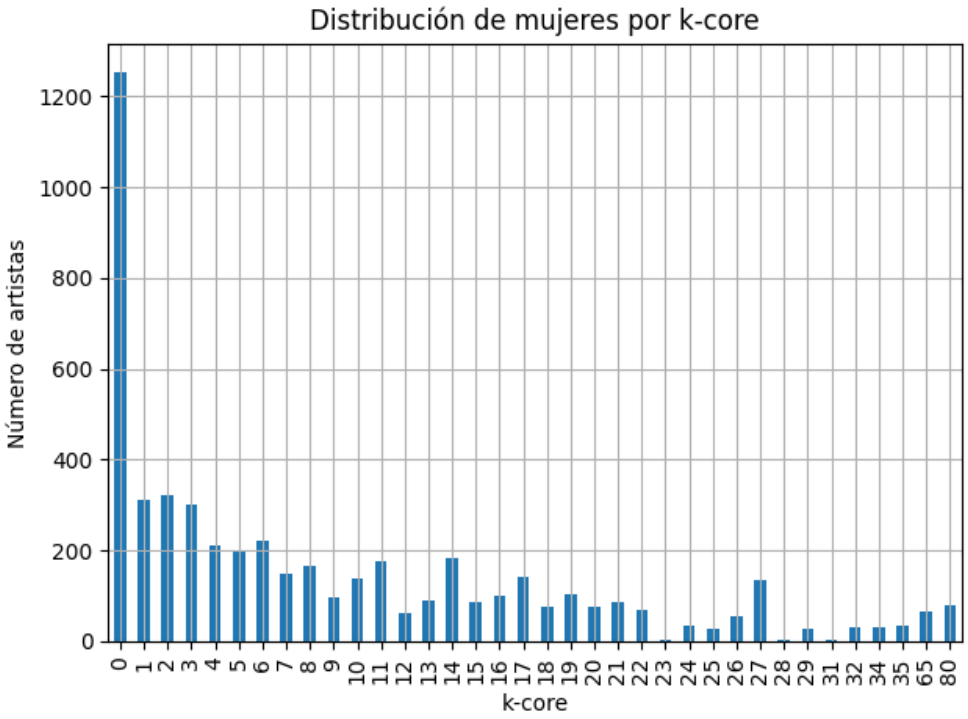


Figura 2. Distribución del número de artistas mujeres según su pertenencia a distintos niveles de k-core en la red.

Conclusiones

El análisis del ecosistema expositivo andaluz desde una perspectiva de red ha permitido visibilizar no solo la magnitud de la brecha de género, sino también algunas de las lógicas estructurales que la sostienen. A pesar del aumento cuantitativo en la participación de mujeres artistas, estas siguen ocupando posiciones periféricas en la arquitectura del sistema, tanto en términos de centralidad como de acceso a instituciones de prestigio. Las redes de coexposición y curaduría revelan que la integración de las mujeres no responde a una transformación estructural, sino a una inclusión parcial que sigue operando bajo lógicas jerárquicas. Los vínculos más frecuentes de las artistas siguen siendo con hombres, y las exposiciones comisariadas por varones continúan siendo las que más visibilidad otorgan a las mujeres mejor posicionadas en la red. Incluso dentro del propio subsistema femenino se reproduce una estructura centro-periferia, donde un reducido núcleo altamente conectado coexiste con una mayoría fragmentada y marginal.

Estos hallazgos permiten cuestionar las políticas culturales centradas exclusivamente en la inclusión numérica, como el citado artículo 26 de la *Ley Orgánica 3/2007* de Igualdad. Reequilibrar los porcentajes no basta si no se transforman las estructuras que sostienen la desigualdad. Comprender las redes que configuran el ecosistema expositivo es esencial para diseñar intervenciones que no solo aumenten la presencia porcentual de las mujeres artistas, sino que también redistribuyan el poder simbólico e institucional en el campo del arte. En

este sentido, las genealogías periféricas no deben entenderse como una anomalía, sino como una parte central del entramado cultural que merece ser visibilizada y fortalecida. Además, es importante hacer una reflexión crítica de cómo la periferia, al ser incluida en el sistema, reproduce jerarquías internas.

Bibliografía

AGUILERA SASTRE, J. (2021). “Las musas suben a la tribuna. Visibilidad y autoridad de las mujeres en el Ateneo de Madrid (1882-1939)”. Anales de la literatura española contemporánea, 46(1), 269-271.

CABRÉ, M. À.; BAYGUAL, J.; et al. (2016). Exposiciones de mujeres en los centros de arte de Barcelona (2011-2015). 4.º Informe Observatorio Cultural de Género – OCG (col. MAV).

MAV – Mujeres en las Artes Visuales (2011-2024). Informes MAV #5, #12, #19, #21. Madrid: MAV. (Disponibles en mav.org.es).

NUALART, Cristina. (2019). Exposiciones temporales del Museo Reina Sofía, 2000-2016, desde una perspectiva feminista (Tesis doctoral). Madrid: Universidad Complutense de Madrid.

RODRIGO VILLENA, I. R. (2018). “Escultoras en un mundo de hombres y su fortuna en la crítica de arte española (1900-1936)”. Arenal. Revista de historia de las mujeres, 25(1), 145-168.

RODRÍGUEZ ORTEGA, Nuria; CRUCES RODRÍGUEZ, Antonio; PINO DÍAZ, José. (2020). «Aportaciones al conocimiento de la presencia de las mujeres en el ecosistema expositivo de Andalucía (2008-2018)». Informes Andalex, 1 (junio).

RODRÍGUEZ ORTEGA, Nuria; et al. (2024). OntoExhibit: An Ontology for Modelling the Field of Art Exhibitions and Its Semantic-Discursive Dimension. En Digital Humanities 2024: Book of Abstracts.

RODRÍGUEZ-ORTEGA, N. (2024). “Contours of Knowledge: Epistemological Implications of Semantic Models in Art Exhibition Representation”. Život Umjetnosti, 114(1), 122-147.

RODRÍGUEZ ORTEGA, Nuria; ROMERO FERRÓN, Bárbara; ORTIZ TELLO, María. (2024). “Challenging Conventional Statistical Metrics for Gender Equity in Exhibitions Ecosystem”. ADHO, Washington University, agosto de 2024. [Actas de congreso].

De los átomos a las ondas: estilometría rítmica para el estudio de la autoría en el teatro del Siglo de Oro

Álvaro Cuéllar
Universitat Autònoma de Barcelona

Palabras clave: Humanidades digitales; Identificación autorial; Aprendizaje automático; Inteligencia artificial; Escansión automática
Keywords: Digital Humanities; Author Identification; Machine Learning; Artificial Intelligence; Automatic Scanning

Introducción

La atribución autorial en el teatro del Siglo de Oro se ha apoyado, con notable éxito, en la estilometría léxica: un enfoque que atomiza los textos en palabras, calcula sus frecuencias relativas y compara los vectores resultantes para inferir proximidades estilísticas. Ese procedimiento ofrece tasas de acierto muy elevadas—cercanas al cien por cien en conjuntos cuidadosamente editados—gracias, en buena medida, a la relativa homogeneidad formal de las comedias áureas (extensión similar, predominio del verso regular, léxico compartido, etc.). Sin embargo, al sustentarse por completo en unidades léxicas, el método deja al descubierto varios frentes potenciales de vulnerabilidad. La temática puede influir: una obra de asunto religioso o mitológico comparte necesariamente léxico especializado con otras de la misma materia, lo que puede enmascarar rasgos propiamente autoriales. Las colaboraciones, frecuentes en la escena barroca, combinan voces de distintos dramaturgos dentro del mismo texto, diluyendo la señal individual. Las modernizaciones ortográficas y las políticas editoriales que uniforman acentuaciones alteran la frecuencia de tokens, y los arcaísmos deliberados que ciertos autores introducen para lograr efectos estilísticos añaden otra capa de ruido. Estos factores no invalidan la estilometría léxica, pero recuerdan que su gran eficacia convive con zonas grises aún por calibrar.

Con el fin de explorar un eje menos expuesto al contenido léxico y más vinculado a la prosodia, proponemos un desplazamiento conceptual *de los átomos a las ondas*. Cada verso puede codificarse como una secuencia binaria de sílabas tónicas (+) y átonas (–), y el agregado de todas esas secuencias configura, a escala de obra, un perfil métrico que parece menos dependiente de la temática o de la modernización gráfica y, presumiblemente, más ligado a hábitos compositivos arraigados. El examen sistemático de dicho perfil—al que denominamos estilometría rítmica—no pretende reemplazar la señal léxica ni

ofrecer certezas absolutas, sino sumar una perspectiva complementaria que pueda reforzar, matizar o, en algunos casos concretos, iluminar las conclusiones derivadas del conteo de palabras. Así, sin desechar la potencia de los enfoques tradicionales, esta línea prosódica abre la posibilidad de contrastar resultados, detectar casos ambiguos y, en última instancia, enriquecer el repertorio metodológico disponible para los estudios de autoría en el teatro aurisecular.

Marco teórico y estado del arte

Diversos trabajos de la última década han puesto de relieve que la prosodia porta una firma autorial tan identificable como la léxica, y en ciertos contextos incluso más estable. En el ámbito hispánico, Kroll y Sanz-Lázaro han mostrado que basta medir la distribución relativa de unos pocos patrones octosilábicos —es decir, observar en qué puntos del verso tienden a concentrarse los acentos— para distinguir con notable precisión a Calderón de la Barca, Lope de Vega o Mira de Amescua, alcanzando márgenes de error muy reducidos. Estudios paralelos de Plecháč, centrados en poesía checa e inglesa, confirman que la métrica funciona como un indicador fiable cuando el vocabulario se degrada por traducción, plagio temático o reforzamiento dialéctico. Este salto cualitativo ha sido posible gracias a la madurez reciente de los algoritmos de escansión automática: herramientas como Rantanplán, Jumper o, de manera especial, libEscansión convierten millones de versos en secuencias binarias tónica-átona con exactitudes próximas al noventa y cinco por ciento, resolviendo de forma programática fenómenos complejos —sinalefas, diéresis, compensaciones hemistiquiales, variación ortográfica— y procesando grandes corpus en tiempos prácticamente lineales. Con todo, la mayoría de pruebas hasta la fecha se han limitado a conjuntos muy reducidos de romances o a comparaciones poéticas puntuales; faltaba aún un estudio de escala intermedia, suficientemente amplio para incluir cien comedias completas y, al mismo tiempo, lo bastante controlado para aplicar protocolos de validación exhaustivos y paralelos sobre características léxicas y rítmicas. Esa laguna metodológica —evaluar, bajo idénticas condiciones experimentales, la potencia relativa de ambos canales— es precisamente la que se propone solventar el presente trabajo.

Corpus y metodología

Se confeccionó un corpus balanceado de cien comedias íntegramente versificadas cuya atribución goza de consenso crítico, procedentes de ediciones académicas recientes que documentan variantes y garantizan una fijación textual homogénea; las obras se distribuyeron entre nueve firmas emblemáticas del teatro áureo —Lope de Vega, Calderón de la Barca, Tirso de Molina, Ruiz de Alarcón, Agustín de Moreto, Francisco de Rojas Zorrilla, Juan Pérez de Montalbán, Guillén de Castro y Miguel de Cervantes— procurando que cada dramaturgo aportara varios títulos y que la cronología cubriera las primeras décadas del seiscientos hasta mediados del siglo. A partir de los originales se

generaron dos canalizaciones paralelas. En la ruta léxica se aplicó una limpieza y homogeneización del corpus a unas mismas variantes. La vía rítmica procesó los cerca de trescientos mil versos resultantes mediante libEscansión, herramienta capaz de anotar tónicas, átonas y licencias (sinalefa, diéresis) con una exactitud verificada del 95 %; cada línea se tradujo a cadenas de signos «+» y «–». Para medir la capacidad discriminativa de ambos dominios se aplicó validación cruzada leave-one-out: en cada iteración se excluyó una comedia, se entrenó con las noventa y nueve restantes y se predijo la etiqueta del texto retenido. Este protocolo garantiza independencia entre entrenamiento y prueba aun con corpus reducido y permite estimar varianza autor-a-autor. Se confrontaron multitud de clasificadores, pero destacan especialmente Delta y SVM. El Delta de Burrows calcula distancias euclidianas sobre puntuaciones z y pondera las varianzas de cada rasgo, y SVM trabaja con máquinas de soporte vectorial lineales de alta dimensionalidad y multicolinealidad. Los resultados se sintetizaron con F_1 macro para la tarea multiclase —promediando la armonía entre precisión y exhaustividad en las nueve categorías— y con la misma métrica, además de en configuraciones binarias «autor frente al resto», todo ello con el fin de explorar tanto la capacidad de atribución global como el rendimiento específico cuando se formula la pregunta “¿es o no de este dramaturgo?”.

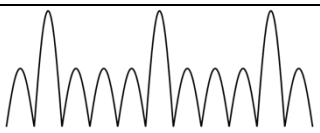
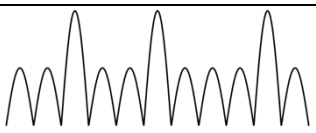
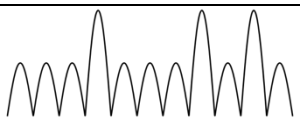
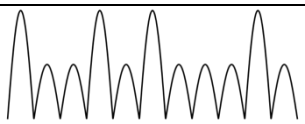
Verso	Onda prosódica	Patrón rítmico
ahora que descansan los aceros		-+---+---+-
de mis nunca olvidados compañeros		--+--+---+-
y el enemigo de furor armado		---+---+-+-
huye vencido o duerme fatigado		+--+-+---+-

Tabla 1. Escansión automática de cuatro versos del teatro del Siglo de Oro

Resultados y discusión

Los experimentos confirman la supremacía del canal léxico en clasificación multiautoral: por encima de quinientas palabras frecuentes, el F_1 supera 0,99 y se mantiene invariante. La señal rítmica, en cambio, alcanza su cenit en torno a ochenta patrones más frecuentes con un F_1 aproximado de 0,72 y desciende cuando se añaden ritmos menos productivos, lo que revela que la riqueza métrica introduce ruido más rápido de lo que añade información. No obstante, en clasificaciones binarias surge el valor diferencial de la métrica: Lope de Vega y

Moreto obtienen resultados superiores a 0,99 con tan sólo setenta ritmos, mientras Calderón se sitúa en 0,98; incluso autores de firma difusa, como Montalbán o Rojas, se acercan al 0,80, margen aceptable dado el reducido vector empleado. Estos resultados avalan que la métrica es extremadamente informativa para algunos dramaturgos y que sirve como test independiente allí donde el léxico se degrada por temática o edición. Además, el análisis de rangos confirma una ley de Zipf métrica: veinte ritmos acumulan más de la mitad de todas las ocurrencias, lo que justifica la efectividad de vectores pequeños y explica por qué errores de escansión aleatorios se diluyen en el cómputo global.

Limitaciones y líneas futuras

Tres factores restringen el alcance de nuestras conclusiones. Primero, la fiabilidad de la escansión depende de las ediciones escogidas; diferencias de sinalefa o de acentuación diatópica pueden sesgar las frecuencias. Segundo, el corpus de cien obras, aunque variado, no representa la enorme diversidad de las tres mil comedias conservadas; será indispensable paralelizar la escansión masiva mediante clusters y silabeadores optimizados para validar la robustez del método a gran escala. Tercero, la interpretabilidad permanece limitada: aunque los SVM lineales ofrecen márgenes claros, aún necesitamos técnicas de explicación para vincular estadística y filología métrica. A corto plazo proponemos ensayar modelos híbridos que concatenen vectores léxicos y rítmicos, explorar transformadas de Fourier sobre las ondas prosódicas para capturar periodicidades globales y generar mapas que combinen ambas distancias para visualizar macrotendencias en todo el teatro áureo. A medio plazo, la aplicación forense del método en atribuciones disputadas permitirá calibrar su potencia probatoria frente a tribunales filológicos.

Conclusión

La estilometría rítmica no aspira a destronar la señal léxica, pero se revela como un complemento valioso, especialmente en configuraciones binarias o cuando el léxico se distorsiona por razones extrínsecas. Con un núcleo de sesenta a cien patrones —en su mayoría octosilábicos— es posible distinguir con casi pleno acierto a ciertos autores, y aun cuando la métrica falla en multiclase, ofrece indicios que la léxica ignora. La prosodia, por tanto, guarda una impronta autorial estable y menos dependiente del contenido, lo que abre una vía fértil para investigaciones sobre datación, género dramático o dialectología interna del Siglo de Oro. El tránsito de *átomos* a *ondas* demuestra que la métrica, leída con herramientas digitales, todavía resuena con información inédita sobre la identidad de quienes la compusieron y ofrece a las Humanidades Digitales un terreno promisorio para futuras exploraciones.

Bibliografía

DE LA ROSA, J., PÉREZ, Á., HERNÁNDEZ, L., ROS, S., & GONZÁLEZ-BLANCO GARCÍA, E. (2020), «Rantanplan, fast and accurate syllabification and scansion of Spanish poetry», *Procesamiento del Lenguaje Natural*, 65, pp. 83–90. https://doi.org/10.26342/2020-65-10

KROLL, S., & 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), pp. 1–18. https://doi.org/10.5944/rhd.vol.7.2022.31620

KROLL, S., & SANZ-LÁZARO, F. (2023), «Ritmo, autoría y género: Nuevas perspectivas sobre teatro lopesco desde las humanidades digitales», *Anuario Lope de Vega. Texto, Literatura, Cultura*, 29, pp. 351–375. https://doi.org/10.5565/rev/anuariolopedevega.491

MARCO REMÓN, G., DE LA ROSA, J., GONZALO, J., ROS, S., & GONZÁLEZ-BLANCO, E. (2021), «Automated metric analysis of Spanish poetry: Two complementary approaches», *IEEE Access*, 9, pp. 51734–51746. https://doi.org/10.1109/ACCESS.2021.3069635

MARCO REMÓN, G., & GONZALO ARROYO, J. (2021), «Escansión automática de poesía española sin silabación», *Procesamiento del Lenguaje Natural*, 66, pp. 77–87. https://doi.org/10.26342/2021-66-6

NAVARRO-COLORADO, B. (2017), «A metrical scansion system for fixed-metre Spanish poetry», *Digital Scholarship in the Humanities*, 33(1), pp. 112–127. https://doi.org/10.1093/llc/fqx009

PLECHÁČ, P. (2021), *Versification and Authorship Attribution*, Prague, Institute of Czech Literature of the Czech Academy of Sciences / Karolinum Press. https://doi.org/10.14712/9788024648903

PLECHÁČ, P., BOBENHAUSEN, K., & HAMMERICH, B. (2018), «Versification and authorship attribution: A pilot study on Czech, German, Spanish, and English poetry», *Studia Metrica et Poetica*, 5(2), pp. 29–54. https://doi.org/10.12697/smp.2018.5.2.02

SANZ-LÁZARO, F. (2023), *Ritmo y estructura de la comedia áurea: Posibilidades y límites del análisis digital automático del teatro español clásico* [Tesis doctoral], Universität Wien. https://doi.org/10.25365/thesis.75102

SANZ-LÁZARO, F. (2024), «libEscansión: A recursive precedence approach to metrical scansion», *Digital Humanities Quarterly*, 18(3). http://www.digitalhumanities.org/dhq/vol/18/3/000739/000739.html

Retrieval Augmented Generation como herramienta para el estudio de la censura teatral de los Siglos de Oro

Joel Reverter Lainez
Universitat Autònoma de Barcelona – PROLOPE

Palabras clave: Siglos de Oro; teatro; censura; inteligencia artificial; RAG
Keywords: Golden Age; theater; censorship; artificial intelligence; RAG

De entre las múltiples facetas del teatro aurisecular, la censura previa a la representación es una de las que más ha atraído la atención de los investigadores. Se trata de una línea de investigación imprescindible tanto para conocer el circuito teatral de los Siglos de Oro como para poder editar fielmente los textos dramáticos de la época. Buena prueba de ello son los hallazgos que se han realizado, por ejemplo, en *El castigo sin venganza*, en cuyo manuscrito Boadas (2024) ha detectado la intervención de Pedro de Vargas Machuca. Estos importantes avances no habrían sido posibles sin la ayuda de técnicas de las humanidades digitales: la investigadora pudo detectar la mano del censor gracias al uso de técnicas de fotografía y espectroscopia, que ya se habían descrito anteriormente (2020). Esto manifiesta la necesidad de servirnos de las últimas tecnologías para intentar desenredar una maraña tan compleja como la que presenta el sistema de censura de los Siglos de Oro.

No es fácil estudiarlo, ya que las generalizaciones tienden a simplificar en exceso un procedimiento en el que hay una multitud de personas implicadas y que se da en una amplia variedad de contextos. Por tanto, cada obra merece un estudio detallado y exhaustivo que parta de los presupuestos generales o más comunes y los contraste con las evidencias de las que disponemos. Por suerte, disponemos de la base de datos CLEMIT, especializada en la censura teatral de los Siglos de Oro, en la que hay a disposición de los investigadores una ficha con los datos más relevantes de cada obra estudiada. Pese a ser un excelente punto de partida, necesitamos más herramientas para estudiar con profundidad la censura.

A menudo se recurre a argumentos conceptuales para justificar la presencia o ausencia de actividad censora en una determinada obra. Véase como ejemplo la ficha de *La cristianísima lis* disponible en el CLEMIT, en la que Sánchez Jiménez expone aquellos fragmentos que, a juzgar por su contenido, pueden ser susceptibles de haber sido censurados. Sin duda, se trata de un primer indicio importante, aunque no definitivo, para poder detectar la actuación censora sobre un manuscrito, o incluso sobre un impreso. Con todo, no es

sencillo delimitar el contenido atajado: cuanto más estudiamos la censura, más ejemplos descubrimos, y más se expanden las posibilidades. En el libro de Urzáiz Tortajada, por ejemplo, se exponen ejemplos de censura religiosa, moral y política, pero también casos en los que el censor amplía sus funciones y ejerce de corrector de estilo, retoca aspectos de puesta en escena o incluso llega a ordenar el cambio de nombres de personajes por estar relacionados con la vida privada del dramaturgo (2023).

Es por esto que hemos intentado construir un sistema de *Retrieved Augmented Generation* (de aquí en adelante RAG) que permita detectar coincidencias temáticas entre las censuras ya detectadas y las que se están estudiando. El RAG es un método de búsqueda de texto cuyo motor es un modelo de inteligencia artificial. Para que funcione, el usuario debe cargar los documentos sobre los cuales quiere hacer la consulta para que la IA los procese y, después, formula una pregunta a la IA del mismo modo que se hace con los chatbots. Es, en definitiva, un método que permite la especialización de un *large language model* (LLM) sin la necesidad de entrenarlo mediante el *fine-tuning*. Aunque existen plataformas como SciSpace o NotebookLM que ofrecen servicios de RAG, estos no son opciones que sirvan para la investigación. Por un lado, son plataformas con una carga limitada de documentos, incluso en las modalidades de pago. Por otro lado, y todavía más importante, son servicios ya configurados en los que el usuario no puede adaptar a sus necesidades el RAG. Es por esto que creemos necesario diseñar un servicio alternativo que se adapte exactamente a nuestras necesidades.

La configuración afecta críticamente al rendimiento de la aplicación RAG, ya que estos parámetros son los que dictan cómo procesa el modelo de IA los documentos que le proporcionamos. Primero, debemos asegurarnos de que los archivos que proporcionemos a la IA son PDF con texto legible, es decir, con un tratamiento OCR. En el RAG, el LLM no lee todos los documentos a la vez, dado que esto ocuparía una gran cantidad de espacio en la memoria caché y se reduciría drásticamente el número de archivos que sería capaz de leer simultáneamente. Es por eso que previamente los PDF se segmentan en trozos más pequeños, en un proceso conocido como *chunking*. Aquí aparece la primera variable, ya que existen varios tipos de segmentación. La más común es determinar un número de tokens y añadir solapamiento para dar más contexto: por ejemplo, se puede segmentar el texto cada 500 tokens con 200 tokens de solapamiento, lo cual daría un segmento de unos 700 tokens en total. Los investigadores de ChromaDB, una de las bases de datos vectoriales más utilizadas y con mejor rendimiento del mercado, estudiaron a fondo los distintos métodos de segmentación de texto para RAG que existen, y publicaron los resultados en un artículo gratuito en línea cuyos resultados son reveladores (2024). En él, demuestran mediante diversas métricas que se obtienen mejores resultados mediante la segmentación semántica; es decir, si se divide el texto atendiendo a su significado mediante una biblioteca de Python especializada para ello. Por tanto, este es el método de segmentación que hemos utilizado.

Posteriormente, este segmento de texto se representa vectorialmente y se almacena dentro de una base de datos vectorial. Este proceso se conoce como *embedding*, y es otro de los parámetros que deben configurarse correctamente. Existen múltiples modelos de *embedding open source* que están disponibles en repositorios como Hugging Face, lo cual permite instalarlos y ejecutarlos localmente. Hemos experimentado con los modelos con mejor relación potencia-rendimiento: all-mpnet-base-v2, all-roberta-large-v1, sentence-transformers/all-distilroberta-v1 y sentence-transformers/all-MiniLM-L6-v2. Cabe apuntar que no se ha podido trabajar con modelos más potentes como Linq-AI-Research/Linq-Embed-Mistral, Alibaba-NLP/gte-Qwen2-7B-instruct, Salesforce/SFR-Embedding-Mistral o infly/inf-retriever-v1, dado que requieren mucha potencia computacional y, aunque se dispusiera parcialmente de ella, el rendimiento sería demasiado lento; son modelos pensados, pues, para ejecutarse en sistemas informáticos profesionales. Es por las limitaciones técnicas que se ha abandonado el entorno local y hemos probado un modelo de *embedding* que se ejecuta en un servidor ajeno a través de una API. Generalmente, estas son de pago, pero la API de Gemini, el LLM de Google, ofrece una versión gratuita limitada pero suficiente para experimentar.

Otro parámetro que afecta directamente al resultado es la recuperación de los *embeddings*. Este corre a cargo de la base de datos vectorial que se utilice. En nuestro caso, hemos optado por ChromaDB por ser la opción *open source* que obtiene mejores resultados en varias métricas de evaluación. Por último, se ha de tener en cuenta que la pregunta del usuario también es relevante a la hora de obtener una respuesta. Con todo, este debería ser el parámetro que, al menos idealmente, afectara en menor medida a los resultados, ya que la aplicación debería ser capaz de sintetizar la pregunta correctamente, independientemente de cómo esté escrita. La pregunta del usuario se procesa también mediante *embeddings* para poder confrontarla con los segmentos de los documentos y recuperar aquellos más relevantes. El número de documentos recuperados es otro parámetro esencial a la hora de obtener resultados.

Nuestra aplicación carga todos los documentos PDF disponibles en el CLEMIT, los procesa y contesta preguntas basándose únicamente en la información que estos contienen. También se ha creado un banco de preguntas contestadas manualmente y se han comparado con los resultados generados por la IA. Adicionalmente, se han analizado los resultados con una biblioteca de métricas de RAG. Aunque la utilidad de estas métricas es relativa porque no hay todavía un estándar de medición para el RAG, sí sirve para saber si las respuestas son correctas o no.

Los resultados han sido esclarecedores. En cuanto a los modelos de *embeddings*, los mejores modelos locales han sido all-mpnet-base-v2 y all-roberta-large-v1, aunque en ambos casos su rendimiento es significativamente peor que el ofrecido por el modelo de Google. El número de documentos recuperados ha reportado mejores resultados cuando se ha establecido entre 20 y 50; si el número es menor, falta contexto, y si es mayor, hay demasiado, por lo

que la precisión disminuye. Los dos modelos de IA que se han probado son las versiones de 14b y 32b de Deepseek, que se ha ejecutado en local a través de Ollama. El que ha dado mejor rendimiento es la versión de 32b, aunque la velocidad de escritura es muy lenta. Hay modelos mucho mejores, pero son de pago y no hemos podido acceder a ellos. Por tanto, la mejor configuración que hemos probado para ejecutar RAG en local de forma gratuita es con el segmentador semántico, generador de *embeddings* de Gemini, base de datos vectorial de ChromaDB, número de documentos recuperados de entre 20 y 50 y el modelo LLM deepseek-r1:32b.

Las conclusiones distan mucho de ser ideales. Incluso con la mejor configuración, las respuestas que ofrece la IA son, en la mayoría de casos, erróneas y, por tanto, evidencian la necesidad de mejorar el sistema para tareas más complejas como el estudio de la censura. Aunque es verdad que hay margen de mejora utilizando servicios de pago, implementando otras funcionalidades como la búsqueda adicional por metadatos o utilizando modelos de *embeddings* especializados para el español del Siglo de Oro, creemos haber topado con una limitación intrínseca al sistema RAG que lo descarta para responder a preguntas de investigación exigentes como las que aquí planteamos: la memoria. Como se ha demostrado con nuestros experimentos, la capacidad del RAG en cuanto a recuperación de segmentos es limitada; por tanto, procesos como la identificación conceptual entre múltiples segmentos rebasan sus posibilidades. Es por ello que pensamos que una potencial línea de investigación puede ser la implementación de memoria a largo plazo en las búsquedas, de modo que el sistema recuerde búsquedas anteriores y pueda mejorarlas progresivamente.

Bibliografía

BOADAS, S. (2020), «Techniques and instruments for studying the autograph manuscripts of Lope de Vega», *Hipogrifo: Revista de literatura y cultura del Siglo de Oro*, 8(2), pp. 509–531. https://doi.org/10.13035/h.2020.08.02.30

BOADAS, S. (2024), «Lope de Vega, Pedro de Vargas Machuca y el final de *El castigo sin venganza*», *Rassegna Iberistica*, 122, pp. 271–291. https://doi.org/10.30687/Ri/2037-6588/2024/23/001

CHROMA RESEARCH (s. f.), «Evaluating Chunking Strategies for Retrieval», recurso en línea. https://research.trychroma.com/evaluating-chunking [Acceso: 06/05/2025].

SÁNCHEZ JIMÉNEZ, R. (s. f.), «La cristianísima *lis*, y azote de herejía», en H. URZÁIZ TORTAJADA (ed.), *Censuras y licencias en manuscritos e impresos teatrales*, CLEMIT, recurso en línea. https://clemit.uv.es/gestion/ficheros/346-1638907109.pdf

URZÁIZ TORTAJADA, H. (2023), *La censura teatral en el Siglo de Oro*, Universidad de Valladolid, recurso en línea. https://uvadoc.uva.es/handle/10324/66059

URZÁIZ TORTAJADA, H., et al. (s. f.), *Censuras y licencias en manuscritos e impresos teatrales*, CLEMIT, recurso en línea. https://clemit.uv.es/

Diseño e integración de dispositivo IoCT en aplicación de Realidad Virtual

Guillem Montalbán Faet
Jaume Segura Garcia
Mario Alberto Montagut Climent
Universitat de València

Palabras clave: IoT; Realidad Virtual; Humanidades Digitales; Patrimonio Cultural; Conservación
Keywords: IoT; Virtual Reality; Digital Humanities; Cultural Heritage; Conservation

En el presente trabajo se presenta el desarrollo de un sistema que integra tecnologías de Internet de las Cosas (IoT) y Realidad Virtual (VR) en el contexto del patrimonio cultural digital. El proyecto consiste en un modelo virtual navegable del Palacio Real de València, al que se vinculan, en tiempo real, datos ambientales recogidos por un dispositivo IoT diseñado por el equipo. Esta integración sitúa el proyecto en el marco del Internet de las Cosas Culturales (IoCT), empleando los datos no solo para monitoreo, sino como variables activas que modifican la experiencia inmersiva.

En los últimos años, el IoT ha emergido como una tecnología clave en la transformación digital de múltiples sectores, incluyendo el ámbito del patrimonio cultural. La Figura 1 muestra un esquema general. Esta tendencia ha sido impulsada por la necesidad de preservar, monitorear y gestionar de manera más eficiente los bienes culturales tanto tangibles como intangibles. Mediante el uso de sensores, redes inteligentes y plataformas conectadas, el IoT permite recolectar datos en tiempo real sobre variables ambientales, estructurales y de comportamiento en museos, monumentos, sitios arqueológicos y colecciones históricas. Esto ha dado lugar al desarrollo de enfoques innovadores como los gemelos digitales (Digital Twins), que combinan modelos virtuales con datos del mundo físico para mejorar la toma de decisiones en conservación preventiva, restauración y gestión del patrimonio.

De esta manera, Laohaviraphap y Waroonkun (2024) han destacado, a través de una revisión sistemática de la literatura, cómo la integración del IoT y la inteligencia artificial permite una monitorización ambiental proactiva y estrategias de conservación adaptativas en entornos patrimoniales, subrayando además la necesidad de soluciones tecnológicas escalables y sostenibles para abordar los riesgos emergentes derivados del cambio climático y la contaminación atmosférica.

En este contexto, el marco propuesto por Darwish y Hassanien (2022) -IoHCT- plantea una integración entre gemelos digitales, redes de sensores, BIM (Building Information Modelling) y blockchain, con un enfoque centrado en la conservación estructural y la salud del patrimonio. Este enfoque coincide parcialmente con el presente proyecto, aunque aquí los datos IoT se utilizan principalmente para fines inmersivos y educativos. Nótese que en este estudio se refieren al término como IoHCT (Internet of Heritage Culture Things), aunque generalmente se utiliza IoCT.

Por su parte, Huang, Zhang y Xue (2022) exploran la integración del IoT con plataformas de realidad virtual en la nube, enfocándose en desafíos técnicos de escalabilidad y rendimiento. Su propuesta inspira nuestro trabajo en cuanto al uso de datos ambientales en tiempo real, aunque con un enfoque más técnico y generalista.

En el trabajo de Zhang et al. (2025) se ofrece una revisión sobre la convergencia entre IoT y tecnologías eXtended Reality (XR) en contextos museográficos, destacando experiencias personalizadas y multisensoriales. A diferencia de ese enfoque más institucional, nuestro proyecto apuesta por una exploración libre e interactiva del entorno digital, centrada en la apropiación cultural activa y la conexión sensorial con el espacio patrimonial mediante elementos de sincronía entre el estado del mundo real y el de la aplicación.

Como muestra la Figura 2, el dispositivo IoCT diseñado y ensamblado por nuestro equipo está construido sobre una placa NodeMCU con microcontrolador ESP32, que proporciona conectividad inalámbrica y capacidad de procesamiento para integrar múltiples sensores. En concreto, se han empleado el SHT31 para medir temperatura y humedad, el TEMPT6000 como sensor de luminosidad, y el MH-Z19C para la detección de niveles de dióxido de carbono (CO_2), cuyas conexiones con la placa microcontroladora están descritas en las Tablas 1, 2 y 3. La combinación de estos sensores permite obtener un perfil ambiental detallado del entorno físico —en este caso, los Jardines del Real de València— cuyos datos son transmitidos y sincronizados dinámicamente con el entorno de realidad virtual. Así, se consigue enriquecer la experiencia inmersiva y vincularla directamente con condiciones reales y cambiantes del espacio patrimonial.

Durante la ejecución de la aplicación de realidad virtual, los datos ambientales recogidos por el nodo IoCT se integran en tiempo real dentro del entorno digital. Los valores de las variables ambientales se presentan al usuario mediante un panel informativo situado discretamente en el campo visual, con el objetivo de contextualizar la experiencia sin interferir con la navegación. Estos datos que actúan como indicadores pasivos pueden emplearse para generar alertas o desarrollar estrategias de conservación preventiva en futuros desarrollos. Por otro lado, el dato proveniente del sensor de luminosidad (TEMPT6000) se utiliza, también, de forma activa. La intensidad luminosa modifica el estado del sistema, modulando la iluminación del entorno virtual, permitiendo que la escena se adapte parcialmente a las condiciones lumínicas reales del espacio físico donde se encuentra instalado el dispositivo. Esta

estrategia refuerza el vínculo entre el mundo físico y el entorno simulado, incrementando la inmersión y la coherencia sensorial de la experiencia.

El funcionamiento de la aplicación se basa en la comunicación entre el nodo IoCT y la plataforma en la nube ThingSpeak, que actúa como servidor intermedio para el almacenamiento y consulta de los datos recogidos. La Figura 3 muestra el esquema de obtención de datos del dispositivo IoCT, que periódicamente -cada minuto- envía los valores capturados por los sensores (temperatura, humedad, CO_2 y luminosidad) a la nube. Por su parte, la aplicación contiene un script que, a una frecuencia también minutal, realiza consultas a la API de ThingSpeak para obtener la última entrada registrada, como se ve en la Figura 4. Estos datos son luego interpretados y aplicados dinámicamente en el entorno virtual. Este flujo continuo permite mantener una sincronización en tiempo casi real entre las condiciones del entorno físico y la representación digital en la aplicación.

Para el desarrollo y validación de la aplicación de realidad virtual se ha utilizado Unity, que permite gestionar tanto la lógica interactiva como la visualización inmersiva. El entorno principal está compuesto por el modelo 3D del Palacio Real de València, elaborado íntegramente por el equipo a partir de documentación histórica y referencias arqueológicas. Este modelo navegable en primera persona constituye la base visual sobre la que se integran dinámicamente los datos IoT, generando una experiencia que combina fidelidad histórica, interactividad y adaptación contextual. Además, como se ve en la Figura 5, la aplicación ofrece al usuario la posibilidad de descargar el histórico completo de datos recogidos en formato CSV, permitiendo así su análisis posterior o su integración en otros sistemas de visualización o documentación.

Como ejemplo de aplicación, la Figura 6 muestra las medias horarias de los valores registrados por el dispositivo IoCT el 31 de julio de 2024 en los Jardines de Viveros. Se observa un pico de temperatura entre las 11:00 y las 14:00 horas, identificando un intervalo crítico para el mantenimiento preventivo. La combinación de temperatura elevada y alta humedad podría activar alertas automáticas, permitiendo una intervención preventiva por parte del personal responsable. Este enfoque demuestra el potencial del sistema para apoyar estrategias de conservación basadas en condiciones ambientales en tiempo real.

En términos de escalabilidad, el diseño del sistema permite la implementación simultánea de múltiples nodos IoCT en distintas zonas, sin interferencias, gracias al uso de canales de comunicación independientes. Cada dispositivo opera en su propio canal, lo que garantiza un rendimiento estable y evita la saturación del sistema. Esta arquitectura facilita la generación de alarmas específicas por zona, lo que resulta especialmente útil en escenarios de monitoreo distribuido.

En suma, la principal contribución de este proyecto se sitúa en su enfoque innovador hacia la conservación preventiva del patrimonio e integración de datos IoT como insumos activos dentro de una experiencia virtual. En particular, se incorpora la luminosidad ambiental como dato que modifica parcialmente la

iluminación del entorno simulado, reforzando la coherencia sensorial entre el espacio físico y el virtual. Más allá de su función informativa, estos datos permiten establecer vínculos en tiempo real entre las condiciones ambientales del entorno físico y su representación digital, habilitando mecanismos de alerta temprana y visualización contextualizada.

Como línea futura de trabajo, se plantea la incorporación de elementos adicionales que enriquezcan aún más la experiencia inmersiva, como la sincronización con la posición del sol, el ciclo de día y noche, y el nivel de nubosidad real mediante servicios meteorológicos conectados. Estas mejoras permitirán reflejar en el entorno virtual las condiciones atmosféricas cambiantes con mayor precisión, fortaleciendo la coherencia entre el mundo físico y digital, y posicionando el proyecto como una referencia en la generación de entornos patrimoniales inteligentes y dinámicos.

Figuras, tablas, mapas o gráficas

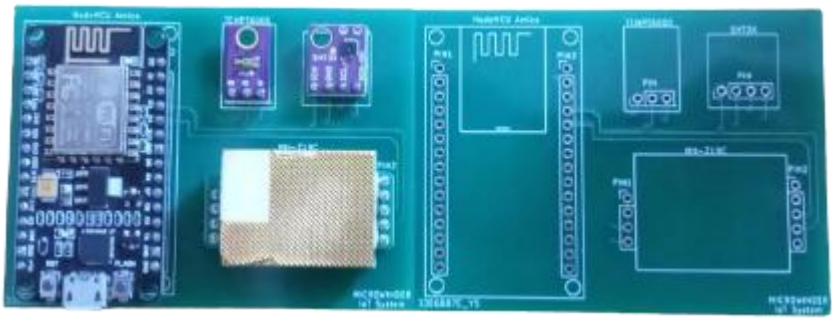


Figura 1. Dispositivo IoCT

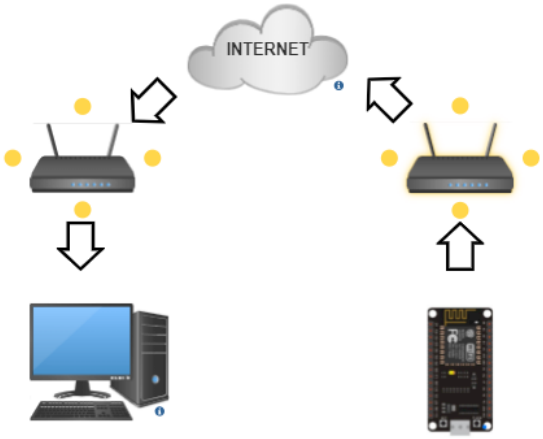


Figura 2. Diagrama de funcionamiento del dispositivo IoCT

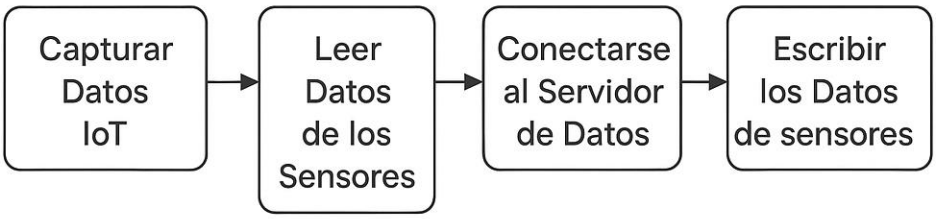


Figura 3. Algoritmo de escritura de datos en servidor en la nube

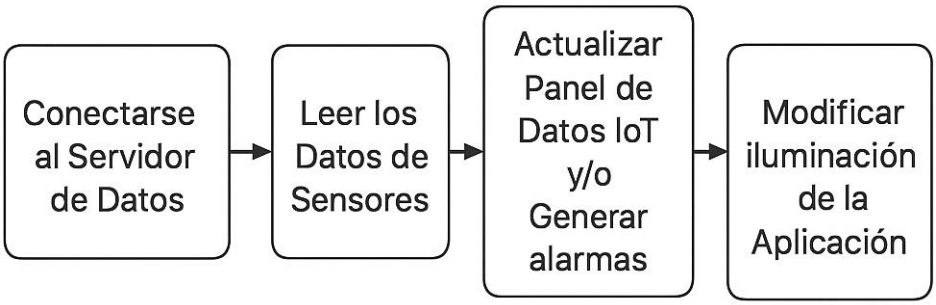


Figura 4. Algoritmo de lectura de datos en la aplicación

Figura 5. Aplicación con panel informativo y botón de descarga de datos IoCT

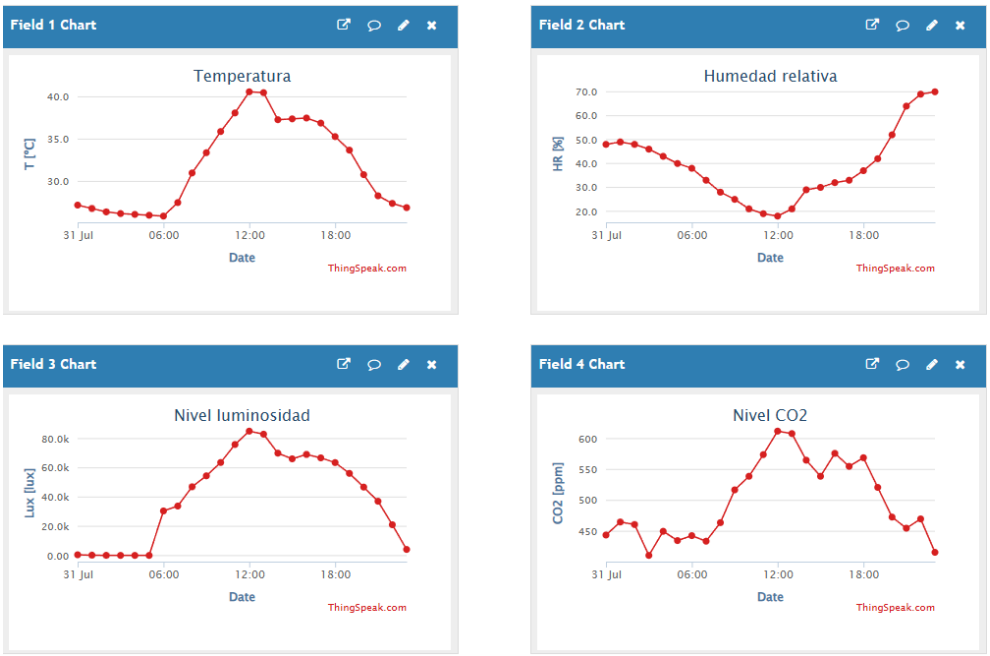


Figura 6. Medias horarias (T, RH, Lux y CO2) medidas con el dispositivo IoCT

SHT31	NodeMCU
SDA	D2
SCL	D1
Vin	3V3
GND	GNS

Tabla 1. Esquema de conexión de pines de sensor SHT31 y el microcontrolador NodeMCU

MHZ19C	NodeMCU
Tx	D4
Rx	D3
Vin	5V
GND	GND

Tabla 2. Esquema de conexión de pines de sensor MHZ19C y el microcontrolador NodeMCU

TEMPT6000	NodeMCU
Out	A0
Vin	3V3
GND	GND

Tabla 3. Esquema de conexión de pines de sensor TEMPT6000 y el microcontrolador NodeMCU

Bibliografía

DARWISH, A., & HASSANIEN, A. E. (2022), «IoHCT: Internet of Cultural Heritage Things digital twins for conservation and health monitoring of cultural in the age of digital transformation», en A. E. HASSANIEN, A. DARWISH, & V. SNASEL (eds.), *Digital Twins for Digital Transformation: Innovation in Industry*, Springer, pp. 1–20. https://doi.org/10.1007/978-3-030-96802-1_1

HUANG, C. F., ZHANG, W., & XUE, L. (2022), «Virtual reality scene modeling in the context of Internet of Things», *Alexandria Engineering Journal*, 61(8), pp. 5949–5958. https://doi.org/10.1016/j.aej.2021.11.022

ZHANG, J., WANG, G., CHEN, H., HUANG, H., SHI, Y., & WANG, Q. (2025), «Internet of Things and extended reality in cultural heritage: A review on reconstruction and restoration, intelligent guided tour, and immersive experiences», *IEEE Internet of Things Journal*. https://doi.org/10.1109/JIOT.2025.3553237

LAOHAVIRAPHAP, N., & WAROONKUN, T. (2024), «Integrating artificial intelligence and the Internet of Things in cultural heritage preservation: A systematic review of risk management and environmental monitoring strategies», *Buildings*, 14(12), art. 3979. https://doi.org/10.3390/buildings14123979

Reconstrucción virtual inmersiva de teatros del *Siglo de Oro* desaparecidos: La Casa de Comedias de l'Olivera (València)

Eugenio Giuseppe Nerelli
Álvaro Egea Benavente
Jaume Segura-Garcia
J. Antonio De Rus
Universitat de València
Mario Montagud
Universitat de València – Fundació i2CAT (Barcelona)

Palabras clave: Realidad virtual; teatros históricos; patrimonio cultural; humanidades digitales; visitas virtuales; tecnologías interactivas
Keywords: Virtual reality; historic theaters; cultural heritage; digital humanities; virtual tours; interactive technologies

Introducción

La integración de entornos analógicos y digitales ofrece nuevas posibilidades a las instituciones del patrimonio cultural y natural (UNESCO Institute for Statistics 2024). En este panorama, es necesario replantear las estrategias para la conservación, estudio y difusión del patrimonio desaparecido. Los teatros históricos, como los corrales de comedias del *Siglo de Oro* español, son un ejemplo paradigmático de este desafío. Su desaparición material limita significativamente la comprensión profunda de las prácticas escénicas, la arquitectura teatral y la experiencia cultural de la época, dejando lagunas en el conocimiento del patrimonio tangible e intangible asociado a ellos.

La Casa de Comedias de l'Olivera, ubicada en València y promovido por el Hospital General en 1584, fue uno de estos espacios teatrales singulares. Representa un caso de estudio relevante por su larga trayectoria, sus diversas remodelaciones documentadas, y su abrupto final con la prohibición de las representaciones teatrales y su posterior demolición en 1750 (Segura et al. 2019). Ante esta ausencia física, la reconstrucción digital emerge como una herramienta poderosa. Investigaciones previas ya habían abordado la reconstrucción virtual 3D de l'Olivera, incluyendo simulaciones acústicas detalladas y explorando aspectos de accesibilidad, aunque orientadas

principalmente a un visionado en pantallas 2D[1] (Montagud, Orero y Fernández 2020; Segura et al. 2019). Este trabajo presenta una evolución significativa de dichas reconstrucciones: su adaptación, extensión y optimización para el consumo inmersivo y la visita interactiva mediante dispositivos de Realidad Virtual (RV). El enfoque actual se ha centrado en rediseñar la interacción y la navegación para permitir una exploración activa y experiencial del espacio teatral desaparecido.

En el ámbito de las humanidades digitales, la RV se consolida como una tecnología de gran potencial para la preservación, exploración y difusión de espacios patrimoniales perdidos o inaccesibles. Permite superar barreras temporales y geográficas inherentes al estudio de estos lugares, facilitando no solo la visualización arquitectónica sino también la visita interactiva (incluyendo narrativas y actuaciones virtuales asociadas), y la adquisición de conocimiento experiencial. La RV ofrece nuevas vías para investigar y comunicar el valor del patrimonio cultural, estableciendo conexiones entre pasado y presente mediante vivencias inmersivas (Montagud, Orero y Fernández 2020). La capacidad de "estar allí" virtualmente, y de moverse e interactuar dentro del espacio, posibilita una comprensión completa, rica y realista del teatro histórico.

1. Metodología

La reconstrucción virtual de la Casa de Comedias de l'Olivera se fundamenta en una rigurosa investigación histórica y documental. El modelo arquitectónico tridimensional se basa en estudios previos sobre la historia del teatro valenciano y los corrales de comedias (Sirera Turó 1981; Ros Andreu 1981; Mouyen 1991b, 1991a; Segura-Garcia et al. 2019; Oleza 2020), complementados con el análisis de documentación original de archivo relativa a su gestión, construcción y reformas.

Para la generación del modelo geométrico se utilizaron herramientas de modelado 3D como *SketchUp,* interpretando planos históricos (Fig. 1), considerando datos de posibles estudios arqueo- lógicos indirectos e investigaciones sobre la estructura (Fig. 2), así como materiales de construcción y capacidad del teatro (García Martínez 1988). Este modelo geométrico detallado fue posteriormente importado al motor de desarrollo *Unity,* plataforma sobre la cual se ha desarrollado la interactividad y se han integrado los componentes visuales y acústicos del entorno virtual.

La transición desde una versión digital previa, pensada para pantallas 2D convencionales, hacia un entorno de RV inmersiva ha supuesto una evolución significativa del proyecto. Este salto ha requerido una revisión sustancial de la interacción con el entorno y con la tecnología asociadas, así como una optimización específica para el consumo en cascos de RV. Se ha pasado de un modelo de observación predominantemente a distancia a uno basado en la

[1] Algunos resultados de dicha visualización en 2D pueden consultarse en el siguiente canal de Youtube: youtu- be.com/@NIMLab, o directamente a través de este enlace: youtu.be/q0sHKD0qx9M

sensación de presencia corporeizada dentro del espacio virtual reconstruido. La adaptación a la RV permite al usuario moverse libremente por el espacio teatral con seis grados de libertad, explorando desde la platea hasta las gradas, los palcos o la “cazuela” (Segura et al. 2019). Esta capacidad ofrece una inmersión directa en el ambiente reconstruido y una vía privilegiada para experimentar un patrimonio arquitectónico desaparecido.

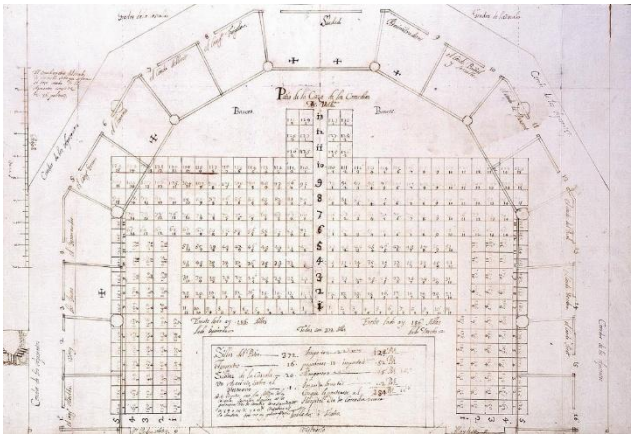


Figura 1. Planos históricos de la Casa de Comedias de l’Olivera (1678).

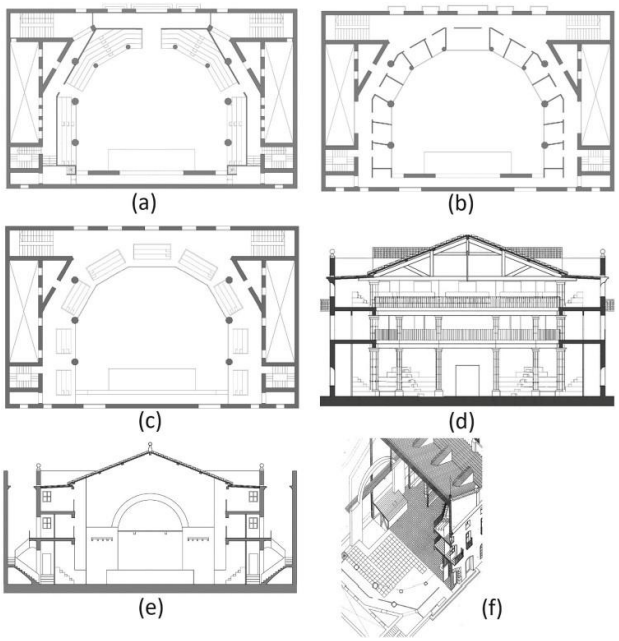


Figura 2. Planos de l’Olivera (Ros Andreu 1981): (a) planta baja, (b) primer piso, (c) segundo piso, (d) fachada, (e) sección desde la fachada y (f) sección hacia el escenario.

1.1. Adaptación y Mejoras para la Experiencia Inmersiva

Las principales adaptaciones para la experiencia inmersiva se han centrado en tres áreas: optimización, navegación e interacción.

Para garantizar una experiencia fluida y mitigar el riesgo de mareos

(*motion sickness*), se ha optimizado la reconstrucción 3D (modelos y texturas) reduciendo la carga computacional. A esto se suman mecanismos de navegación adaptados para una exploración efectiva. Se ofrece una combinación de desplazamiento continuo a velocidad controlada —que puede incluir un efecto de viñeta para reducir el campo de visión periférico— y opciones de teletransporte. Estas permiten tanto el desplazamiento instantáneo a Puntos de Interés (PoI) predefinidos (Fig. 3) como el teletransporte libre mediante un sistema de apuntado y selección en el suelo.

Figura 3. Visualización del sistema de teletransporte mediante apuntado en el entorno virtual de la Casa de Comedias de l'Olivera, cuya reconstrucción se basa en el trabajo de Oleza (2017). Se permite al usuario seleccionar el destino de su desplazamiento con trayectoria, mientras que los iconos circulares con la '*i*' representan los Puntos de Interés (PoI) que proporcionan información contextual.

En tercer lugar, se han adaptado las Interfaces de Usuario (*UI*, por sus siglas en inglés *User Interface*) y las interacciones con las mismas. En concreto, la *UI* se ha rediseñado mediante el uso de componentes tridimensionales integrados de forma diegética en el espacio virtual con tal de posi- bilitar una interacción más natural e intuitiva utilizando los controladores de RV. El menú principal permite al usuario gestionar opciones de navegación, como seleccionar PoI para teletransportarse directamente a ellos (Fig. 4), o activar elementos informativos contextuales asociados a diferentes partes del teatro. Se ha priorizado un diseño de interfaz claro y accesible desde cualquier punto del entorno virtual.

Esta adaptación a entornos de RV no solo implica cambios técnicos, sino que transforma fundamentalmente la interacción del usuario con el espacio histórico reconstruido. A diferencia de las visualizaciones convencionales en pantalla, donde el usuario mantiene una posición de observador externo, la RV

permite una experiencia corporeizada donde el visitante se convierte en agente activo dentro del entorno. Los controladores de RV facilitan una manipulación directa e intuitiva de las interfaces tridimensionales integradas en el espacio, permitiendo una interacción natural con los elementos informativos y de navegación. La capacidad de experimentar físicamente las proporciones, distancias y perspectivas del espacio teatral ofrece perspectivas imposibles de obtener mediante representaciones bidimensionales.

Figura 4. Interfaz de usuario (UI) para la navegación mediante puntos predefinidos en el teatro virtual. El sistema permite seleccionar ubicaciones específicas de interés (PoI) tanto desde un menú desplegable como desde un mapa interactivo del teatro, facilitando el teletransporte directo a estas posiciones estratégicas.

2. Resultados

El resultado principal es un entorno virtual inmersivo de la Casa de Comedias de l'Olivera que permite una exploración libre y visita interactiva, con PoI y visualización de obras grabadas con video estereoscópico (Fig. 5), optimizado para su uso con cascos de RV. Desde la perspectiva de las Humanidades Digitales, el proyecto demuestra cómo la RV puede transformar la investigación y la difusión del patrimonio cultural. Este tipo de reconstrucción digital ofrece un medio único para estudiar aspectos del teatro histórico difíciles de analizar por otras vías. La capacidad de moverse libremente dentro de edificios perdidos permite un acceso virtual sin precedentes a estos espacios patrimoniales (Segura et al. 2019).

Figura 5. Integración de actuaciones teatrales virtuales en el entorno reconstruido de la Casa de Comedias de l'Olivera, basado en Oleza (2017). La imagen muestra cómo los actores reales, grabados mediante técnicas de croma y posteriormente procesados, son reproducidos como texturas transparentes sobre el escenario virtual, creando la ilusión de una representación teatral en vivo dentro del espacio histórico reconstruido.

3. Desafíos y trabajo futuro

El desarrollo no ha estado exento de desafíos, particularmente en cuanto a lograr una optimización del rendimiento gráfico y acústico en tiempo real con el fin de garantizar una experiencia de RV fluida y de alta fidelidad, así como en cuanto a garantizar mecanismos de interacción intuitivos y efectivos. Lograr un equilibrio entre fidelidad histórica, detalle visual, rendimiento computacional, y usabilidad sigue siendo un proceso iterativo.

Las líneas de trabajo futuro se orientan a enriquecer la interactividad y la dimensión experiencial, consolidando el enfoque en la visita virtual como herramienta de exploración y adquisición de conocimiento cultural e histórico. Estas incluyen:

- mejorar la interacción del usuario con elementos específicos del entorno virtual, permitiendo manipulaciones sencillas que aporten información contextual multimodal;
- explorar estrategias de gamificación para guiar la exploración y el aprendizaje de forma lúdica;
- integrar fragmentos de obras teatrales, utilizando representaciones virtuales de actores basadas en formatos de vídeo inmersivo;
- posibilitar experiencias multiusuario, que permitan visitas guiadas virtuales, actividades educativas colaborativas o la asistencia conjunta a representaciones virtuales simuladas;

- determinar la receptividad y niveles de inmersión que este tipo de reconstrucciones virtuales otorga a usuarios finales e incluso profesionales (en el ámbito de la educación y cultura).

Conclusiones

La reconstrucción virtual inmersiva de la Casa de Comedias de l'Olivera representa un ejemplo referente del uso de tecnologías digitales avanzadas para el estudio, preservación y difusión del patrimonio cultural e histórico. Iterando desde trabajos previos centrados en el modelado 3D y la acústica virtual (Segura et al. 2019), esta fase ha priorizado la adaptación a entornos de RV con tal de facilitar visitas interactivas e inmersivas, dando la sensación de teletransporte presencial al entorno visitado. Combinando una investigación histórica rigurosa con técnicas de optimización para RV y diseño de interacción centrado en el usuario, se ha generado un entorno que permite explorar libremente un espacio cultural e histórico desaparecido, comprendiendo mejor su arquitectura, espacialidad y elementos clave del mismo. El proyecto contribuye a hacer más comprensible un elemento clave del patrimonio escénico valenciano, utilizando la RV como herramienta para la exploración, el aprendizaje experiencial y el disfrute cultural.

Agradecimientos

Este trabajo ha sido parcialmente financiado por el Programa Horizonte Europa de la Unión Europea, bajo el acuerdo nº 101135637 (proyecto HEAT), por la Agencia Estatal de Investigación (AEI), en el marco del Proyecto Estratégico Orientado a la Transición Ecológica y a la Transición Digital (TED) 2021, bajo el proyecto TED2021-131690A-C33 (REVOLUTION), y en el marco de Proyectos Generación de Conocimiento 2022, bajo el proyecto PID2022-140749OBI0 (EVOLVE), así como por la Generalitat Valenciana, en el marco de las Subvenciones del programa Prometeo para grupos de investigación de excelencia, bajo el proyecto CIPROM/2024/82 (TECLAVAL). El trabajo de Mario Montagud ha sido financiado por MCIN/AEI/10.13039/501100011033, bajo la subvención RYC2020-030679-I y el trabajo de Juan Antonio De Rus por el mismo MICIN, bajo la beca FPU20/05384, ambos también por el "Fondo Social Europeo (FSE) - Invirtiendo en tu futuro". Los autores también agradecen a Joan Oleza, Teresa Ferrer y su equipo el apoyo para lograr una reconstrucción integral y fidedigna del espacio virtual presentado.

Bibliografía

GARCÍA MARTÍNEZ, S. (1988), *Homenatge al doctor Sebastià Garcia Martínez*, Valencia, Conselleria de Cultura, Educació i Ciència, Generalitat Valenciana.

MONTAGUD, M., ORERO, P. & FERNÁNDEZ, S. (2020), «Immersive media and accessibility: Hand in hand to the future», *ITU Journal*.

MOUYEN, J. (1991a), «El corral de la Olivera de Valencia en 1678 y 1682: tentativa de definición sociológica de su público», en M. V. DIAGO & T. FERRER (eds.), *Comedias y comediantes*, Valencia, Universidad de Valencia, pp. 407–432.

MOUYEN, J. (1991b), «Las casas de comedias de Valencia», *Cuadernos de Teatro Clásico*, 6, pp. 91–122.

OLEZA, J. (2017), «Cap a una reconstrucció virtual de la Casa de Comèdies de l'Olivera (1618–1749). El treball de Josep Lluís Sirera i l'estat de la qüestió», en J. L. CANET VALLÉS, M. HARO CORTÉS, J. LONDON & G. SANSANO BELSO (eds.), *Teatro hispánico y su puesta en escena: estudios en homenaje a Josep Lluís Sirera Turó*, València, Universitat de València, pp. 261–290.

OLEZA, J. (2020), «La casa de comedias de La Olivera. Reconstrucción virtual de un teatro distinto», en *Teatro, personaje y discurso en el siglo de oro*.

ROS ANDREU, J. L. (1981), *L'efímer en la formació del barroc valencià 1586–1632*, Tesis doctoral, Universitat Politècnica de València.

SEGURA, J., MIRASOL, S., MONTAGUD, M. & OLEZA, J. (2019), «Beyond Virtual Reality: Virtual Acoustic Reconstruction of ancient theatres in the Golden Age. Towards a virtual experience of theatre at the dawn of ultra-realism», *Diablotexto Digital*, 3(8), pp. 87–110.

SEGURA-GARCIA, J., MONTAGUD, M., MIRASOL-MENACHO, S. & OLEZA-SIMÓ, J. (2019), «Theatrical Virtual Acoustic rendering with head movement interaction», *AI EDAM*, 33(3), pp. 359–368.

SIRERA TURÓ, J. L. (1981), *El teatro en Valencia durante los siglos XVI y XVII: la comedia barroca*, Tesis doctoral, Universitat de València.

TRIMBLE INC. (2025), *SketchUp*, recurso en línea. https://www.sketchup.com

UNESCO INSTITUTE FOR STATISTICS (2024), *UNESCO Framework for Cultural Statistics*. Draft Framework. Draft version, forthcoming in 2025, UNESCO.

UNITY TECHNOLOGIES (2025), *Unity*, recurso en línea. https://unity.com

Geolingüística digital: el *Atlas Lingüístico Galego* en línea

Xulio Sousa
Pedro García-Louzao
Instituto da Lingua Galega –
Universidade de Santiago de Compostela

Palabras clave: Dialectología; atlas lingüístico; humanidades digitales; cartografía digital
Keywords: Dialectology; linguistic atlas; digital humanities; digital cartography

Introducción y contexto

El Atlas Lingüístico Galego (ALGa) representa uno de los hitos más significativos en la historia reciente de la lingüística gallega. Su importancia radica no solo en la exhaustividad y rigor de sus materiales, sino también en su capacidad para documentar la diversidad lingüística de Galicia y de los territorios limítrofes donde se habla gallego. El proyecto fue concebido en la década de 1970 por un grupo de investigadores del Instituto da Lingua Galega (ILG) de la Universidade de Santiago de Compostela, con el objetivo principal de analizar la variación diatópica del gallego en sus diferentes niveles: fonético, morfosintáctico y léxico (Sousa 2017).

La labor de documentación se desarrolló entre 1974 y 1977, abarcando 167 localidades del dominio lingüístico gallego: 152 en Galicia, 7 en Asturias, 5 en León y 3 en Zamora. Estos trabajos de campo permitieron recoger un corpus de registros lingüísticos que, en su conjunto, supera el medio millón de datos, cada uno acompañado de información contextual relevante sobre los informantes y sus respuestas. La riqueza de este material lo convierte en una fuente imprescindible para el estudio de la lengua gallega, tanto desde una perspectiva diacrónica como sincrónica.

Publicación y trascendencia del ALGa

El primer volumen del ALGa se publicó en 1990, marcando el inicio de una serie editorial que, hasta la fecha, ha visto la luz en ocho volúmenes impresos: tres dedicados a la variación gramatical y cinco centrados en distintos ámbitos léxicos. El proyecto prevé completarse con un total de doce volúmenes, lo que consolidará aún más la posición del ALGa como referencia fundamental para la investigación dialectológica y sociolingüística en el ámbito gallego.

La trascendencia del ALGa va más allá de su valor documental. Su publicación ha permitido avanzar en el conocimiento de las variedades dialectales del gallego, facilitando la identificación de isoglosas, la delimitación de áreas lingüísticas y el estudio de los procesos de cambio y difusión lingüística. Además, el ALGa ha servido como base para numerosos trabajos de investigación, tanto en lingüística teórica como aplicada, y ha contribuido a la formación de nuevas generaciones de investigadores en el campo de la dialectología.

Los atlas lingüísticos tradicionales: retos y limitaciones

A pesar de su indudable valor, los atlas lingüísticos tradicionales han enfrentado importantes desafíos en su explotación científica. Investigadores en geografía lingüística suelen lamentar que estos tesoros documentales sean infrautilizados en la investigación lingüística, tanto teórica como aplicada. Entre las razones que explican esta situación, destaca la dificultad de manejo de los materiales: la información aparece dispersa en volúmenes de gran tamaño, lo que complica la recopilación y el análisis sistemático de los datos (Sousa 2020).

Esta limitación no es exclusiva del ALGa, sino que afecta a la mayoría de los atlas lingüísticos publicados en el siglo XX (Julià Luna 2021). La naturaleza misma de estos proyectos, basados en trabajos de campo extensos y en la edición de volúmenes impresos, dificulta el acceso rápido y eficiente a la información, limitando así su potencial para la investigación contemporánea.

La digitalización de los atlas lingüísticos: tendencias y ejemplos

En las últimas décadas, la irrupción de las tecnologías digitales ha transformado el panorama de la geografía lingüística (Lameli 2010). Numerosas iniciativas han surgido con el objetivo de facilitar el acceso a los materiales de los atlas lingüísticos a través de aplicaciones web y bases de datos interactivas. Estas iniciativas pueden agruparse en dos grandes categorías:

- **Versiones digitalizadas de los atlas impresos**. En muchos casos, los proyectos se limitan a ofrecer versiones digitales de los volúmenes originales, a veces complementadas con herramientas de búsqueda. Ejemplos destacados en el ámbito ibérico incluyen el *Atles Lingüístic del Domini Català* (ALDC), el *Petit Atles Lingüístic del Domini Català* (PALDC) y el *Euskararen Herri Hizkeren Atlasa* (EHHA).
- **Aplicaciones dinámicas y bases de datos interactivas**. En otros casos, los proyectos ofrecen aplicaciones web avanzadas que permiten realizar búsquedas complejas, visualizar representaciones cartográficas preeditadas y generar mapas personalizados. Entre los ejemplos más relevantes se encuentran el *World Atlas of Language Structures* (WALS), el *Atlas Lingüístico de la Península Ibérica* (ALPI; García Mouton 2018), el *CartoDialect* (edición digital del Atlas Linguistique de la France; Gillieron & Edmont 1902-1910), el proyecto

Verba Alpina (Colcuc 2020) y el *Corpus de los Atlas Lingüísticos* (CORPAT; Julià Luna 2025), que integra nueve atlas regionales españoles.

Estas iniciativas han demostrado el enorme potencial de las tecnologías digitales para la difusión y explotación de los datos geolingüísticos, facilitando el acceso tanto a investigadores como al público general (Upton 2011).

El ALGa digital: objetivos, metodología y desarrollo

En este contexto, el Instituto da Lingua Galega, con el apoyo de la Fundación Barrié, puso en marcha en 2024 el desarrollo de una versión digital del ALGa. El objetivo principal de este proyecto es ofrecer acceso a una parte significativa de los materiales recopilados en los volúmenes impresos a través de una aplicación web adaptada tanto a las necesidades de los usuarios expertos como a la curiosidad del público no especializado.

El proceso de desarrollo del ALGa digital ha implicado varias etapas clave:

- **Selección y tratamiento de los datos.** Se identificaron los materiales más relevantes de la base de datos original y se procedió a su normalización y estructuración para su integración en la plataforma digital.
- **Diseño y desarrollo de la aplicación.** Se diseñó una interfaz intuitiva y accesible, pensada para facilitar la navegación y la visualización de los datos.
- **Preedición de las representaciones cartográficas.** Se prepararon mapas interactivos que permiten visualizar la distribución de las variantes lingüísticas de manera dinámica.
- **Evaluación de las versiones provisionales.** Se realizaron pruebas de usuario y se recogieron valoraciones para mejorar la usabilidad y la funcionalidad de la plataforma.

Principales características del ALGa digital

El ALGa digital es un atlas lingüístico interactivo que ofrece al usuario la posibilidad de visualizar mapas similares a los de los volúmenes impresos, pero con funcionalidades avanzadas que permiten personalizar la representación de las variantes seleccionadas. Entre las principales características de la aplicación destacan:

- **Visualización cartográfica dinámica.** El usuario puede ocultar o resaltar símbolos, modificar formas y colores, y filtrar por variantes específicas.
- **Búsquedas avanzadas.** Los usuarios expertos pueden realizar consultas complejas sobre el corpus (también usando caracteres fonéticos) y descargar los resultados en formatos estandarizados para su posterior análisis con herramientas externas.

- **Accesibilidad y usabilidad.** La plataforma está diseñada para ser accesible tanto para investigadores como para el público general, promoviendo así la divulgación científica y la valorización del patrimonio lingüístico gallego.

Impacto y perspectivas de futuro

El desarrollo del ALGa digital es un claro ejemplo de cómo el avance tecnológico (especialmente en el ámbito de los sistemas de información geográfica (GIS) y las bases de datos) está transformando la difusión y explotación de la investigación en geografía lingüística. La plataforma no solo facilita el acceso a información de gran valor científico y patrimonial, sino que promueve nuevas formas de colaboración y análisis en el campo de las humanidades digitales.

El proyecto abre perspectivas de futuro muy prometedoras. Entre ellas, la posibilidad de integrar nuevos datos, ampliar las funcionalidades de la plataforma y establecer colaboraciones con otros proyectos similares a nivel nacional e internacional. Todo ello contribuirá a consolidar el ALGa como una herramienta de referencia para el estudio y la preservación de la diversidad lingüística en el siglo XXI.

Conclusión

El ALGa digital representa un paso adelante en la explotación y difusión de los atlas lingüísticos, demostrando el potencial de las humanidades digitales para revitalizar y revalorizar el patrimonio lingüístico. Su desarrollo constituye un modelo a seguir para otros proyectos de digitalización en el ámbito de la geografía lingüística.

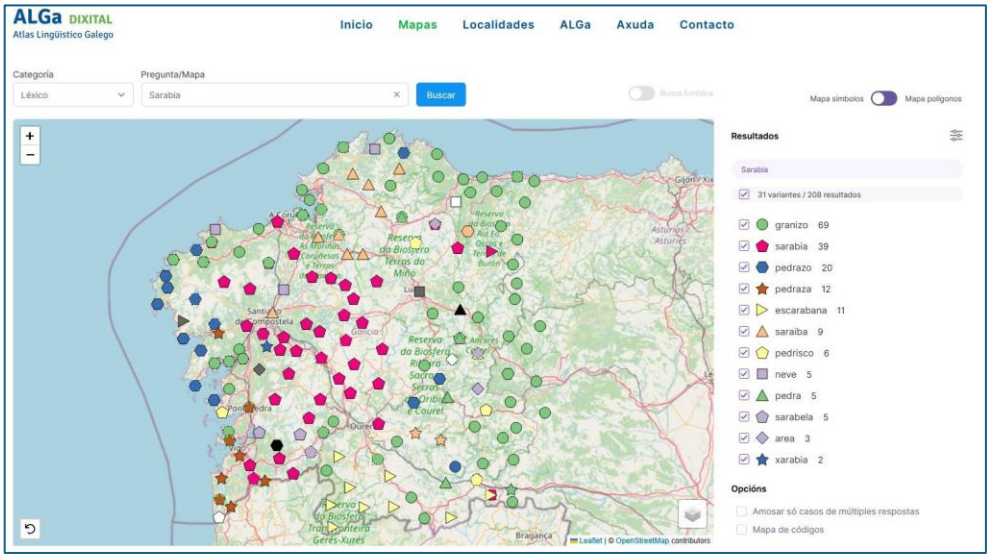


Figura 1. Muestra de un mapa léxico del ALGa dixital (versión de prueba).

Bibliografía

COLCUC, B. (2020), «La geolinguistica digitale e le sfide lessicografiche nell'era delle digital humanities: l'esempio di VerbAlpina», en C. MARRAS, M. PASSAROTTI, G. FRANZINI & E. LITTA (eds.), *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, Milano, Università Cattolica del Sacro Cuore, pp. 74–81. https://aiucd2020.unicatt.it/aiucd-Colcuc.pdf

GARCÍA MOUTON, P. (2017), «El *Atlas Lingüístico de la Península Ibérica* (ALPI) en línea. Geolingüística a la carta», *Estudis Romànics*, 39, pp. 335–342. https://raco.cat/index.php/Estudis/article/view/321054

GARCÍA MOUTON, P. (2018), «Geolingüística y Humanidades digitales: el *Atlas Lingüístico de la Península Ibérica*», en D. CORBELLA DÍAZ, A. FAJARDO AGUIRRE & J. LANGENBACHER-LIEBGOTT (eds.), *Historia del léxico español y Humanidades digitales*, Berlín, Peter Lang, pp. 191–204.

GILLIÉRON, J., & EDMONT, E. (1902–1910), *Atlas linguistique de la France*. Versión digital en *Cartodialect* 5.4.6: http://lig-tdcge.imag.fr/cartodialect5/#/

JULIÀ LUNA, C. (2021), «Del atlas lingüístico tradicional al corpus geolingüístico digital: diseño de un proyecto», *Scriptum Digital*, 10, pp. 109–147. https://doi.org/10.5565/rev/scriptum.116

JULIÀ LUNA, C. (dir.) (2025), *CORPAT: Corpus de los atlas lingüísticos*, recurso en línea. http://corpat.es [Acceso: 21/05/2025].

LAMELI, A. (2010), «Linguistic atlases – traditional and modern», en P. AUER & J. SCHMIDT (eds.), *Language and Space. Theories and Methods*, vol. 1, Berlín, Mouton de Gruyter, pp. 567–592. https://doi.org/10.1515/9783110220278.567

SOUSA, X. (2017), «From field notebooks to automatic mapping: the *Atlas Lingüístico Galego* database», *Dialectologia et Geolinguistica*, 23(1), pp. 1–22.

SOUSA, X. (2020), «Humanidades digitales y geografía lingüística: la edición digital del *Atlas Lingüístico de la Península Ibérica*», en Á. J. GALLEGO BARTOLOMÉ & F. ROCA URGELL (eds.), *Dialectología digital del español*, Santiago de Compostela, USC Editora, pp. 139–158. https://dx.doi.org/10.15304/9788418445316

UPTON, C. (2011), «Designing maps for non-linguists», en A. LAMELI, R. KEHREIN & S. RABANUS (eds.), *Language Mapping: An International Handbook of Linguistic Variation*, Berlin/New York, De Gruyter Mouton, pp. 142–166. https://doi.org/10.1515/9783110219166.1.142

El reconocimiento de entidades y la investigación histórica en portugués

Renata Vieira
Ana Sofía Ribeiro
Fernanda Olival
Universidade de Évora – CIDEHUS
Helena Freire Cameron
Universidade de Évora – CIDEHUS – Instituto Politécnico de Portalegre

El reconocimiento de entidades ha sido bien estudiado, principalmente en áreas como la política, la economía, la biología y la medicina. En el ámbito de las Humanidades Digitales, esta tarea también ha demostrado valiosas contribuciones a la investigación en Historia. Este trabajo discute la importancia de esta herramienta en particular para la investigación histórica, basada en el estudio de textos en portugués precontemporáneo y divulga las estrategias y metodologías desarrolladas en el ámbito del ChronosLab de CIDEHUS, o sea en su laboratorio de Humanidades Digitales.

La tarea de reconocimiento de entidades nombradas (NER) implica la identificación de nombres propios o vocabulario especializado. Los nombres propios requieren un tratamiento distinto, ya que pueden no estar presentes en el léxico tradicional. Además, dado que las entidades nombradas pueden ser referenciadas por una secuencia de palabras, es importante identificar la expresión completa en el texto, como medio para no descontextualizar el token. Estas secuencias deben ser reconocibles no solo para un etiquetador de frases parciales, sino también para la extracción de información y otras tareas semánticas. Las ontologías también se benefician del reconocimiento de entidades. Son importantes para modelar el conocimiento de forma estructurada y consistente, lo que facilita la comunicación y la integración de datos en sistemas complejos, identificando instancias y conceptos relevantes en relación con un dominio específico.

Para los estudios en humanidades e historia, las entidades nombradas pueden desempeñar un rol crucial para el análisis de realidades históricas en que las personas, las instituciones y la sociedad se organizaron, estructuraron y comportaron de manera diferente a los patrones actuales. Una característica es que las categorías pueden cambiar según los dominios, objetivos de

investigación y escalas analíticas. Este artículo evalúa las contribuciones que ofrece a quienes estudian las sociedades del pasado.

En los estudios históricos, hay un conjunto básico de casos (personas y grupos, lugares y organizaciones, sistemas de producción y modos de consumo, el intercambio de bienes e ideas, el estilo de vida y la religiosidad) se conceptualizan e investigan con mayor frecuencia. Aunque todos estos tienen un contexto profundo y complejo, un tema importante para un historiador, el objetivo principal del análisis histórico es proporcionar respuestas a las siguientes preguntas esenciales: ¿quién?, ¿dónde?, ¿qué?, ¿cómo? y ¿por qué?

Desde la década de 1980, con la incorporación de tecnologías digitales en la investigación académica, los historiadores han utilizado una variedad de bases de datos prosopográficas digitales y con otros formatos para contestar, de forma sistemática, a sus cuestiones orientadoras. Esto ha permitido, sobre todo, el tratamiento estadístico y relacional de grandes cantidades de datos, mientras también ha aportado un gran esfuerzo de recursos humanos y tiempo. Las herramientas digitales han mejorado los estudios seriales basados en muy grandes cantidades de datos, contribuyeron a la organización de dichos datos y a diversos análisis cuantitativos y cualitativos. Los historiadores se beneficiaron del procesamiento digital de *big data*, lo que ha profundizado la comprensión de diversos tipos de sistemas humanos complejos [J. Wang y Hu, 2024]. También los datos espaciales han sido muy valorados por los historiadores, resultando en la conversión de referencias espaciales documentales en coordenadas geográficas, aprovechando el potencial abierto por la API de Google Earth, a partir de 2005, y otras tecnologías de geovisualización más recientes. Todo ello facilitó la participación de los historiadores en el «giro espacial», creando las «geo humanidades» [Presner y Shepard, 2016]. Esto implica combinar y relacionar, en el tiempo, observaciones y análisis espacial, lo que ha contribuido a historificar espacios y paisajes. En última instancia, ha permitido la producción de atlas históricos y geo bibliotecas o repositorios históricos que posibilitaron una historización del espacio mismo (por ejemplo, el World Historical Gazetteer https://whgazetteer.org/).

Las bases de datos antes mencionadas también deben tener en cuenta los análisis de redes, que ocasionalmente se combinan con abstracciones geoespaciales. Cuando una base de datos se organiza en matrices adyacentes, basadas en biografías individuales relacionadas entre sí, los historiadores finalmente pueden observar y analizar cómo los sistemas políticos, sociales, económicos, financieros o culturales se organizaron, y comparar comportamientos económicos y sociales transnacionales a una escala cuantitativa sin precedentes [Ribeiro y Pacheco 2018] [Lemercier 2023]. Aunque no es una exacta reproducción de sistemas sociales y jerarquías del pasado, herramientas digitales de visualización y el análisis permitieron a los historiadores comprender mejor las variables que vinculan y relacionan a las personas entre sí en diferentes momentos, espacios y contextos culturales.

Los avances recientes en procesamiento de lenguaje natural e inteligencia artificial han propiciado una evolución significativa en el análisis de texto, lo que cambió el perfil de muchos conjuntos de datos. Estos avances han transformado la manera en que se comprende el análisis y la creación de textos y han dado lugar a herramientas cada vez más sofisticadas, también para historiadores. La "ingeniería" sobre texto nunca ha llegado tan lejos como ahora. Normalmente, operan con textos y, hoy en día, cada vez se digitalizan más fuentes. Por estas razones, la historia digital [Fridlund et al., 2020] es una disciplina en plena expansión. Entre las numerosas aplicaciones de IA y PLN, el reconocimiento de entidades desempeña un papel destacado.

La extracción de información contribuye a la automatización del reconocimiento y extracción de información de los documentos sobre el quién, dónde y cuándo, y rellenando los espacios en blanco sobre los atributos y características de las personas, lugares, instituciones o organizaciones colectivas [Claro et al. 2024]. Esto supone una recopilación y sistematización de datos más eficiente en menor cantidad de tiempo, permitiendo grandes cantidades de textos a procesar. Además, estas tareas ayudan a los investigadores a recopilar datos estructurados que alimentarán posteriores estudios, incluso pueden llevar a mejorar los estudios sobre lingüística histórica. Con respecto a los análisis de datos, las entidades permiten un análisis semántico del material textual. También permiten generar automáticamente glosarios de entidades anotadas, que pueden ser vinculados a datos adicionales para obtener nueva información (externa al corpus documental) relativa a un individuo o un lugar.

Actualmente, la literatura sobre reconocimiento de entidades para la investigación histórica es abundante. Existen trabajos en diversos idiomas, temas y cronologías, por ejemplo, entre otros igualmente pertinentes, para el chino [Liu et al. 2021], armenio [Tambuscio y Andrews 2021], ruso [Mozharova y Loukachevitch 2016], español [Díez Platas et al., 2021]. En cuanto a las cartas latinas, estas han sido analizadas en [Chastang et al., 2021]. La historia de la Segunda Guerra Mundial se estudia con el apoyo de reconocimiento de entidades en [Heino et al., 2017]. En [Ehrmann et al. 2023] podemos encontrar Un *survey* sobre reconocimiento de entidades en documentos históricos.

Con el desarrollo de sistemas de reconocimiento de entidades aplicados a estudios históricos, surgen nuevos desafíos. El lenguaje en textos precontemporáneos suele estar repleto de variantes con diferentes ortografías, diferencias en el uso del vocabulario o en la semántica. También, hay formas de palabras ya antiguas o, en relación con la contemporaneidad, palabras actualmente en desuso, lo que puede reducir el rendimiento de los modelos. Además, los modelos deben considerar las diferencias entre las variantes contemporáneas, especialmente las normas brasileña y portuguesa o otras, para que el portugués sea completamente procesable, a pesar de la diversidad existente.

Otros desafíos se relacionan con la definición de categorías y subcategorías más finas y sutiles que se requieren según los diferentes dominios y períodos de estudio. Un tercer desafío es integrar todo en modelos de mayor escala que permitan flujos de trabajo más rápidos, desde el archivo hasta el procesamiento de grandes datos. Finalmente, es crucial generar resultados con estructuras más efectivas para ser procesados inmediatamente por los historiadores, que puedan ser leídos directamente por software cartográfico o de otro procesamiento y producir todo lo que pueda ser reutilizado por otros usuarios.

En este contexto, pueden mencionarse dos proyectos desarrollados por las autoras, que consideran dos tipologías de fuentes históricas muy distintas entre sí. Las primeras son los libros das Monções del inicio del siglo XVII, un conjunto de cartas entre los virreyes portugueses del Estado da India (un espacio que comprendía territorios desde la cuesta oriental africana hasta el Japón, con su cabeza en la ciudad de Goa) y el rey o otras instituciones de la administración central del reino portugués.

Las segundas, son las *Memórias Paroquiais* de 1758-61. Tras el terremoto de Lisboa de 1755, la corona portuguesa envió a todos los párrocos una encuesta para obtener datos sobre la tierra, las montañas y el río, así como sobre los efectos del terremoto en el territorio en cada parroquia. La intención era también dar continuidad a un proyecto de diccionario geográfico, que un padre oratoriano había empezado unos años antes y se habían perdido los materiales con el sismo. Estos textos pasaron más tarde a ser conocidos como las Memorias Parroquiales. Contienen información valiosa sobre las tierras y la demografía, las costumbres y tradiciones, y la ocupación del territorio, incluyendo fauna y flora. Las Memorias Parroquiales son un repositorio de datos locales, pero para todo el Reino de Portugal, ó sea datos a una escala que se puede considerar micro.

Estos proyectos, cooperando entre sí, han definido su propio grupo de categorías de anotación, han avanzado con anotación manual y automática. Para la anotación manual, se verificó consenso entre los anotadores. Para la anotación automática se han desarrollado modelos a partir del aprendizaje de máquina, con aporte de modelos de lenguaje como el BERT. En esta presentación las estrategias y metodologías adaptadas a cada uno de estos conjuntos documentales serán enseñadas, mostrando cómo la investigación histórica puede adaptar el reconocimiento de entidades para, de forma eficiente, extraer información sumamente relevante para la producción del conocimiento, de una manera más eficiente y fácilmente compartible. Los recursos producidos están disponibles públicamente (www.monsoon.uevora.pt), y han sido divulgados en (Albuquerque et al., 2024) y (Nunes et al., 2025). La anotación de los *Libros das Monções* tenía como objetivo facilitar la búsqueda de información en la fuente a partir de las entidades. En cuanto al proyecto Memorias Parroquiales, se crearon conjuntos de datos que están disponibles en el repositorio Zenodo y en GitHub, junto con varias publicaciones. La orientación

es extraer información para conectarla con otros datos similares pero actuales, dando profundidad retrospectiva al conocimiento disponible online.

Por supuesto, existen otras tareas de extracción de información relevantes para las humanidades. En este trabajo se presta especial atención al reconocimiento de entidades, una técnica aún poco explorada en relación con las fuentes históricas portuguesas.

El reconocimiento de entidades y la Historia es una relación en construcción, ya que las herramientas disponibles no pueden simplemente aplicarse; estas tienen que adaptarse a las realidades pasadas y a su complejidad. También es necesario desarrollar mejores resultados para proporcionar información en formatos legibles. Encontrar el mejor modelo de reconocimiento de entidades, evaluado cualitativamente, es solo el primer paso. Aun así, el reconocimiento de entidades es un paso importante en el objetivo de los historiadores de vincular las entidades extraídas con repositorios históricos disponibles.

El REN es una tarea de PNL de larga trayectoria que también ayuda a los historiadores a construir sus conjuntos de datos y a trabajar con grandes volúmenes de información de forma más eficiente. Además, ayuda a que el pasado esté más presente en los datos actuales.

Agradecimientos

Este trabajo recibió apoyo financiero de la Fundación para la Ciencia y la Tecnología (FCT) en el contexto de los proyectos CEECIND/01997/2017, 2022.07730.PTDC y UIDB/00057/2025.

https://doi.org/10.54499/UIDB/00057.

https://doi.org/10.54499/2022.07730.PTDC

Bibliografía

ALBUQUERQUE, G. C., SOUZA, M., VIEIRA, R., & RIBEIRO, A. S. (2024), «Applying event classification to reveal the Estado da Índia», en *Proceedings of the 16th International Conference on Computational Processing of Portuguese (PROPOR 2024)*, Santiago de Compostela, Association for Computational Linguistics, pp. 247–254. https://aclanthology.org/2024.propor-1.25.pdf

CHASTANG, P., TORRES AGUILAR, S. O., & TANNIER, X. (2021), «A named entity recognition model for medieval Latin charters», *Digital Humanities Quarterly*, 15(4).

CLARO, D. B., SANTOS, J., SOUZA, M., VIEIRA, R., & PINHEIRO, V. (2024), «Extração de informação», en H. M. CASELI & M. G. V. NUNES (eds.), *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*, BPLN.

DÍEZ PLATAS, M. L., ROS MUÑOZ, S., GONZÁLEZ-BLANCO, E., RUIZ FABO, P., & ÁLVAREZ MELLADO, E. (2021), «Medieval Spanish (12th–15th

centuries) named entity recognition and attribute annotation system based on contextual information», *Journal of the Association for Information Science and Technology*, 72(2), pp. 224–238.

EHRMANN, M., HAMDI, A., PONTES, E. L., ROMANELLO, M., & DOUCET, A. (2023), «Named entity recognition and classification in historical documents: A survey», *ACM Computing Surveys*, 56(2).

FREITAS, C., & SANTOS, D. (2023), «Gender depiction in Portuguese: Distant reading in Brazilian and Portuguese literature», *Journal of Computational Literary Studies*, 2(1), pp. 1–29.

FRIDLUND, M., OIVA, M., & PAJU, P. (2020), *Digital Histories: Emergent Approaches within the New Digital History*, Helsinki, Helsinki University Press.

GRILO, S., BOLRINHA, M., SILVA, J., VAZ, R., & BRANCO, A. (2020), «The BDCamões collection of Portuguese literary documents: A research resource for digital humanities and language technology», en *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, Marseille, ELRA, pp. 849–854.

HEINO, E., TAMPER, M., MÄKELÄ, E., LESKINEN, P., IKKALA, E., TUOMINEN, J., KOHO, M., & HYVÖNEN, E. (2017), «Named entity linking in a complex domain: Case Second World War history», en *Language, Data and Knowledge (LDK 2017)*, Galway, Springer, pp. 120–133.

WANG, J., ZHANG, Y.-J., C. X., J. L., J. S., J. X., L. F., T. Z., & HU, Y. (2024), «Reconstructing the evolution history of networked complex systems», *Nature Communications*, 15, 2849.

LEMERCIER, C. (2023), «Historical and archaeological network data», en T. BRUGHMANS, B. J. MILLS, J. MUNSON, & M. A. PEEPLES (eds.), *The Oxford Handbook of Archaeological Network Research*, Oxford, Oxford University Press, pp. 347–362.

LIU, S., YANG, H., LI, J., & KOLMANIČ, S. (2021), «Chinese named entity recognition method in history and culture field based on BERT», *International Journal of Computational Intelligence Systems*, 14, pp. 1–10.

MOZHAROVA, V., & LOUKACHEVITCH, N. (2016), «Two-stage approach in Russian named entity recognition», en *2016 International FRUCT Conference on Intelligence, Social Media and Web (ISMW FRUCT)*, IEEE, pp. 1–6.

NUNES, R. O., SANTOS, J., SPRITZER, A., BALREIRA, D. G., FREITAS, C. M. D. S., OLIVAL, F., CAMERON, H. F., & VIEIRA, R. (2025), «Assessing European and Brazilian Portuguese LLMs for NER in specialised domains», en A. PAES & F. A. N. VERRI (eds.), *Intelligent Systems. BRACIS 2024*, Lecture Notes in Computer Science, 15412, Cham, Springer, pp. 215–230. https://doi.org/10.1007/978-3-031-79029-4_15

PRESNER, T., & SHEPARD, D. (2016), «Mapping the geospatial turn», en *A New Companion to Digital Humanities*, Wiley, pp. 199–212.

RIBEIRO, A. S., PINHEIRO, F. L.; SANTOS, F. C.; POLONIA, A. & PACHECO, J. M. (2018), «Structural and temporal patterns of the first global trading market», *Royal Society Open Science*, 5, 180577.

TAMBUSCIO, M., & ANDREWS, T. L. (2021), «Geolocation and named entity recognition in ancient texts: A case study about Ghewond's Armenian *History*», en *Proceedings of the Computational Humanities Research Conference (CHR 2021)*, pp. 136–148.

Experiencias con el proyecto COST *Action Distant Reading for European Literary History* (CA16204) y el desarrollo del corpus ELTeC

Borja Navarro Colorado
Universidad de Alicante

Palabras clave: Literatura europea; estudios literarios computacionales; corpus literario; *distant reading*; proyecto COST Action
Keywords: European literature; computational literary studies; literary corpus; distant reading; COST Action project

El proyecto europeo COST Action *Distant Reading for European Literary History* (CA16204 2017- 2022, https://www.distant-reading.net/) se ideó para formar una amplia red de investigadores interesados en la aplicación de métodos computacionales a los estudios literarios (*Computacional Literary Studies*, Schöch et al 2023), con un ambicioso objetivo a largo plazo: cambiar la forma en que se escribe la historia de la literatura europea. Este cambio viene motivado por el enfoque: aprovechar los modelos computacionales actuales (inteligencia artificial, procesamiento del lenguaje natural, minería de textos) para analizar grandes colecciones de texto literario y extraer datos que permitan tanto establecer rasgos comunes como desarrollar estudios comparados. Este enfoque se enmarca en lo que se ha dado en llamar *Distant Reading* (Moretti 2013, Jockers 2013, Underwood 2019): en este caso, un estudio global y unificado de la literatura europea.

El proyecto fue liderado por Christof Schöch (Univesidad de Trier, Alemania) y por Maciej Eder (Polish Academy of Sciences, Polonia). En el proyecto participaron investigadores de prácticamente todos los países europeos (31 en total). El primer objetivo (formar la red de investigadores) se cumplió perfectamente, y durante las reuniones de trabajo se pudieron ver, compartir y comentar las idiosincrasias de las tradiciones lingüísticas y literarias de cada país.

Además de largas y fructíferas discusiones sobre “best practices” en la aplicación de métodos computacionales para el análisis de texto literario, la validez científica del análisis de datos en los estudios literarios, o los fundamentos teórico-críticos del *distant reading*, en el proyecto se trabajó mucho en lo que es la base de todo estudio literario computacional: el corpus de análisis. Para poder desarrollar ese cambio en cómo se escribe la historia de la literatura europea, es necesario en primer lugar contar con un corpus de texto literario que sea representativo de esa literatura europea. El resultado de este esfuerzo es la

European Literary Text Collection, o corpus ELTeC (Odebrecht et al 2019, Burnard et al 2021, Burnard et al 2022).

Para crear el corpus ELTeC se formó un grupo de trabajo específico liderado por Carolin Odebrecht (Universidad Von Humboldt de Berlín, Alemania), Lou Burnard (Reino Unido), Martina Scholger (Universidad de Graz, Austria) y Borja Navarro Colorado (Universidad de Alicante, España). Además de representativo, el corpus debía permitir la realización de estudios computacionales y comparados. Para ello, los líderes del grupo establecieron unos criterios comunes en cuanto a la selección y marcado de los textos, así como el método y flujo de trabajo. A partir de esos criterios comunes, investigadores de cada país fueron seleccionando los textos literarios correspondientes y desarrollando todo el proceso de compilación, edición y marcado.

El diseño del corpus ELTeC fue más complejo de lo que en un principio podría parecer, precisamente por la idiosincrasia de la tradición literaria de cada país. En la presentación se darán detalles de los criterios de selección y marcado, así como problemas encontrados y las soluciones adoptadas. Se comentarán también problemas específicos aparecidos en diferentes países, con especial atención al caso español.

A modo de resumen, el primer criterio fue delimitar el corpus a un género literario y a una época concreta: la novela europea publicada entre 1840 y 1920. La novela europea en este periodo presenta una homogeneidad en cuanto a rasgos literarios que no está presente en otros géneros o en épocas anteriores. Esta homogeneidad favorece la aplicación de métodos computacionales para su análisis, así como la comparación de datos. Por ello se decidió acotar el corpus a ese género y periodo.

A partir de aquí se determinaron unos criterios de selección comunes a todos los países, de tal manera que el corpus resultante fuera, dentro de lo posible, compatible y comparable. Así, el corpus está compensado en cuanto a lugar de publicación (100 novelas por tradición literaria), a la fecha de publicación (25 novelas por cada 20 años), al sexo del o la novelista (entre el 10 y 50%, según la situación de cada país), a su éxito editorial (30% de novelas con una única edición), etc. En la presentación se mostrarán los detalles, así como los problemas surgidos y las soluciones adoptadas.

De estos criterios, el de “éxito editorial” fue el más polémico. En un principio se utilizó un criterio de canonicidad: que hubiera una cantidad similar de novelas canónicas y no canónicas. Pero el concepto de “canon” resultó muy polémico, en algunos países con connotaciones políticas negativas. Por ello se decidió cambiar ese criterio por el de “cantidad de reimpresiones”: incluir en el corpus, por cada tradición literaria, al menos 30 novelas con una única edición (la primera: novelas que entrarían, por tanto, en la categoría de “raros y curiosos”); y al menos 30 novelas con edición moderna en papel (es decir, novela que aún hoy día son leídas por un conjunto más o menos amplio de lectores).

Si bien el corpus se sigue ampliando, al finalizar el proyecto se alcanzaron las 1200 novelas previstas. En concreto, en la versión oficial del corpus ELTeC hay 100 novelas por cada uno de los siguientes idiomas: checo, alemán, inglés, francés, suizo-alemán, húngaro, polaco, portugués, rumano, esloveno, español y serbio. Otros idiomas, por diferentes motivos, no llegaron a las 100 novelas comprometidas. Un resumen del estado actual del corpus se puede en esta web: https://distantreading.github.io/ELTeC/.

El corpus ELTeC está ya en fase de explotación, pero la extracción de datos de interés para los estudios literarios no resulta fácil. Entre otros problemas, el hecho de estar formado por textos en diferentes idiomas sigue siendo una barrera para aplicar métodos computacionales. En la presentación se comentarán éste y otros problemas en su explotación para lo estudios literarios computacionales.

Ese gran objetivo de escribir una nueva historia global de la literatura europea desde el enfoque del *distant reading* aún está por conseguir. El corpus ELTeC y su explotación son solo los primeros pasos de un largo y, esperemos, fructífero camino. El corpus está disponible en el repositorio de desarrollo GitHub[1], en Zenodo[22] y TEXT Grid Repository[3]. También está accesible desde las infraestructuras CLARIN[4] y DARIAH[5], y desde la propia web del proyecto[6].

Consideramos que la comunidad de HDH puede sacar un gran provecho científico de este recurso, tanto en la explotación y análisis del corpus de novela española en sí mismo, como por los análisis comparativos con las novelas en otros idiomas. Por ejemplo, extraer datos sobre similitudes narratológicas entre novelas de diferentes países que ayuden a conformar un mapa más ajustado de las relaciones literarias europeas. Por otro lado, la experiencia alcanzada en el desarrollo del proyecto, así como los procesos definidos para la compilación y anotación de un corpus literario multilingüe de esas características, puede ser útil tanto para su ampliación con otras lenguas ibéricas (el caso del vasco está en proceso, no así el catalán o gallego), como para su ampliación con la producción novelística americana (incluyendo novelas escritas en portugués brasileño, inglés norteamericano, francés quebequés o las variantes del español, entre otros idiomas).

Agradecimientos

Proyecto financiado por la Unión Europea a través de la *COST Action Distant Reading for European Literary History* (CA16204 - Distant-Reading): www.distant-reading.net. Este trabajo también ha recibido ayudas del proyecto de I+D+i CORTEX: Generación consciente de texto «PID2021- 123956OB-I00»,

[1] https://github.com/COST-ELTeC
[2] https://zenodo.org/
[3] https://textgridrep.de/project/TGPR-99d098e9-b60f-98fd-cda3-6448e07e619d
[4] https://www.clarin.eu/
[5] https://www.dariah.eu/
[6] https://www.distant-reading.net/eltec/

financiado/a por MCIN/ AEI/10.13039/501100011033/ y por “FEDER Una manera de hacer Europa”.

Bibliografía

BURNARD, L.; NAVARRO-COLORADO, B.; ODEBRECHT, C.; & SCHOLGER, M. (2022), «Collaborative creation of a multi-lingual literary corpus. Challenges and best practices for corpus design», comunicación presentada en *COST Action Distant Reading Closing Conference*, Cracovia/en línea, 21–22 de junio de 2022.
BURNARD, L.; SCHÖCH, C.; & ODEBRECHT, C. (2021), «In Search of Comity: TEI for Distant Reading», *Journal of the Text Encoding Initiative*, 14. https://doi.org/10.4000/jtei.3500
JOCKERS, M. (2013), *Macroanalysis: Digital Methods and Literary History*, Urbana/Chicago/Springfield, University of Illinois Press.
MORETTI, F. (2013), *Distant Reading*, London, Verso.
ODEBRECHT, C.; BURNARD, L.; NAVARRO-COLORADO, B.; EDER, M.; & SCHÖCH, C. (2019), «The European Literary Text Collection (ELTeC)», comunicación presentada en *Digital Humanities 2019 (DH 2019)*, Utrecht, 9–12 de julio.
SCHÖCH, C.; DUDAR, J.; & FILEVA, E. (eds.) (2023), *Survey of Methods in Computational Literary Studies*, recurso en línea. https://methods.clsinfra.io — https://doi.org/10.5281/zenodo.7892112
UNDERWOOD, T. (2019), *Distant Horizons: Digital Evidence and Literary Change*, Chicago, University of Chicago Press.

Salvando el desajuste entre formación y necesidades del ámbito laboral en las Humanidades Digitales

María Goicoechea de Jorge
Universidad Complutense de Madrid

Palabras clave: Empleabilidad; competencias; formación en Humanidades Digitales; buenas prácticas; DARIAH
Keywords: Employability; skills; Digital Humanities training; best practices; DARIAH

La sostenibilidad de las Humanidades Digitales (HD) como campo depende de encontrar el equilibrio adecuado entre la realización de investigación académica fundamental y la atención a las necesidades formativas y educativas del sector empresarial. En la última década, ha habido una notable aparición de programas de grado y posgrado que integran las prácticas de las Humanidades Digitales en el sistema educativo, desempeñando un papel fundamental en el establecimiento del campo como un área de estudio reconocida dentro del mundo académico. Sin embargo, la velocidad de las transformaciones tecnológicas está alterando el equilibrio entre la oferta y la demanda en el mercado laboral. En general, los investigadores creen que la actual transformación digital puede dar lugar a un desequilibrio de competencias acumulado a largo plazo que requerirá una reorientación masiva del sistema educativo (Curci et al., 2024; Pouliakas, 2024; Cedefop, 2024a). A pesar de la proliferación de programas en Humanidades Digitales, persiste un desfase entre las competencias que buscan las empresas y la educación que reciben los estudiantes. La rápida evolución de la tecnología de la IA y su progresiva integración en el entorno laboral ha aumentado también la necesidad de trabajadores con competencias digitales que comprendan tanto las ventajas como los riesgos potenciales que esta tecnología conlleva. Por lo tanto, es imperativo en este momento evaluar el progreso de estos programas de HD para los futuros graduados en términos de oportunidades profesionales no sólo en los sectores académico y público, sino también en el ámbito de la empresa y el sector privado.

En esta contribución queremos presentar el libro blanco “Bridging the Gap between Industry and Education: Best Practices for the Digital Humanities” (Goicoechea et al. 2025), el cual presenta las conclusiones de una serie de estudios realizados entre 2023 y 2025, en los que se examina la adecuación entre la oferta de competencias universitarias y la demanda de los sectores públicos y privados. La investigación se centró específicamente en los

programas de Humanidades Digitales en Europa y Norteamérica, con el objetivo de identificar tanto las necesidades del sector como ejemplos de buenas prácticas. El objetivo es ofrecer recomendaciones que mejoren la empleabilidad de los graduados en Humanidades Digitales y promuevan el intercambio de conocimientos dentro del campo.

Metodología

El estudio se llevó a cabo empleando un enfoque de métodos mixtos combinando técnicas de investigación cuantitativas y cualitativas para explorar los programas de Humanidades Digitales (DH) en Europa, EE.UU. y Canadá, e identificar sus estrategias para aumentar las oportunidades de empleo de sus graduados. La metodología incluyó el análisis de datos proporcionados por la API del Digital Humanities Course Registry, la investigación documental, la organización de entrevistas, cuestionarios y dos talleres con coordinadores de programas y antiguos alumnos que trabajan en los sectores académicos, públicos y privados.

Para obtener información sobre los programas de HD en Europa, se llevaron a cabo entrevistas a 34 coordinadores de programas de máster y a 41 egresados, además se realizó un análisis de 54 programas europeos a través de investigación documental. Para obtener información sobre los programas de HD en Norteamérica, se recopilaron datos de 120 instituciones norteamericanas. Además, se realizó una encuesta mediante un cuestionario (22 respuestas) y se entrevistó a 7 coordinadores de programas de HD norteamericanos.

Resultados

Los resultados del proyecto se han publicado en diferentes foros a lo largo de los dos años de duración del proyecto:

2022-2023	**DARIAH2023 (Budapest):** Preliminary Report: "Digital Humanities and Industries: Identifying Employment Niches" **CLARIN2023 (Leuven): CLARIN in Training and Education.** In *Selected Papers from the CLARIN Annual Conference 2023*. https://doi.org/10.3384/ecp210011
2023-2024	**DARIAH2024 (Lisbon):** Workshop: "Emerging Job Profiles for DH Graduates: Bridging Gaps Between Industry & Education" / Padlet / CLARIN & DARIAH Blog Post **DH2024 (Washington DC):** Short paper "Identifying Challenges and Best Practices for DH Courses: A Comparative Analysis of European and North American Implementations" **CLARIN2024 (Barcelona):** CLARIN Bazaar Poster
2025	**DARIAH2025 (Göttingen):** White Paper: Bridging the Gap between Industry and Education: Best Practices for the Digital Humanities.

El libro blanco reúne las conclusiones del estudio, presentando las opiniones de las principales partes interesadas en el campo de las Humanidades Digitales: los coordinadores de programas de HD y los antiguos alumnos que ya han hecho la transición al mundo profesional, y ofreciendo orientaciones prácticas para los responsables de la toma de decisiones en las universidades, los coordinadores de programas de HD, los estudiantes y graduados, los empleadores que buscan talento y las infraestructuras de investigación.

De la comparación entre los enfoques europeos y norteamericanos se extraen una serie de diferencias que pueden aportar estrategias alternativas tanto a un lado como a otro del Atlántico. Una diferencia principal es que los programas europeos de HD están dirigidos al nivel de postgrado o doctorado, con la mayoría de los programas ofrecidos a nivel de máster (Fig.1), mientras que la mayoría de los programas de HD en América del Norte son *Minors* (18-24 horas de crédito aproximadamente) y *Graduate Certificates* (12-16 horas de crédito aproximadamente) (Fig.2). Esta diferencia tiene implicaciones importantes a la hora de abordar los estudios en Humanidades Digitales como formación complementaria (enfoque norteamericano), o fundamentalmente profesionalizante (enfoque europeo).

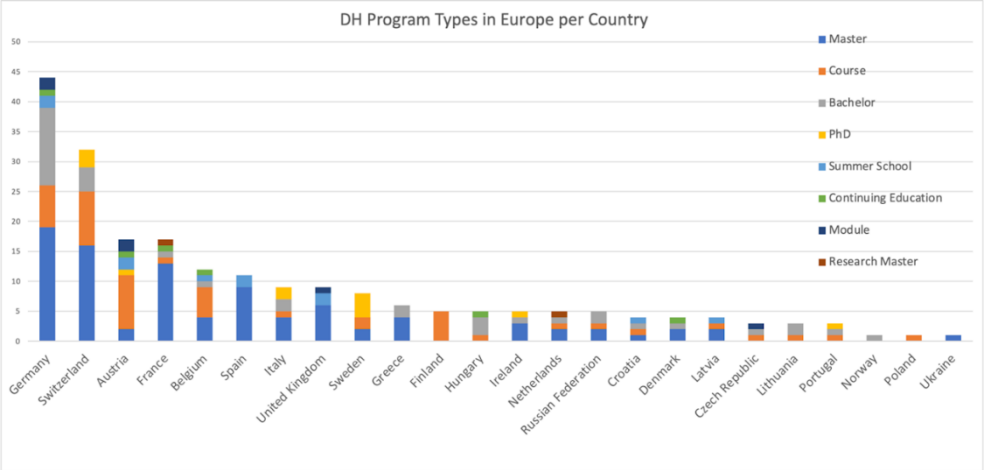


Figura 1. Visión general de los tipos de programas de HD en Europa por país.

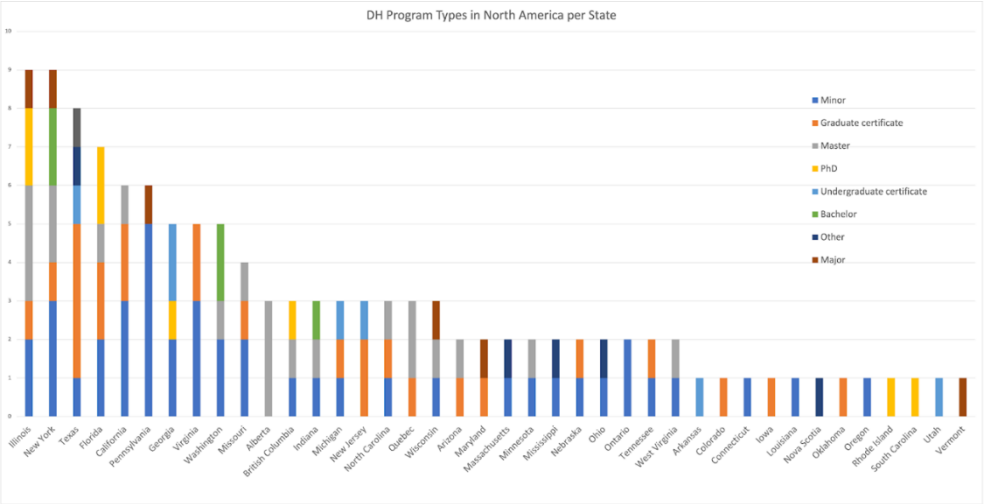


Figura 2. Visión general de los tipos de programas de HD en Norteamérica por estado.

Los programas europeos de HD se caracterizan por un enfoque interdisciplinar que abarca de forma homogénea todas las áreas de estudio identificadas (Fig.3), con un fuerte énfasis en la comunicación digital para las industrias culturales, las habilidades de programación en lenguaje natural y las prácticas y colaboraciones con la industria. Sin embargo, existen diferencias entre los países. Los países del norte de Europa, como Alemania o Austria, cuentan con sólidas infraestructuras de investigación que pueden absorber a los licenciados en su seno. Los países del Sur miran hacia las instituciones públicas y las empresas privadas como empleadores potenciales y adaptan sus programas para alimentar mejor las necesidades del mercado.

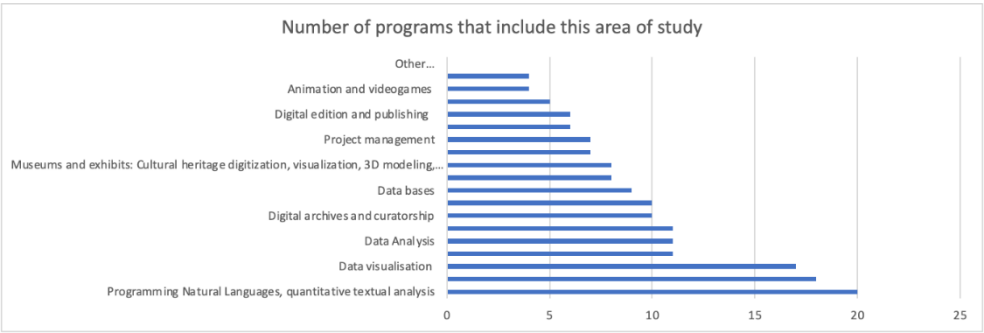


Figura 3. Número de programas europeos que incluyen esta área de estudio

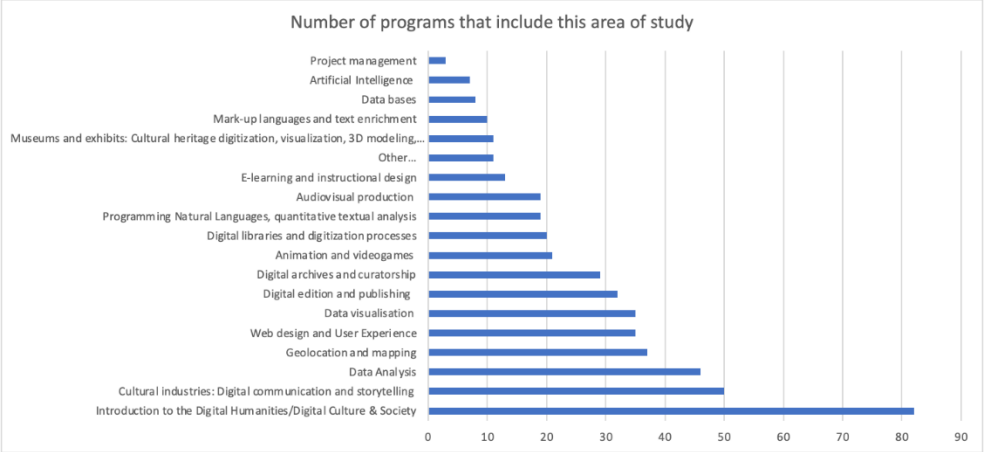


Figura 4. Número de programas norteamericanos que incluyen esta área de estudio

Dado que los *Minors* son más frecuentes en Norteamérica, es de esperar que los cursos introductorios sean más comunes. La inteligencia artificial ya está presente en los programas norteamericanos, mientras que no aparece como curso frecuente en los europeos; sin embargo, esta ausencia podría deberse también a la falta de homogeneidad en los títulos de los cursos, ya que los programas europeos ofrecen cursos, como «Programación de lenguajes naturales», que integran elementos de IA generativa.

En comparación con los programas norteamericanos más generalistas, los másteres europeos requieren cursos a medida para un número limitado de estudiantes (20-25). Son difícilmente sostenibles para las universidades. Sin embargo, las universidades norteamericanas aprovechan muy bien sus recursos ofreciendo una amplia gama de cursos ya disponibles, a menudo como optativos, y también son muy flexibles ofreciendo nuevos cursos a un ritmo rápido.

Otra diferencia significativa es que muchos programas de máster en Europa optan por las modalidades de impartición en línea o híbrida. Mientras que la modalidad presencial es la preferida en las universidades norteamericanas. Sin embargo, queda la duda de que esto siga siendo asequible para las generaciones futuras.

Conclusiones

Es evidente que el rápido desarrollo del campo ha repercutido en la aparición de nuevos perfiles profesionales a los que un programa de HD da acceso. Incluso a los coordinadores de estos programas les resulta difícil estar siempre al tanto de algunos de esos nuevos perfiles, por eso es perentorio que los coordinadores hagan un seguimiento de las trayectorias profesionales de los antiguos alumnos y de sus experiencias, ya que éstas se convertirán en una aportación muy valiosa para la formación de los nuevos estudiantes.

Uno de los principales dilemas a los que se enfrentan los programas de HD es si seguir siendo generalistas u optar por un mayor grado de especialización.

Identificar las diferentes universidades por sus áreas de especialización parecería la forma natural de evolucionar y diversificar la oferta para evitar la competencia entre programas, sin embargo, los empleadores parecen apreciar la visión general que los graduados en HD obtienen de los enfoques generales de las Humanidades Digitales, su capacidad para actuar como mediadores entre clientes y programadores y sitúan las habilidades blandas en el centro.

Los programas de HD también deben ser flexibles y versátiles para ser competitivos en este entorno. Para afrontar los retos del futuro, somos conscientes de que es necesario un enfoque global más coordinado de la formación. Esto podría lograrse, por ejemplo:

- Compartiendo e intercambiando información sobre los programas a través de un registro común (como el Digital Humanities Course Registry de DARIAH).
- Compartiendo buenas prácticas: por ejemplo, manteniendo el contacto con antiguos alumnos de HD, incorporando sus experiencias y los conocimientos recién adquiridos a las actividades de los programas o estableciendo marcos de colaboración con la industria.
- Utilizando las infraestructuras de investigación y las redes de HD establecidas, por ejemplo, HDH, DARIAH y CLARIN, para compartir conocimientos y experiencia.

Este proyecto ha sido financiado por el Plan de Financiación DARIAH para Actividades de Grupos de Trabajo 2023-2025 y es un esfuerzo conjunto del Grupo de Trabajo “Digital Humanities Course Registry” y el Grupo de Trabajo “DARIAH Community Engagement”, junto con muchos colaboradores.

Bibliografía

CEDEFOP (2024a), *Human-centred digital transitions: skill mismatches in European workplaces*, recurso en línea. https://www.cedefop.europa.eu/en/events/human-centred-digital-transitions-skill-mismatches-european-workplaces [Acceso: 14/10/2025].

CURCI, Y., GREENAN, N., & NAPOLITANO, S. (2024), «The path toward a human-centric digital transformation: what the learning capacity of organisations can do for jobs’ skills matching», comunicación presentada en *Workshop “Human-Centred Digital Transitions: Skill Mismatches in European Workplaces”*, Tesalónica, 11–12 de diciembre de 2024. https://www.cedefop.europa.eu/files/curci-greenan-napolitano.pdf

GOICOECHEA DE JORGE, M., SANZ CABRERIZO, A. D. R., VAN DER LEK, I., GHELDOF, T., JOFFRES, A., GARNETT, V., WEIS, J., KURZMEIER, M., & ANTLOGA, Š. (2025), *White Paper: Bridging the Gap between Industry and Education: Best Practices for the Digital Humanities*, *DARIAH Annual Conference*, Göttingen, Zenodo. https://doi.org/10.5281/zenodo.15633464

POULIAKAS, K. (2024), *Cedefop’s European skills and jobs survey*, comunicación presentada en *Workshop “Human-Centred Digital*

Transitions: Skill Mismatches in European Workplaces”, Tesalónica, 11–12 de diciembre de 2024. https://www.cedefop.europa.eu/files/esjs2_workshop_11-12_dec_10_12_2024.pdf

Data driven learning revisitado: del corpus a la IA generativa en el aula de español

Virginia González García
Universitat de València – Grupo Val.Es.Co

Palabras clave: Data-Driven Learning; aprendizaje inductivo; corpus discursivos; inteligencia artificial generativa; ELE
Keywords: Data-Driven Learning; inductive learning; discourse corpora; generative artificial intelligence; ELE

Problema de investigación

El enfoque Data-Driven Learning (DDL) o aprendizaje a través del manejo de datos, desarrollado por Tim Johns en los años noventa, propone un modelo inductivo de aprendizaje basado en la observación y análisis de datos lingüísticos reales. Principalmente, Johns (2002) postula la actuación de los estudiantes como "detectives" que investigan en la naturaleza de diversos fenómenos lingüísticos descubriendo regularidades lingüísticas y estableciendo generalizaciones a través de la consulta directa de corpus. El proceso característico del DDL parte de la (1) observación del aprendiz para pasar al (2) análisis, (3) inducción y posterior (4) conclusión (Johns, 2002).

A lo largo del tiempo, se han establecido dos formas principales de entender el DDL: (1) DDL en sentido estrecho, que implica el uso directo del corpus por parte del estudiante (hands-on), mediante herramientas como concordancias, para observar patrones, formular hipótesis y generalizar reglas; (2) DDL en sentido amplio, que integra el corpus con otros recursos y herramientas (tareas guiadas, gramáticas, visualizaciones y adaptaciones de resultados de concordancias, mediación docente) que facilitan el descubrimiento desde un enfoque pedagógico más guiado y con una apariencia o interfaz más amable (Vyatkina, 2016; Lusta et al., 2023).

En los últimos años, la aparición de la inteligencia artificial generativa (IAG), especialmente con modelos como ChatGPT, ha introducido una tercera vía emergente para el aprendizaje por descubrimiento. Aunque los modelos de IAG también permiten detectar patrones lingüísticos, parece complejo determinar cuál es el papel del procesamiento inductivo, fundamental en el DDL, al aplicar estos modelos. Como señalan Crosthwaite & Baisa (2023), la pregunta que surge es si con la IAG el estudiante procesa la información de manera inductiva o es si es la herramienta de inteligencia artificial la que funciona inductivamente. Resulta

complejo, a primera vista, diseñar actividades que guien el proceso con IAG de la misma manera que se hace con el uso de corpus discursivos. Conviene, por lo tanto, analizar y comparar las posibilidades de ambos recursos para ver cómo se aplican en el aula.

Metodología

En consecuencia, este trabajo propone una reflexión crítica y comparativa entre el uso de corpus y de IAG como herramientas para el aprendizaje por descubrimiento en la enseñanza del español como lengua adicional. Para ello se ha diseñado una secuencia didáctica experimental dirigida a estudiantes del máster de postgrado de Estudios Hispánicos Avanzados de la Universitat de València, en la asignatura Español como Lengua Extranjera: nuevas perspectivas. Esta secuencia experimental se enmarca en la formación de futuros profesores de español y propone integrar el aprendizaje por descubrimiento con herramientas digitales, combinando el uso de corpus lingüísticos e inteligencia artificial generativa (IAG). La propuesta se inspira en el enfoque Data-Driven Learning (DDL), lo compara con las posibilidades didácticas que ofrece la inteligencia artificial y, finalmente, explora una vía híbrida que evalúa el uso conjunto de corpus e inteligencia artificial generativa para la adquisición y el aprendizaje del componente gramatical y el componente léxico en español.

La secuencia se organiza en torno a tres fenómenos lingüísticos: (1) el uso y la posición de los relativos en español; (2) la afinidad semántica entre los verbos saber y conocer; y (3) el uso del sufijo apreciativo -azo. En el caso 1, se ofrece a los estudiantes una secuencia didáctica del uso de los relativos, tanto *hands-on* como *hands-off*, a partir del uso y explotación de corpus discursivos, concretamente el CORPES XXI. A partir de este modelo inicial, los estudiantes debe diseñar tareas que faciliten los mismos procesos inductivos y generalización de reglas, pero esta vez basados en el uso de la inteligencia artificial (*ChatGPT*). En los casos 2 y 3, los estudiantes diseñan tareas inductivas, primero con herramientas de corpus (*CORPES XXI*, *Corpus del español*, *Sketch Engine*) y luego con IAG (*ChatGPT*), y evalúan los resultados desde una perspectiva crítica. En la última fase de esta propuesta exploratoria se invita a los estudiantes a seleccionar un nuevo caso a partir de un repertorio proporcionado por la docente y elaborar una secuencia híbrida que combine el uso de corpus e inteligencia artificial generativa.

Para validar y evaluar el trabajo de los participantes en torno a estos tres fenómenos lingüísticos, se emplean varios instrumentos complementarios. Por un lado, se utiliza un diario reflexivo individual, donde los futuros docentes documentan sus decisiones didácticas, dificultades encontradas y aprendizajes derivados del uso tanto del corpus como de la IAG. Por otro, se aplica una plantilla de observación comparativa, que permite analizar ambos enfoques. Además, el alumnado realizará una revisión por pares a través de una rúbrica, basada en criterios centrados en el aprendizaje inductivo y por descubrimiento.

Finalmente, los participantes valorarán críticamente las fortalezas y riesgos percibidos en cada enfoque, a través de un cuestionario.

Resultados

Los resultados esperables, en línea con estudios recientes (Boulton & Vyatkina, 2021; Pérez-Paredes & Boulton, 2025), apuntan a una mejora en la competencia metalingüística, una mayor autonomía docente en la selección y diseño de materiales, y una conciencia más clara sobre las limitaciones de las herramientas generativas así como de aquellas propiedades que las hacen aptas para el aprendizaje por descubrimiento. Este enfoque prioriza la formación crítica y la experimentación controlada, no el uso automático y acrítico de tecnologías emergentes. De acuerdo con investigaciones previas en el ámbito de DDL (O'Keeffe, 2021; Boulton & Vyatkina, 2021), se considera que el proceso de diseño, contraste y reflexión aporta un aprendizaje más valioso que el producto final en sí.

Con respecto a la aplicación final de un modelo híbrido de DDL que incluya la IAG, nuestra propuesta se alinea con los principios de la teoría del procesamiento del input formulada por Lee y VanPatten (1995), que subraya la necesidad de diseñar tareas que ayuden a los aprendices a focalizar la atención en aspectos formales del lenguaje que de otro modo pasarían desapercibidos. Al promover la observación de patrones en contextos auténticos mediante corpus y facilitar comparaciones con output generado por IAG, las tareas diseñadas fomentan el procesamiento consciente del *input*. Así, esta versión extendida del DDL se presenta no solo como un enfoque de descubrimiento, sino como una vía para promover la adquisición de forma y significado de manera integrada.

Bibliografía

BOULTON, A., & VYATKINA, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology*, 25(3), 66–89.

CROSTHWAITE, P., & BAISA, V. (2023). Generative AI and the end of corpus-assisted data-driven learning? *Applied Corpus Linguistics*, 3, 100066.

JOHNS, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *ELR Journal*, 4, 1–16.

LEECH, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann et al. (Eds.), *Teaching and Language Corpora* (pp. 1–23). Longman.

LUSTA, A., DEMIREL, Ö., & MOHAMMADZADEH, B. (2023). Language corpus and data-driven learning (DDL) in language classrooms: A systematic review. *Heliyon*, 9, e22731.

O'KEEFFE, A. (2021). Data-driven learning: A call for a broader research gaze. *Language Teaching*, 54(2), 259–272.

PÉREZ-PAREDES, P., & BOULTON, A. (2025). *Data-Driven Learning in and out of the Language Classroom*. Cambridge University Press.

VYATKINA, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL*, 28(2), 207–226.

JOHNS, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 105–117). Rodopi.

LEE, J. F., & VANPATTEN, B. (1995). *Making Communicative Language Teaching Happen*. McGraw-Hill.

Mejora de la escritura en español mediante Inteligencia Artificial y reglas explícitas en contextos académicos

Natalia Centeno Alejandre
Pedro Manuel Moreno-Marcos
Marina Serrano Marín
Universidad Carlos III de Madrid

Palabras clave: Inteligencia artificial generativa; corrección de errores lingüísticos; analítica del aprendizaje; escritura académica; retroalimentación personalizada.
Keywords: Generative artificial intelligence; linguistic error correction; learning analytics; academic writing; personalised feedback.

Introducción

En la actualidad, la lengua escrita en español, tanto de hablantes nativos como no nativos, presenta errores gramaticales complejos que dependen en gran medida del contexto, como el tiempo verbal o el registro utilizado. En una era donde la escritura está más presente que nunca —en correos electrónicos, redes sociales y otras plataformas digitales—, estos errores se han intensificado (Condori Cajchaya & Coapaza Mamani, 2023). La tendencia a escribir de forma coloquial, imitando el habla, especialmente en redes sociales, ha contribuido a normalizar fallos como el empleo del tú impersonal en contextos formales ("tú puedes concluir" en vez de "se puede concluir"), señalados en estudios anteriores (Tusón, 1991). Estos errores, al no ser identificados adecuadamente por las herramientas actuales, persisten incluso en niveles académicos avanzados (Gumiel Molina, 2023). Ante esta situación, la inteligencia artificial (IA), especialmente la IA generativa (Yan, Martinez-Maldonado, & Gašević, 2024), se plantea como una herramienta poderosa para corregir estos errores, además de proporcionar explicaciones claras que puedan ayudar a los usuarios a comprender y mejorar su uso del idioma (Bañuelos Márquez & Romero Martínez, 2024).

Objetivos

El objetivo principal de este estudio es explorar la capacidad de un modelo de inteligencia artificial generativa, sin ajustes específicos, para apoyar a los usuarios en la mejora de la calidad de su escritura en español en contextos educativos. En concreto, se busca:

1. Diseñar un sistema de inteligencia artificial generativa para la corrección de errores lingüísticos en español, proporcionando además una explicación breve que facilite la comprensión del error por parte del usuario.
2. Evaluar la capacidad del sistema de IA creado para corregir específicamente el error del uso incorrecto del tú impersonal en textos escritos, con el fin de analizar su utilidad como herramienta de apoyo lingüístico y pedagógico.

Metodología

En esta investigación se ha evaluado el desempeño de modelos de inteligencia artificial generativa en la detección y corrección de errores lingüísticos en español, concretamente del uso incorrecto del tú impersonal en contextos que requieren construcciones impersonales formales. A continuación, se detalla la estrategia metodológica en función de los objetivos específicos planteados.

1. Para cumplir con el primer objetivo, se ha probado un modelo de inteligencia artificial generativa llamado 'Mistral', al que se le ha proporcionado un conjunto de oraciones creadas intencionadamente con errores relacionados con el uso del "tú" impersonal en contextos formales donde se requiere un uso impersonal. Se han llevado a cabo dos procesos separados: primero, el modelo corrige cada enunciado a partir de un *prompt* (instrucción específica); y segundo, utilizando la frase original con error junto con su corrección, se le solicita al modelo que genere una explicación breve del fallo identificado.
2. Para alcanzar este objetivo, se realiza una evaluación automática de las correcciones realizadas. Se utiliza la métrica GLEU, utilizada en tareas de corrección gramatical, la cual permite cuantificar el grado de coincidencia entre la corrección producida por el modelo y la versión de referencia previamente definida. Para la evaluación de la retroalimentación, se realiza una valoración cualitativa basada en la utilidad y fiabilidad de las respuestas.

Resultados obtenidos

En esta sección se presentan los resultados obtenidos en la evaluación de distintos enfoques de *prompting* aplicados al modelo de inteligencia artificial 'Mistral'. Se han probado tres niveles de detalle en las instrucciones: indicaciones sin ejemplo, indicaciones con ejemplos e indicaciones con ejemplos y un *system prompt* que aporta contexto al modelo, indicando el rol y la tarea asignados. Estas mismas estructuras de *prompt* se han empleado tanto para la corrección de errores como para la generación de retroalimentación.

El enfoque que ha ofrecido mejores resultados en ambos casos ha sido el último, instrucciones con ejemplo, acompañadas de un *system prompt*. El *prompt*

utilizado para la corrección se muestra en la Figura 1, y el correspondiente a la generación de retroalimentación en la Figura 2.

Con respecto a la corrección, en una evaluación con 300 ejemplos específicamente diseñados para contener errores relacionados con el uso impersonal del pronombre “tú”, el mejor *prompt* ha alcanzado un rendimiento medio de 0,8469 en la métrica GLEU entre la frase generada y la corrección esperada. Esta métrica mide la correspondencia entre la frase generada por el modelo y la corrección de referencia, asignando un valor entre 0 y 1, donde 1 significa una coincidencia perfecta. GLEU es especialmente útil para correcciones lingüísticas, ya que analiza no solo palabras individuales, sino también grupos consecutivos de palabras (llamados n-gramas). Al comparar tanto coincidencias como diferencias, detecta con precisión cambios sutiles pero importantes.

Por tanto, estos resultados muestran que el modelo tiene una capacidad razonable para transformar frases personales en construcciones impersonales correctas, aunque aún comete algunos errores relacionados con la concordancia o la estructura de la frase. Por ejemplo, ante la frase: “Tú identificaste patrones en los datos que antes no se habían considerado”, el modelo generó: “Se identificaron patrones en los datos que antes no se habían considerados”. En este caso, aunque realiza bien el cambio a una forma impersonal con “se identificaron”, comete un error de concordancia al escribir “considerados” en plural, posiblemente influido por el sustantivo plural “patrones”, cuando lo correcto sería “considerado”, en singular, para mantener la coherencia con la construcción pasiva refleja.

En cuanto a la retroalimentación, el enfoque principal consiste en proporcionar al modelo la frase original y la corrección previamente realizada para obtener una explicación breve del error. En la mayoría de los casos, el modelo describe la naturaleza del error y propone una manera adecuada de evitarlo.

Un ejemplo de retroalimentación para la frase incorrecta “Tú debes seguir un estilo formal y preciso" es: “La frase original utiliza el pronombre "tú" (debes) en un contexto formal, lo que resulta inapropiado. En textos académicos, es recomendable utilizar construcciones impersonales para mantener la objetividad y la distancia necesarias para un estilo formal. En lugar de "tú", se prefiere "se" para el pronombre personal, ya que este es más adecuado para referirse a un sujeto abstracto o impersonal. Además, es fundamental ajustar el verbo en tercera persona singular para mantener la concordancia verbal y evitar errores de número.

Conclusiones

Los resultados de este estudio muestran que los modelos de inteligencia artificial generativa, incluso sin ajustes específicos, tienen un alto potencial para detectar y corregir errores gramaticales complejos en español, como el uso

inapropiado del pronombre “tú” en contextos formales que requieren impersonalidad. El rendimiento alcanzado con los mejores *prompts*, especialmente aquellos que combinan instrucciones claras con ejemplos representativos, indica que el modelo puede ofrecer correcciones razonablemente precisas, aunque todavía comete errores sutiles que evidencian sus limitaciones actuales. Además, la generación de retroalimentación puede contribuir a mejorar la comprensión del error por parte del usuario, facilitando su aprendizaje de forma consciente. Estas explicaciones ayudan a los estudiantes a comprender mejor por qué cometieron el error y cómo evitarlo en un futuro.

Esta investigación demuestra que la inteligencia artificial generativa puede ser una herramienta útil para apoyar a los usuarios en la mejora de su escritura en español, al combinar corrección automática con su explicación correspondiente. Sin embargo, es fundamental seguir mejorando la precisión de las correcciones y la calidad de la retroalimentación para potenciar su impacto educativo.

Agradecimientos

Este trabajo ha sido financiado por la Comunidad de Madrid a través del convenio-subvención para el fomento y la promoción de la investigación y la transferencia de tecnología en la Universidad Carlos III de Madrid (PALABRIA-CM-UC3M).

Bibliografía

CONDORI CAJCHAYA, L., & COAPAZA MAMANI, M. Y. (2023), *Redes sociales y ortografía: un estudio con universitarios*, Editorial Idicap Pacífico.

TUSÓN, A. (1991), «Las marcas de la oralidad en la escritura», *Signos. Teoría y práctica de la educación*, 3, pp. 14–19.

GUMIEL MOLINA, S. (2023), «Por qué la inteligencia artificial no puede sustituir al lenguaje humano», *The Conversation*, recurso en línea.

YAN, L., MARTINEZ-MALDONADO, R., & GAŠEVIĆ, D. (2024), «Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle», en *Proceedings of the 14th International Learning Analytics and Knowledge (LAK) Conference*, New York, Association for Computing Machinery, pp. 101–111.

BAÑUELOS MÁRQUEZ, A. M., & ROMERO MARTÍNEZ, E. (2024), «Retroalimentación formativa con inteligencia artificial generativa: Un caso de estudio», *Wímb lu*, 19(2), pp. 1–20.

Visualización 3D para la documentación, análisis y conservación de cuevas rituales ibéricas

Agustín Diez Castillo
Sonia Machause López
Daniel Negueruela Villanueva
Universitat de València

Palabras clave: Virtualización; 3D; cuevas rituales ibéricas; arqueología; fotogrametría
Keywords: Virtualization; 3D; Iberian ritual caves; archaeology; photogrammetry

Las cuevas han sido consideradas como lugares simbólicos desde la Prehistoria. En época ibérica (ss. VI-I a.n.e.), determinadas cavidades naturales, generalmente situadas lejos de los asentamientos, fueron frecuentadas con fines rituales, principalmente entre los ss. V-IV a.n.e. El registro arqueológico conservado —cerámicas, restos de fauna, carbones, metales y, en algunos casos, restos humanos— nos muestran pautas de ofrendas y prácticas rituales desarrolladas en estos espacios.

Desde los trabajos pioneros de M. Gil-Mascarell (1975), que identificó una serie de rasgos comunes en estas cavidades, el conocimiento sobre estos espacios ha avanzado significativamente. Los estudios de las últimas décadas han evidenciado la gran heterogeneidad de prácticas y depósitos asociados a estos lugares (Machause 2019 con referencias). Sin embargo, la interpretación de estos espacios cuenta con unas limitaciones relacionadas con las características del registro material, ya que muchas de ellas fueron intervenidas con metodologías antiguas, o bien sus materiales proceden de recogidas superficiales, a menudo llevadas a cabo por aficionados. Esta situación ha provocado importantes sesgos interpretativos y ha dificultado la reconstrucción precisa de los contextos de deposición.

A pesar de estos obstáculos, en los últimos años se ha producido un notable impulso en la revisión e investigación en contexto de estas cavidades desde enfoques actualizados. En el área valenciana, destacan los ejemplos de la Cueva del Sapo (Chiva, València) (Machause et al. 2014) y la Cueva Merinel (Bugarra, València) (Martínez Perona 1992). Estas dos cuevas se encuentran en el límite sur del territorio ibérico controlado por la ciudad ibérica de Edeta (Tossal de Sant Miquel de Llíria, Valencia) (Bonet et al. 2008; Machause y Skeates 2022; Machause y Diez 2022). Las intervenciones arqueológicas llevadas a cabo desde 2018, en el caso del Sapo, y desde 2024, en el caso de Merinel, han

mostrado el enorme potencial que encierran estos espacios para comprender las prácticas rituales ibéricas y su articulación territorial. Los análisis interdisciplinares en curso están permitiendo profundizar en el carácter ritual y funerario de estas cavidades durante la Edad del Hierro, así como en sus usos previos y posteriores a esta época.

Sin embargo, para seguir profundizando en el conocimiento de estos lugares de culto, precisamos de intervenciones arqueológicas con metodología actual que nos permita comprender el uso y evolución de estos espacios. Estas metodologías deben permitir captar tanto la dimensión material como espacial y simbólica de estos enclaves. En este sentido, las tecnologías digitales, especialmente la fotogrametría y el modelado 3D, ofrecen herramientas clave no solo para el análisis y la documentación, sino también para su conservación y difusión (Diez Castillo et al. 2017 y sus referencias). En este trabajo presentamos la aplicación de estas técnicas en el marco del proyecto “Deep in caves: ritualidad ibérica en contexto” (CIGE 2022/94), centrándonos en los casos del Sapo y Merinel, con el objetivo de explorar sus posibilidades para la digitalización.

Metodología

En las cuevas de Merinel y El Sapo se ha aplicado, por vez primera, una metodología de documentación ‘paperless’, continuando la línea de trabajo iniciada en la cueva de La Cocina (Diez Castillo et al. 2017). A tal fin, en los dos casos se comenzó con el escaneo tridimensional de cada una de las cavidades. El escaneo de alta resolución, realizado en este caso por la empresa Global Mediterránea, es el paso esencial para la georeferrenciación de cada uno de los sondeos realizados en las cuevas. Los sondeos se ubican dentro de la cavidad en un sistema de coordenadas locales, para facilitar el trabajo con los modelos tridimensionales, con la ayuda de una estación local estacionada a partir de las bases con coordenadas ETRS89. El siguiente paso consiste en la colocación de puntos de apoyo en cada uno de los sondeos, cuyas coordenadas locales servirán para la posterior georreferenciación de los cada uno de los modelos tridimensionales.

Antes de comenzar a excavar cada unidad estratigráfica (UE) se realiza un modelo tridimensional tanto fotogramétrico como con un sensor LiDAR. El proceso se repite al acabar la excavación de cada una de las UUEE. En la misma cueva, se descargan los modelos 3D y se procede, con la ayuda de los puntos de apoyo, a su georreferenciación tanto en coordenadas locales, como en coordenadas ETRS89. Para la georreferrenciación se utiliza el programa Meshlab (Cignoni et al. 2008), en el caso de coordenadas locales, y CloudCompare 2.13.2 (2025) en el caso de coordenadas globales.

En el caso de Meshlab, la georreferenciación se realiza cargando un fichero con las coordenadas de los puntos de apoyo y seleccionando en el modelo cada uno de los puntos en un orden secuencial. La herramienta necesaria se denomina “reference scene”, el fichero de puntos solo puede tener un

identificador y las coordenadas x,y,z. Se han de utilizar coordendas locales para evitar errores en el modelo y al acabar el proceso, después de generar la rototraslación hay que bloquear la posición de la malla (‘freeze current matrix’). Después de estos pasos, se puede exportar el modelo.

En CloudCompare el proceso es similar, con la ventaja, de que las coordenadas pueden ser globales y el fichero de puntos puede tener además de las coordenadas y el ID, una etiqueta y tres bandas de color RGB. Si utilizamos coordenadas globales, CloudCompare nos propone trasladar lo que hacemos equivalente a las coordenadas locales. El proceso se denomina ‘align -points pairs picking-’ y consiste en identificar visualmente los puntos del modelo y los de referencia. En CloudCompare, la textura se convierte en RGB y nos permite rasterizar el modelo para obtener modelos digitales del terreno (MDT) y ortofotos de cada uno de los pasos de la excavación.

Resultados

El resultado (modelos 3D, MDT y ortofotos) sustituye, de forma ventajosa, la forma tradicional de realizar el trabajo arqueológico, tomando fotografías y cotas iniciales y finales de cada UE y permite completar el registro estratigráfico que, en nuestro caso, se realiza con una metodología madura de bases de datos relacionales que combinan la información estratigráfica (ficha de UUEE, información topográfica y croquis).

La identificación de áreas de deposición ritual y el estudio de la compartimentación simbólica del espacio constituyen ejes fundamentales para comprender las prácticas socio-culturales en contextos rupestres. Este análisis no solo delimita zonas específicas de actividad (como ofrendas, o espacios rituales), sino que revela la estructura cognitiva y social de las comunidades íberas.

Esta estrategia metodológica abre importantes posibilidades para el análisis relacional y sensorial de los espacios. Mediante herramientas SIG es posible reconstruir relaciones espaciales (proximidad, visibilidad, accesibilidad) y explorar la experiencia sensorial (sonidos, olores, iluminación) que definían el carácter ritual de las cuevas de Merinel y El Sapo. En definitiva, esta forma de documentar nuestras excavaciones enriquece la interpretación de la funcionalidad y sacralidad del espacio.

La virtualización de unos contextos arqueológicos frágiles y de difícil acceso, mediante fotogrametría o LiDAR, permite el acceso a los mismos a todo tipo de público mediante estrategias avanzadas de difusión y musealización digital. En un futuro, nos planteamos generar modelos 3D interactivos que muestren las actividades rituales a través de reconstrucciones inmersivas (realidad virtual/aumentada) para experiencias sensoriales guiadas, poniendo marcha en plataformas digitales que contextualice los hallazgos, vinculando el análisis relacional con narrativas pedagógicas accesibles.

Figuras, tablas, mapas o gráficas

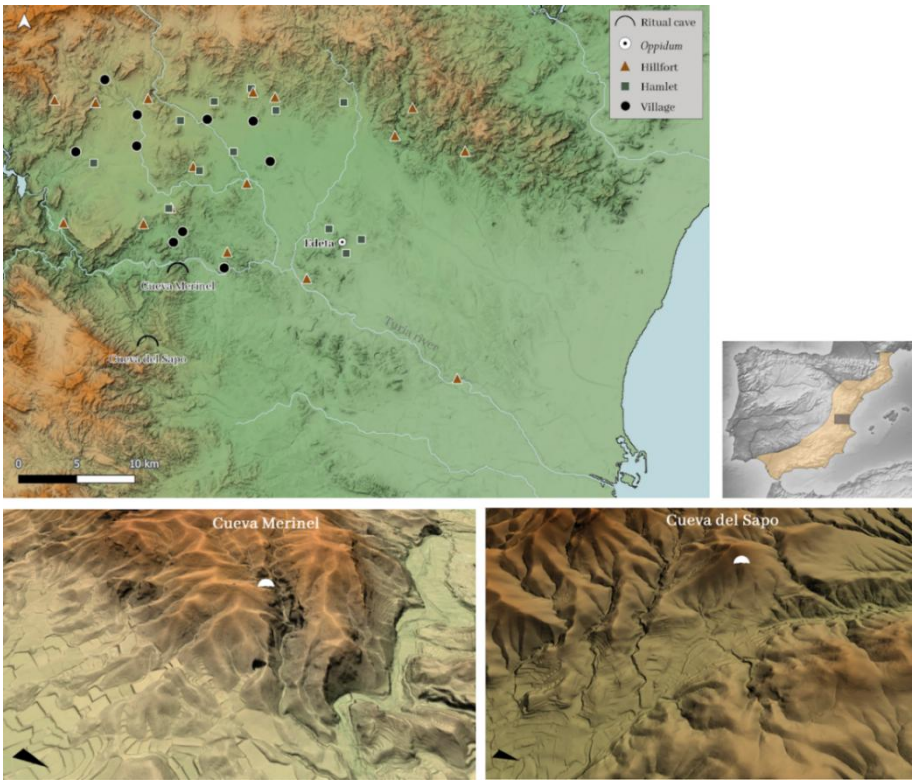


Figura 1. Localización de los casos de estudio presentados (arriba) y visualización 3D del entorno de ambas cavidades (abajo)

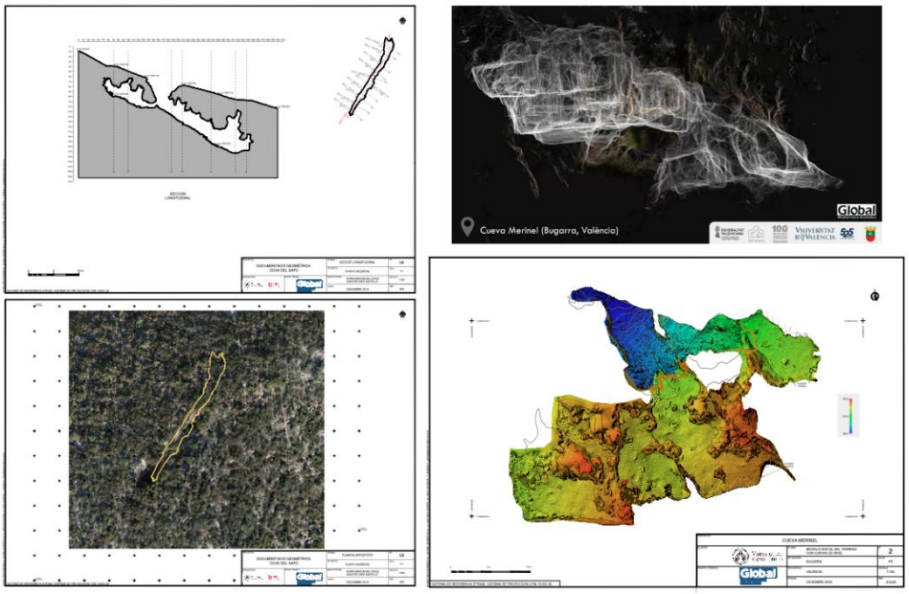


Figura 2. Documentación 3D de ambas cuevas: perfil y planta de la Cueva del Sapo (izq.) y planta y captura del vídeo “Sumérgete en el pasado con 3D: la Cueva Merinel (Bugarra, València)”:https://www.youtube.com/watch?v=RVX-fO7Wmj8 (der.).

Figura 3. Virtualización y toma de datos fotogramétricos en la Cueva del Sapo (1-2) y Cueva Merinel (3-5). Escaneado 3D de piezas en el Laboratorio Milagro Gil-Mascarell (6).

Bibliografía

BONET, H.; MATA, C.; & MORENO, A. (2008), «Iron Age Landscape and Rural Habitat in the Edetan Territory, Iberia (4th–third centuries BC)», *Journal of Mediterranean Archaeology*, 21(2), pp. 165–189. https://doi.org/10.1558/jmea.v21i2.165

CIGNONI, P.; CALLIERI, M.; CORSINI, M.; DELLEPIANE, M.; GANOVELLI, F.; & RANZUGLIA, G. (2008), «MeshLab: an Open-Source Mesh Processing Tool», en V. SCARANO, R. DE CHIARA & U. ERRA (eds.), *Eurographics Italian Chapter Conference*, Aire-la-Ville (CH), The Eurographics Association.

CLOUDCOMPARE (v. 2.13.2) [software GPL] (2025), recurso en línea: http://www.cloudcompare.org/ [Acceso: 30/03/2025].

DÍEZ CASTILLO, A.; CORTELL NICOLAU, A.; GARCÍA PUCHOL, O.; & ESCRIBÁ RUIZ, P. (2017), «3D environment for the analysis and virtual reconstruction of the archaeological fieldworks at Cocina Cave (Dos Aguas, Valencia, Spain)», *Virtual Archaeology Review*, 8(17), pp. 75–83. https://doi.org/10.4995/var.2017.7028

GIL-MASCARELL, M. (1975), «Sobre las cuevas ibéricas del País Valenciano. Materiales y problemas», *SAGVNTVM-PLAV*, 11, pp. 281–332.

MACHAUSE, S. (2019), *Las cuevas como espacios rituales en época ibérica. Los casos de Kelin, Edeta y Arse*, Jaén, Colección Arqueologías, Serie Íbera, UJA Editorial.

MACHAUSE, S.; & DÍEZ, A. (2022), «Analysing the symbolic landscape in the Iberian Iron Age: GIS, caves and ritual performance», *Zephyrus*, XC, pp. 135–158. https://doi.org/10.14201/zephyrus202290135158

MACHAUSE, S.; PÉREZ, A.; VIDAL, P.; & SANCHIS, A. (2014), «Prácticas rituales ibéricas en la Cueva del Sapo (Chiva, Valencia): Más allá del caliciforme», *Zephyrus*, LXXIV, pp. 157–179.

MACHAUSE, S.; RUEDA, C.; GRAU, I.; & ROURE, R. (eds.) (2021), *Rock and Ritual: Caves, Rocky Places and Religious Practices in the Ancient Mediterranean*, Montpellier, Mondes Anciens (PULM).

MACHAUSE, S.; & SANCHIS, A. (2015), «La ofrenda de animales como práctica ritual en época ibérica: la Cueva del Sapo (Chiva, Valencia)», en A. SANCHIS & J. L. PASCUAL (eds.), *Preses petites i grups humans en el passat. II Jornades d'Arqueozoologia del Museu de Prehistòria de València*, València, Museu de Prehistòria de València/Diputació de València, pp. 261–286.

MACHAUSE, S.; & SKEATES, R. (2022), «Caves, senses and ritual flows in the Iberian Iron Age: the territory of Edeta», *Open Archaeology*, 8(1), pp. 1–29. https://doi.org/10.1515/opar-2022-0222

MARTÍNEZ PERONA, J. V. (1992), «El santuario ibérico de la Cueva Merinel (Bugarra). En torno a la función del vaso caliciforme», en J. CABANILLES (coord.), *Estudios de Arqueología ibérica y romana. Homenaje a Enrique Pla Ballester*, Serie de Trabajos Varios del SIP, 89, València, Diputació de València, pp. 261–282.

RANZUGLIA, G.; CALLIERI, M.; DELLEPIANE, M.; CIGNONI, P.; & SCOPIGNO, R. (2013), «MeshLab as a complete tool for the integration of photos and color with high resolution 3D geometry data», en *CAA 2012 Conference Proceedings*, Amsterdam, Pallas Publications – Amsterdam University Press, pp. 406–416.

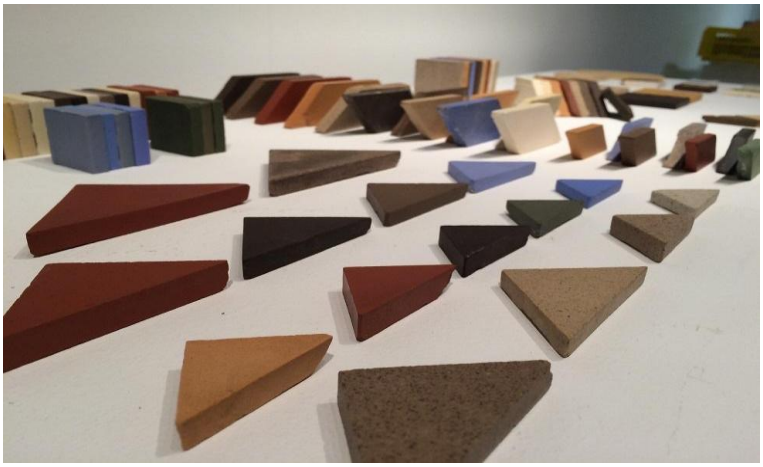
Cartografiar la cerámica arquitectónica valenciana: tecnología y difusión en Nollamap

Vanesa García López de Andújar
Universitat de València
Angela López Sabater
Xavier Laumain
Centro de Investigación y Difusión de la Cerámica Nolla

Palabras clave: Cerámica Nolla; mapeado; participación ciudadana; patrimonio arquitectónico; difusión cultural
Keywords: Nolla Ceramics; mapping; citizen participation; architectural heritage; cultural dissemination

Introducción

La cerámica Nolla, producida desde mediados del siglo XIX en Meliana (Valencia), es uno de los elementos arquitectónicos y artísticos más interesantes del Modernismo en nuestro país. De altísima resistencia y gran valor estético, estas teselas de gres monocromático de geometrías variadas (Fig.1) se colocaban una a una, formando diseños geométricos y figurativos que revestían suelos (Fig. 2), fachadas, zócalos y mobiliario urbano. Material habitual en las viviendas burguesas del siglo XIX en el levante español, también está presente en edificios de uso público y espacios urbanos. Su éxito comercial sobrepasó ampliamente el ámbito local, e incluso nacional, con la consiguiente repercusión histórico-artística, económica y social, y se han encontrado testimonios de este material en países como Portugal, Argentina, Brasil o México.



A pesar de su relevancia técnica e histórica, hasta hace pocos años este material no había sido estudiado en profundidad, por lo que la cerámica Nolla no siempre ha sido valorada ni reconocida por la ciudadanía. Este desconocimiento ha provocado que en muchos casos se hayan ocultado, desmontado o demolido pavimentos o zócalos de este tipo, amenazando la conservación de los revestimientos y del saber artesanal que su colocación implicaba.

Desde 2010 se vienen desarrollando trabajos de investigación en torno a la cerámica Nolla, su fábrica y el Palauet Nolla en Meliana. Estos estudios abordan todas las vertientes de este patrimonio: industrial, arquitectónica, artística, histórica y social, e incluyen desde investigaciones sobre la producción e historia del material hasta tareas de difusión orientadas a su reconocimiento, como talleres, publicaciones, exposiciones e incluso dos congresos internacionales. Estas acciones se han canalizado habitualmente a través del CIDCeN (Centro de Investigación y Difusión de la Cerámica Nolla), que desde 2015 ha desarrollado un amplio abanico de actividades a nivel local, nacional e incluso internacional. En la última década, desde el centro se ha constatado un creciente interés e implicación por parte de instituciones, restauradores, diseñadores y particulares, que han contactado con el CIDCeN con consultas y propuestas de colaboración.

Dentro de las acciones llevadas a cabo por la asociación, y gracias a una subvención del Ministerio de Cultura de España y de València Capital Mundial del Diseño, en 2022 surge el proyecto NollaMap**.** Se trata de una plataforma web y una app móvil que funcionan como una gran base de datos georreferenciada con edificios o lugares donde se encuentran ejemplos de mosaico Nolla alrededor del mundo. Es un mapa participativo (Fig. 3) en el que la ciudadanía puede subir ejemplos de cerámica Nolla mediante un sencillo formulario online que genera una ficha por cada elemento. Estas contribuciones se completan,

validan y publican desde el CIDCeN, resultando una herramienta que ofrece una visión clara, visual y accesible de la proyección de este patrimonio tangible e intangible a nivel global. El objetivo principal del proyecto es contribuir a poner en valor este legado de origen valenciano mediante una herramienta digital que permite localizar, registrar, compartir y consultar información desde cualquier lugar del mundo, promoviendo su conocimiento y concienciando acerca de su relevancia patrimonial.

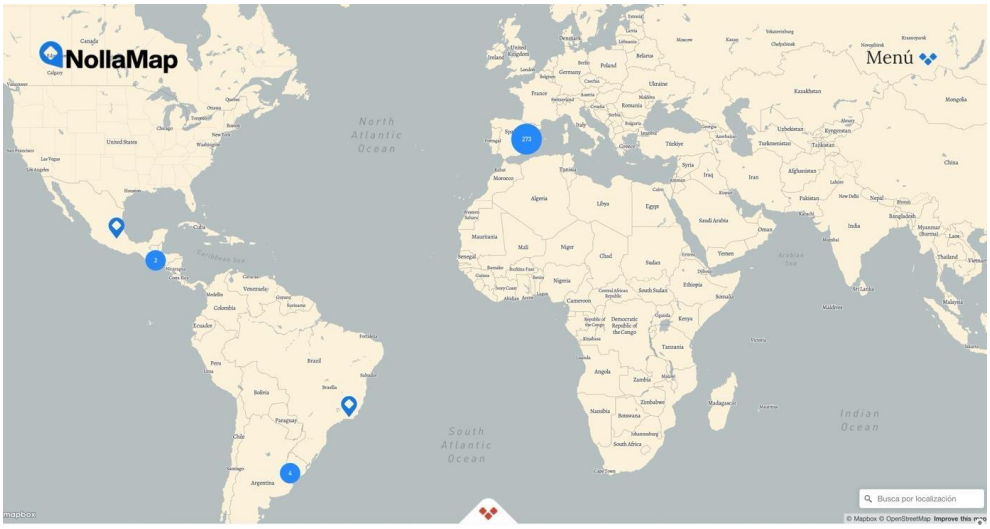


Figura 3. Mapa actualizado a 30 de junio de 2025. www.nollamap.org

Metodología

La metodología del proyecto se organizó en cuatro fases:

Investigación y selección de casos:

Esta fase se inició mucho antes del lanzamiento de NollaMap. Durante años, el CIDCeN identificó edificios y entornos urbanos donde se conservaban elementos de cerámica Nolla. Se realizaron investigaciones documentales en archivos, museos y bibliotecas (inventarios, catálogos, prensa histórica) y visitas de campo en Valencia, Barcelona, Meliana y otros municipios de alrededor. Se catalogaron casos representativos de esta cerámica colocada en edificios públicos, como el Mercado Central o el Ayuntamiento de Valencia, viviendas privadas del ensanche valenciano y del barrio de El Cabanyal, así como ejemplos en otros países, como Portugal o Argentina.

Diseño conceptual y gráfico:

Una vez concedido el proyecto, se desarrolló una fase multidisciplinar dividida en dos niveles.

Por un lado, se estableció la información necesaria para la creación de un ítem publicable, definiendo los campos que compondrían cada ficha (ubicación, tipología, cronología, etc.), se trabajó en el sistema de registro de usuarios, los

requisitos para la colaboración ciudadana y los criterios técnicos (normalización, interoperabilidad) que contribuirían a completar una aportación. También se estableció el protocolo de revisión y validación de las aportaciones ciudadanas por parte del CIDCeN. Por otro lado, se abordó el diseño gráfico y de experiencia de usuario. A partir de un trabajo realizado con personas expertas en diseño gráfico y web, se creó una interfaz clara y funcional para la app (Fig. 4) y la web, priorizando la usabilidad en dispositivos móviles y una navegación intuitiva con fuerte presencia de la imagen.

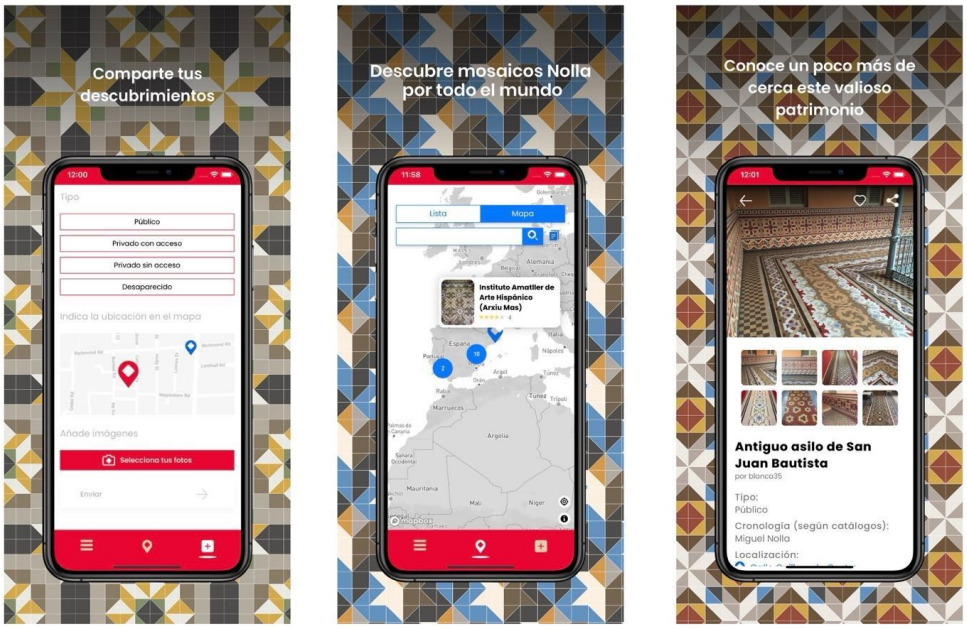


Figura 4. Pantallas de la APP para la explicación de su funcionamiento. Fuente: CIDCeN

Desarrollo técnico:

Con el apoyo de un programador, se desarrolló una app híbrida para iOS y Android con Cordova, un sitio web público implementado en Nuxt.js y una aplicación de gestión interna con Vue.js. El backend, basado en una API REST en PHP con Laravel, sirve tanto a la aplicación móvil como al sitio web y la plataforma de gestión. Además, se estableció un protocolo editorial desde el CIDCeN para verificar cada alta en la plataforma, que permite completar la información aportada por el usuario asignando valores cronológicos que la ciudadanía habitualmente no conoce, garantizando así el rigor patrimonial del contenido.

Activación participativa y difusión:

La última parte del proyecto consistió en la activación participativa, donde la dimensión colaborativa fue clave. Se organizaron sesiones presenciales y virtuales con profesionales del patrimonio y ciudadanía interesada. En ellas, se presentaron la plataforma, sus objetivos y funcionamiento, y se guio a las

personas en cómo realizar contribuciones. También se creó una cuenta de Instagram específica para dar a conocer el proyecto, con el objetivo de reforzar el alcance del mismo. En pocas semanas, se registraron más de 100 localizaciones de mosaicos Nolla alrededor del mundo. Este componente cooperativo sirvió también para validar la herramienta como recurso de formación básica en documentación digital y cartografía participativa, y contó con la participación de agentes formales y redes informales, generando cauces de diálogo entre ciudadanía y expertos.

Resultados

Actualmente, NollaMap cuenta con un registro cartográfico de más de 280 localizaciones (Fig. 5), con información estructurada, documentación fotográfica, fichas técnicas (Fig. 6) y referencias históricas. Además de su valor documental, ha generado un notable impacto social: funciona como herramienta educativa, ha generado conexiones con otros países y ha contribuido al desarrollo de redes patrimoniales activas, a la vez que ha inspirado iniciativas de conservación y recuperación en varios países del mundo que, además de participar en el mapa colaborativo, han realizado consultas y mostrado interés en la protección patrimonial de esta cerámica arquitectónica en lugares como Chile, o Uruguay.

El proyecto, además, cuenta con reconocimiento institucional, gracias al apoyo del Ministerio de Cultura de España, València Capital Mundial del Diseño y, actualmente, del Ayuntamiento de Meliana, que contribuye con el mantenimiento de la plataforma. También ha generado una red de colaboradores: agentes culturales, vecinales, docentes y ciudadanía en general, implicados en su sostenibilidad y crecimiento.

NollaMap es un claro ejemplo de cómo una estructura tecnológica accesible y colaborativa puede poner en valor un patrimonio específico, fomentar la formación básica en documentación patrimonial y promover la co-construcción de la memoria patrimonial. Constituye una evidencia del papel crucial de las herramientas digitales en la democratización patrimonial como recurso poderoso para dinamizar redes locales de conocimiento con posibilidad de ser replicado por otras iniciativas.

No obstante, tras varios años en funcionamiento, se han identificado diversos aspectos susceptibles de revisión o mejora. En primer lugar, la escalabilidad del sistema, pues el creciente número de ubicaciones registradas haría recomendable una mayor automatización en los procesos de validación. También se ha detectado la necesidad de mejorar la visibilidad pública del proyecto, fortaleciendo las estrategias de difusión con el fin de alcanzar a públicos más amplios. Por último, se considera necesario asegurar la sostenibilidad a largo plazo del proyecto, contemplando tanto el mantenimiento técnico como el acompañamiento y dinamización comunitaria.

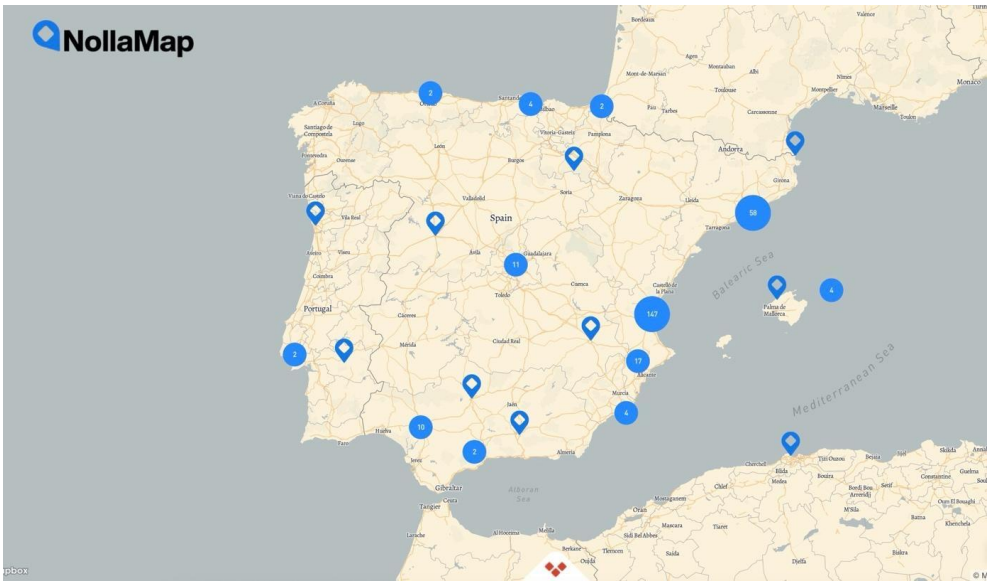


Figura 5. Detalle del mapa web de España con las aportaciones validadas a 30 de junio de 2025. www.nollamap.org

Figura 6. Modelo de ficha publicada en www.nollamap.org

Como líneas de trabajo a futuro se valora la posibilidad de extender el modelo a otros materiales cerámicos similares producidos en distintas partes del mundo, con el objetivo de analizar sus similitudes y diferencias patrimoniales, y ampliar así el alcance y utilidad de la plataforma.

Conclusiones

El proyecto NollaMap demuestra que es posible construir una plataforma colaborativa de difusión patrimonial con impacto real, validación institucional y respaldo comunitario. Su enfoque sobre la cerámica Nolla articula una

aproximación técnica, sensible y social al patrimonio, ya que contribuye a poner en valor un recurso arquitectónico y cultural que ha sido históricamente olvidado, al tiempo que combina conocimiento experto y saber popular, reconociendo la legitimidad de ambos. Además, contribuye a generar comunidad y conocimiento compartido, favoreciendo así la conservación y la valoración sostenible de este patrimonio.

Bibliografía

LAUMAIN, X.; & LÓPEZ SABATER, Á. (2016), «El mosaico Nolla. Un patrimonio valenciano de excepción», en *Congreso Internacional El modernismo en el Arco Mediterráneo (CIMAM2016)*, Cartagena, Universidad Politécnica de Cartagena, pp. 643–650. ISBN 978-84-16325-26-9.LÓPEZ SABATER, Á.; LAUMAIN, X.; & GARCÍA LÓPEZ DE ANDÚJAR, V. (2021), *The Nolla community, a mosaic of Faro Convention good practices*, en *People, places, stories. Faro Convention inspired experiences*, Communication and Publications Service (Jaume I University) & Council of Europe, pp. 7–17.

FLAT MAGAZINE (2021, 15 de octubre), «NollaMap, un mapeo para seguir la pista del mosaico Nolla», recurso en línea: https://flatmagazine.es/diseno/nollamap-un-mapeo-para-seguir-la-pista-del-mosaico-nolla/ [Acceso: 14/10/2025].

YOROKOBU (2021, 12 de diciembre), «NollaMap: al rescate del patrimonio arquitectónico valenciano», recurso en línea: https://yorokobu.es/nollamap [Acceso: 14/10/2025].

YOUTUBE (2021, 15 de octubre), «NollaMap, el mosaico de Nolla en el mundo» [Video], recurso en línea: https://www.youtube.com/watch?v=KfM33vcksTA [Acceso: 14/10/2025].

HILAME frente al espejo. Una mirada al futuro desde el presente

Agurtzane Paz Moro
Universidad Pública de Navarra
Ana Galdós Monfort
Universidad del País Vasco
Arsenio Dacosta
Universidad de Salamanca

Palabras clave: Interoperabilidad; prosopografía; sostenibilidad; transdisciplinariedad; usabilidad
Keywords: Interoperability; prosopography; sustainability; transdisciplinarity; usability

El proyecto “Hidalgos, Labradoras, Mercaderes. Una prosopografía de los territorios cantábricos en la baja Edad Media” (HILAME) es un proyecto de prosopografía centrado en la recopilación y análisis de datos biográficos de individuos que vivieron en la Edad Media. Su principal objetivo radica en proporcionar una herramienta para la comprensión de la sociedad medieval, que permita la recomposición de las redes sociales que unían a dichos individuos entre sí. También busca proporcionar un instrumento de investigación que favorezca la formulación de nuevas preguntas a partir del tratamiento, gestión y visualización de los datos recuperados mediante el análisis documental (Dacosta y Díaz de Durana, 2018). Entre sus propósitos asimismo se encuentran su interés por contribuir al debate sobre las Humanidades Digitales y la convergencia con otros proyectos para hacer interoperables sus bases de datos. Para alcanzar esos objetivos, HILAME ha puesto en práctica diversas estrategias que, aunque han supuesto un desafío, también han suscitado nuevas vías de exploración, y han impulsado un profundo proceso de renovación. Presentar esos retos y los caminos por los que estamos transitando para abordarlos es el propósito de esta comunicación.

Comentemos primero la interoperabilidad. No constituye ninguna novedad recordar que HILAME ha llevado a cabo con éxito la convergencia con *Scripta manent*, alcanzando la interoperabilidad de algunos de sus datos desde 2019 (Dacosta y Jular, 2020). Prueba de ello son los registros ya introducidos —en torno a 800— consultables a través de búsquedas generales o particulares. No obstante, a lo largo de este año (2025), esa interoperabilidad se va a convertir en bidireccional gracias a la inclusión de forma íntegra en *Scripta manent* de algunos documentos ya importados en HILAME.

Figura 1. Interoperabilidad entre HILAME y *Scripta manent* (https://www.hilame.info)

Para abordar al objetivo de contribuir al debate sobre las Humanidades Digitales resulta imprescindible plantear la sostenibilidad de los contenidos digitales, un desafío creciente por la retirada de sus principales investigadores o la obsolescencia del software. Por ello, HILAME se ha integrado, junto con otros diez proyectos liderados por distintas universidades y centros de investigación, en "Humanidades en Común, dispositivo de ciencia abierta" (HeC), un concentrador de proyectos de investigación en Humanidades que nació en 2023 por iniciativa del grupo "Sociedades, Procesos, Culturas (siglos VIII al XVIII)" de la EHU. Todos ellos comparten su compromiso con el desarrollo de las Humanidades Digitales y también de la Ciencia Abierta, integrando, además de a humanistas de distintas disciplinas, a otros especialistas con perfiles técnicos.

Figura 2. Presentación de HeC (https://humanidadesencomun.eu)

Respecto a la prosopografía de HILAME, algunos de los retos en los que estamos trabajando vienen derivados de su crecimiento, y de la necesidad de

aplicar sus procesos metodológicos a nuevas fuentes de datos (Dacosta, Paz Moro y Díaz de Durana, 2025). Uno importante tiene que ver con el uso del *factoide*, que podemos definir, según Smythe (2007), y Pasin y Bradley (2015), como toda información que involucra a una o varias personas y que aparece reflejada en una fuente documental medieval. Para poder materializar este concepto es necesario proceder a una categorización de la información contenida en las fuentes, clasificando los datos en virtud de los acontecimientos que marcan el desarrollo vital de los individuos documentados. La creación inicial de estas categorías se realizó a partir de una importación con datos relativos a las luchas de bandos en los territorios del actual País Vasco (Guerra, 1930). Su proceso de creación fue coherente y atendiendo a la naturaleza de la fuente, aportando un elemento cohesionador al conjunto de *factoides*. Sin embargo, una gran parte de las categorías generadas durante la génesis de la base de datos respondieron a la necesidad de documentar los enfrentamientos banderizos y no tuvieron en cuenta, al menos no desde una perspectiva amplia, otros escenarios.

En este contexto, el aumento de datos en HILAME, que se ha producido paulatina pero constantemente, ha permitido diversificar la información contenida. Nos referimos a instituciones religiosas femeninas o a linajes rurales y urbanos, entre otros, cuya información procede de colecciones diplomáticas ya publicadas; así como a la idiosincrasia de la ciudad bajomedieval, documentada gracias al trabajo de edición de las Actas Municipales de Vitoria, fechadas en el siglo XV. Esta realidad ha requerido incrementar el número de *factoides* disponibles. No obstante, a diferencia de la uniformidad inicial, en este caso, la clasificación ha respondido a necesidades puntuales durante el procedimiento de importación. A consecuencia de ello, el conjunto de *factoides* usados inicialmente en la base de datos prosopográfica de HILAME están siendo revisados y actualizados.

No es el único reto metodológico que buscamos afrontar en HILAME. Sin lugar a dudas, la visualización de datos también es un asunto pendiente (Jular y Dacosta, 2018). Es indudable que el desafío deriva de la naturaleza de los datos y de la estructuración de los mismos, y se manifiesta a través de las citadas categorías. Una correcta categorización es esencial para evitar confusiones en la representación de individuos dadas las dificultades vinculadas con lagunas informativas, ambigüedades, nomenclatura u homonimia, entre otros; pero también en el establecimiento de los nexos entre personas, de naturaleza e intensidad variadas. Las conexiones entre individuos, como parentescos, alianzas políticas y vínculos comerciales, generan redes de relaciones densas y difíciles de visualizar. Aunque la plataforma ya aporta un prototipo Gephi personalizado, que ejemplifica un modelo de representación de datos que consideramos será clave en el futuro, el resultado final todavía dista mucho de estar completado.

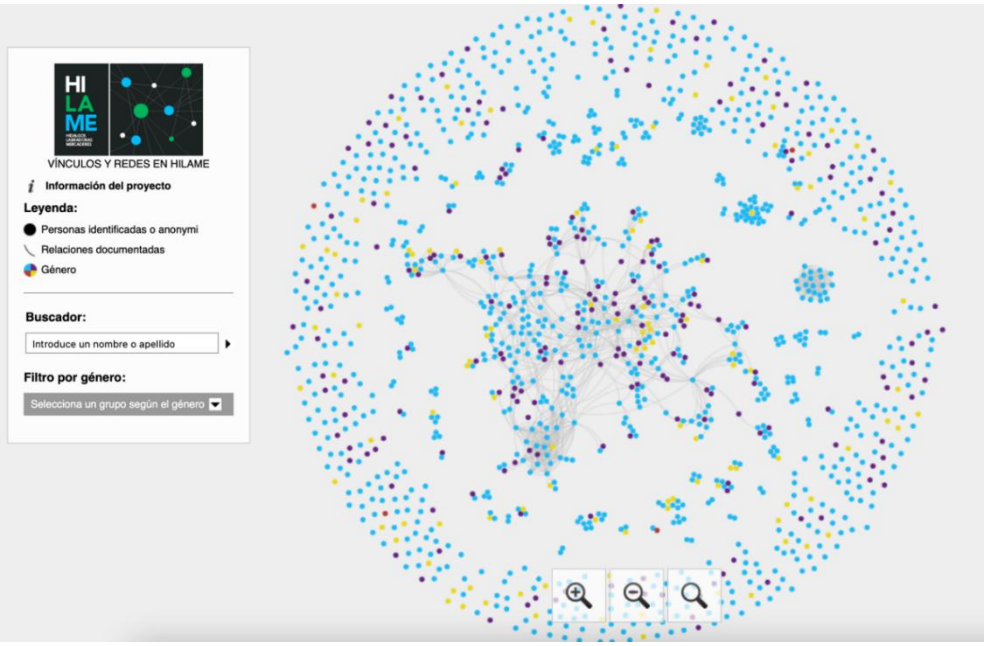


Figura 3. Prototipo Gephi para la visualización de redes sociales

En el contexto actual de HILAME, hay que destacar dos asuntos más: el desarrollo específico para mejorar la experiencia del alumnado en prácticas y la implicación del equipo en tareas de divulgación y transferencia. Ni que decir tiene que la participación casi ininterrumpida del alumnado desde 2021 refleja el carácter formativo e integrador de HILAME, y consideramos que constituye uno de los pilares del proyecto. Los precursores fueron un grupo de cuatro estudiantes de la EHU que procesaron los datos contenidos en las Actas Municipales de Vitoria durante los años 80 del siglo XV, y cuyos primeros resultados presentamos al Congreso de HDH2021 (Paz Moro y Jular Pérez-Alfaro, 2021). Aunque la práctica finalizó de forma exitosa, hemos detectado una serie de limitaciones en el sistema empleado para la importación de datos, lo que nos ha llevado a reformular ese mecanismo para favorecer tanto la usabilidad y la futura importación de los datos a HILAME, como la experiencia de enseñanza/aprendizaje en beneficio del grupo de estudiantes participantes. Se trata de una nueva aplicación a la que hemos denominado “HILAME para el alumnado que investiga” (HAI) que además de ofrecer un sistema de documentación más amigable e intuitivo, permite al docente realizar y comunicar correcciones en tiempo real a través de un chat personal, facilitando la realización de las modificaciones pertinentes.

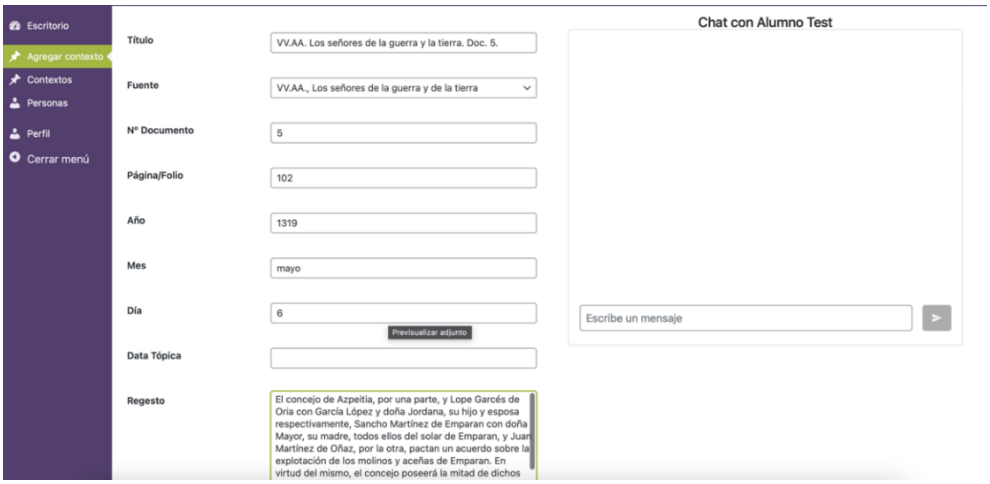


Figura 4. Espacio de trabajo en HAI

Finalmente, el desafío de la comunicación, para la divulgación científica y la transferencia de conocimiento. Esta labor no se ha circunscrito exclusivamente a difundir los avances que el grupo y proyecto de investigación vinculado con HILAME ha protagonizado, sino, más importante aún si cabe, nos sirve para dar a conocer su contenido. Son diversas las iniciativas en las que el equipo de HILAME se ha involucrado: la divulgación de artículos en un medio generalista donde, a partir del trabajo directo con fuentes documentales, construye relatos accesibles que permiten acercar el conocimiento a un público amplio, sin renunciar al rigor histórico; la perspectiva de género participando en un proyecto bautizado como WikiDamas, que, coincidiendo con el 8 de marzo de 2025, ha permitido visibilizar en Wikipedia a 16 mujeres de los linajes de Velasco y Ayala.

A modo de colofón, solo resta señalar que queda demostrada la apuesta de quienes integramos HILAME por la ciencia abierta y colaborativa (con voluntad de acercarnos al cumplimiento de los principios FAIR y la integración de nuestros datos en las plataformas europeas que se están desarrollando), por la convergencia y la interoperabilidad, por la integración de nuevas generaciones que garanticen la sostenibilidad, y por la transferencia y comunicación tanto de nuestro trabajo como de los contenidos que conforman nuestra base de datos. Por ello, estamos inmersos en una constante renovación y mejora de HILAME para intentar plantear soluciones innovadoras que, ¿por qué no? puedan constituir un modelo para su replicación.

Bibliografía

DACOSTA, A., & DÍAZ DE DURANA, J. R. (2018), «Prosopografía y bases de datos. Desafíos teóricos y metodológicos para el estudio de la Edad Media», en A. I. CARRASCO MANCHADO (ed.), *El historiador frente a las palabras: Lenguaje, poder y política en la sociedad medieval: Nuevas herramientas y propuestas*, Lugo, Editorial Axac, pp. 191–217.

DACOSTA, A., & JULAR, T. (2020), «HILAME ante los desafíos de las Humanidades Digitales: Convergencia e interoperabilidad», en S. DE LA

TORRE, E. ETXEBARRIA GALLASTEGI & J. R. DÍAZ DE DURANA (eds.), *«Valer más en la tierra». Poder, violencia y linaje en el País Vasco bajomedieval*, Madrid, Sílex, pp. 87–100.

DACOSTA, A., PAZ MORO, A., & DÍAZ DE DURANA, J. R. (2025), «Goldsmithing Data: Features and Methodological Challenges of the Use of Prosopography», en *Digital Humanities in Medieval and Early Modern Spanish Texts*, London, Routledge, pp. 253–269.

GUERRA, J. C. (1930), *Oñacinos y gamboinos. Rol de banderizos vascos, con la mención de las familias pobladoras de Bilbao en los siglos XIV y XV*, San Sebastián.

JULAR, T., & DACOSTA, A. (2018), «HILAME (Hidalgos, Labradoras, Mercaderes): procesamiento y visualización de datos prosopográficos», en *Humanidades Digitales. Miradas hacia la Edad Media*, Berlín/Boston, De Gruyter, pp. 79–89.

PASIN, M., & BRADLEY, J. (2015), «Factoid-based prosopography and computer ontologies: Towards an integrated approach», *Digital Scholarship in the Humanities*, 30(1), pp. 86–97.

PAZ MORO, A., & JULAR PÉREZ-ALFARO, T. (2023), «HILAME: de la persona al dato, del dato a la persona», en *Scire vias: humanidades digitales y conocimiento*, A Coruña, Universidade da Coruña, pp. 325–345.

SMYTHE, D. (2007), «A Whiter Shade of Pale: Issues and Possibilities in Prosopography», en K. KEATS-ROHAN (ed.), *Prosopography. Approaches and Applications. A Handbook*, vol. 13, Oxford, Oxford University Press, pp. 127–137.

Canon clásico y academia hispanohablante: patrones de canonización en textos académicos del siglo XXI

Luisa Ripoll-Alberola
Andreas Niekler
Manuel Burghardt
Computational Humanities – Universidad de Leipzig

Palabras clave: Canon; humanidades digitales; modelado de temas; PLN; cienciometría.
Keywords: Canon; digital humanities; topic modeling; NLP; scientometrics.

El término latino canon comenzó a utilizarse en el siglo XVIII para referirse a listas selectivas de autores antiguos. En clásicas, "canon" y "canónico" designan autores seleccionados desde la antigüedad según diversos criterios (Pfeiffer 1968). Dado que la erudición desempeña un papel clave en la configuración del canon, estamos interesados en estudiar la presencia de autores grecorromanos en corpus de artículos académicos contemporáneos. Estas referencias pueden indicar autoridad y pertenencia a la tradición intelectual, mientras que su ausencia o uso selectivo puede reflejar cambios en las prioridades disciplinarias. El objetivo de este trabajo es explorar cómo se propaga el conocimiento de autores canónicos antiguos en la academia hispanohablante de las primeras dos décadas del s. XXI, y si existen diferencias respecto a nuestros resultados previos en textos académicos en inglés (Ripoll-Alberola et al., 2025). Esta investigación forma parte del proyecto MECANO, que explora «la mecánica de la formación del canon y la transmisión del conocimiento de los textos de la antigüedad grecorromana». Este tema es central en los estudios literarios y las metodologías computacionales están demostrando ser interesante para rastrear la presencia o ausencia de autores en la evolución de los cánones (Romanello 2013; Berti 2024). En este contexto, cuando nos referimos a “autores canónicos” será siempre de la antigüedad clásica, y con “referencias canónicas” nos referimos a las referencias a dichos autores.

Metodología y corpus

En etapas anteriores de nuestra investigación, estudiamos la presencia de referencias canónicas en un corpus académico en inglés incluyendo diversas disciplinas provenientes de JSTOR (Luhmann y Burghardt, 2022), usando como patrón comparativo una lista de ocurrencias de autores extraída de la bibliografía

(autoritativa para los estudios clásicos) l'Année Philologique[1] (en adelante APh).

Usaremos el CORPES (Corpus del Español del s. XXI), creado por la Real Academia Española: un corpus extenso que incluye varios géneros y es accesible en línea. Como no podemos acceder a través de una API, se realizó una búsqueda autor por autor, siguiendo el orden de la lista de APh y filtrando los textos académicos por el campo "Tipología". Se descargaron metadatos y concordancias de ~80 tokens (un párrafo), totalizando casi 300.000 tokens[2] (Fig. 1). Después de crear este subcorpus, aplicaremos modelado de temas y otras técnicas de Procesamiento de Lenguaje Natural (PNL) para extraer información respecto a las referencias canónicas.

	REFERENCIA	BIBLIOGRAFÍA	AUTOR	TÍTULO	FECHA	CRITERIO	BLOQUE	MEDIO	SOPORTE	TEMA	PAÍS	ZONA	TIPOLOGÍA	NOTAS	CONCORDANCIA
0	Propercio	Rico, Francisco: «Antiguos y modernos». Los di...	Rico, Francisco	Los discursos del gusto. Notas sobre clásicos ...	2003	Primera edición	No ficción	Escrito	Libro	Ciencias sociales, creencias y pensamiento	España	NaN	Académico	Texto perteneciente a la colección de trabajos...	: la ineludible búsqueda en los clásicos de un...
1	Propercio	Gil, Juan: El burlador y sus estragos. Madrid:...	Gil, Juan	El burlador y sus estragos	2011	Primera edición	No ficción	Escrito	Libro	Ciencias sociales, creencias y pensamiento	España	NaN	Académico	El texto contiene citas ajenas al autor	en la epopeya romana. De Dido solo se sabe que...
2	Lucano	Muñoz Machado, Santiago; Sánchez Ron, Manuel: ...	Muñoz Machado, Santiago; Sánchez Ron, Manuel	Los itinerarios de la libertad de palabra	2013	Primera edición	No ficción	Escrito	Libro	Política, economía y justicia	España	NaN	Académico	NaN	aquellos autores en la literatura americana de...
3	Lucano	Egido, Aurora: La búsqueda de la inmortalidad ...	Egido, Aurora	La búsqueda de la inmortalidad en las obras de...	2014	Primera edición	No ficción	Escrito	Libro	Ciencias sociales, creencias y pensamiento	España	NaN	Académico	El texto contiene citas ajenas al autor, así c...	Muchos habían sido los que anteriormente volar...
4	Cicerón	Rojo, Guillermo: El lugar de la sintaxis en la...	Rojo, Guillermo	El lugar de la sintaxis en las primeras gramát...	2001	Primera edición	No ficción	Escrito	Libro	Ciencias sociales, creencias y pensamiento	España	NaN	Académico	El texto contiene citas ajenas al autor	actualidad. Reduciéndonos al punto concreto qu...

Figura 1. Vista del subcorpus de textos académicos con menciones a autores canónicos creado a partir del CORPES.

La construcción del corpus arrojó resultados preliminares significativos. Para crear este subcorpus realizamos búsquedas por lemas (ej. "sofocl*" por "Sófocles") para incluir adjetivos derivados. La exploración del corpus en inglés arrojó menos ejemplos que el CORPES. El español presenta mayor tendencia derivativa: "tragedia sofoclea", "humor aristofánico", "orientación hipocrático-galénica". En ocasiones se requerían búsquedas múltiples para capturar variaciones consonánticas en los lemas, como "Séneca", que genera formas con "seneq*" (senequismo, senequista). La prevalencia de estas derivaciones es notable: "Aristóteles" aparece 788 veces y sus adjetivos "aristotélic[oa]s" 284 veces (26,5% del total). Esta característica puede atribuirse al registro más literario del español académico frente al inglés, evidenciado también en expresiones como "digno de Homero" o "Homero, el gran rapsoda".

La desambiguación presenta características distintivas respecto al inglés debido a la mayor prevalencia de nombres clásicos en América Latina contemporánea, como "Plinio", "Apuleyo" o "Juan Crisóstomo". Estos casos resultan relativamente sencillos de resolver al ir acompañados de apellidos. La complejidad surge con nombres como "Agustín", donde se priorizan sistemáticamente las referencias al autor y su obra sobre las menciones toponímicas o hagiográficas, lo cual requiere lectura cercana.

De los primeros 100 autores según APh, 21 no registraron ocurrencias.

[1] https://github.com/luisarip/DH2025/blob/main/MECANO_authors.csv
[2] https://github.com/luisarip/HDH2025/blob/main/CORPES_subcorpus.csv

Esta ausencia obedece a autores con producción reducida (Tibulo, Filodemo) o de interés específicamente filológico (historiadores como Casio Dion, o autoría dudosa como el anónimo de los Himnos Homéricos). Esto confirma que el subcorpus abarca disciplinas humanísticas más amplias que la filología clásica.

Resultados

El CORPES registra 3717 menciones a 88 autores canónicos en 368 documentos, con picos en 2003 y 2013 (Fig. 2). La distribución por autor (Fig. 3) revela una estructura jerárquica donde ciertos autores canónicos alcanzan niveles de reconocimiento significativamente superiores, mientras la mayoría permanece en rangos de atención considerablemente menores, confirmando la presencia de un "hipercanon" en el sentido de Damrosch (2006). Los textos provienen mayoritariamente de España (Fig. 4). Estudiando el ranking de autores por países, se observa que son homogéneos en todo el mundo hispanohablante, con la clara primacía hipercanónica de Platón y Aristóteles.

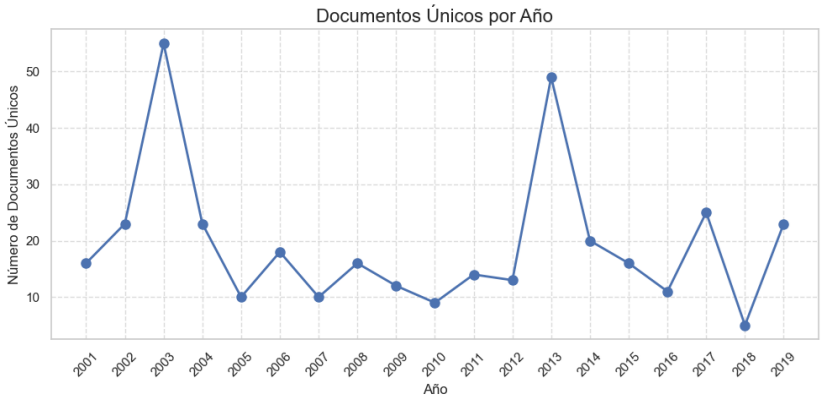


Figura 2. Distribución temporal de documentos con al menos una mención a autores canónicos.

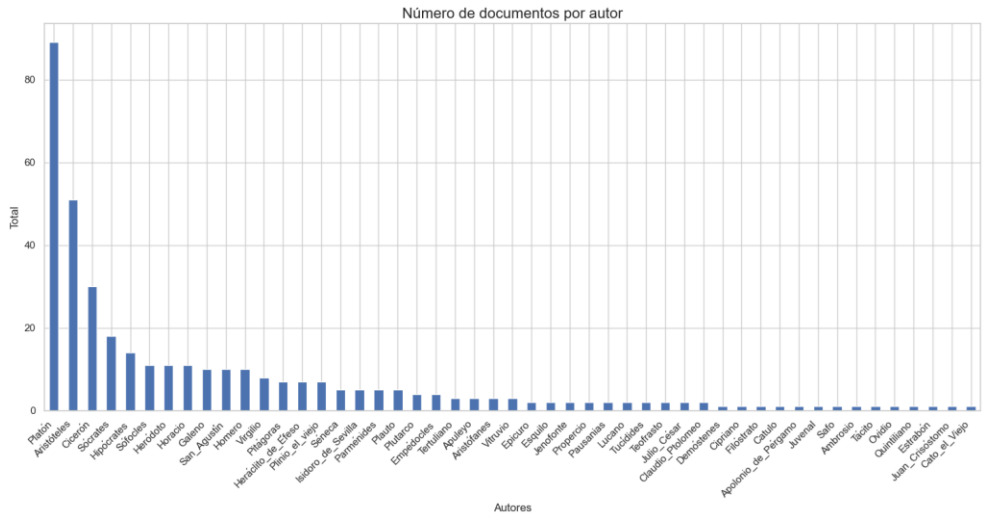


Figura 3. Número total de documentos con al menos una referencia al autor.

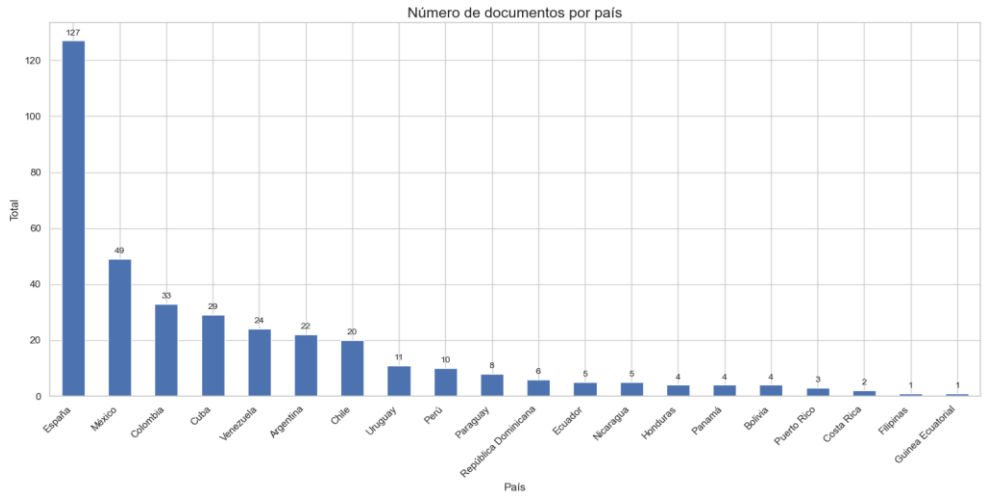


Figura 4. Distribución geográfica de los documentos del subcorpus.

A continuación, comparamos los rankings de los 20 autores más presentes en cada corpus (Fig. 5). Esta comparación es pertinente al contrastar corpus académicos en diferentes idiomas con uno disciplinar que refleja exclusivamente intereses de la comunidad de clasicistas. Con tal de obtener una medida más objetiva, calculamos el coeficiente de correlación de rango Spearman. Mientras que la comparación entre los corpus de JSTOR y APh es significativa (Ripoll-Alberola et al.), no lo son las comparaciones con el CORPES, por lo que no podemos aceptar que exista correlación. Esto obedece a la abundante presencia de autores únicos en el CORPES. La no-correlación evidencia que la canonización intradisciplinar (representada por APh) es un fenómeno diferenciado. La diferencia entre inglés y español también afecta a la presencia de autores canónicos.

Sócrates y Plinio el Viejo están presentes en JSTOR y CORPES, pero no en APh. Plauto es compartido por CORPES y APh, pero no por JSTOR. San Agustín aparece más elevado en comparación con JSTOR, posiblemente reflejando la mayor presencia cultural del catolicismo en el mundo hispanohablante. Aparece en el CORPES otro autor del inicio del cristianismo, Isidoro de Sevilla, cuyo análisis contextual revela su abordaje desde perspectivas médicas y científico-enciclopédicas. El papel destacado de la medicina en el corpus, con Hipócrates y Galeno en posiciones elevadas, puede deberse tanto a un sesgo específico del CORPES —que incluye numerosos textos de Pedro García Barreno, médico y miembro de la RAE— como a la significativa presencia de humanidades médicas (Macías Llanes, 2011), antropología médica (Martínez-Hernáez et al., 2015) y bioética (Álvarez-Díaz, 2012) en la academia hispanohablante, campos propensos a citar autores canónicos. Por otro lado, la aparición de Heráclito y Parménides refleja la influencia presocrática en la filosofía contemporánea hispana.

El desplazamiento de Homero resulta notable dado su estatus fundacional en la tradición grecolatina. Ello podría explicarse por la menor intensidad de la

actividad filológica en Iberoamérica comparada con otros contextos europeos (Italia, Alemania…). Las aproximaciones a estos autores en el ámbito hispanohablante proceden antes de la filosofía que de la historia antigua o la filología clásica.

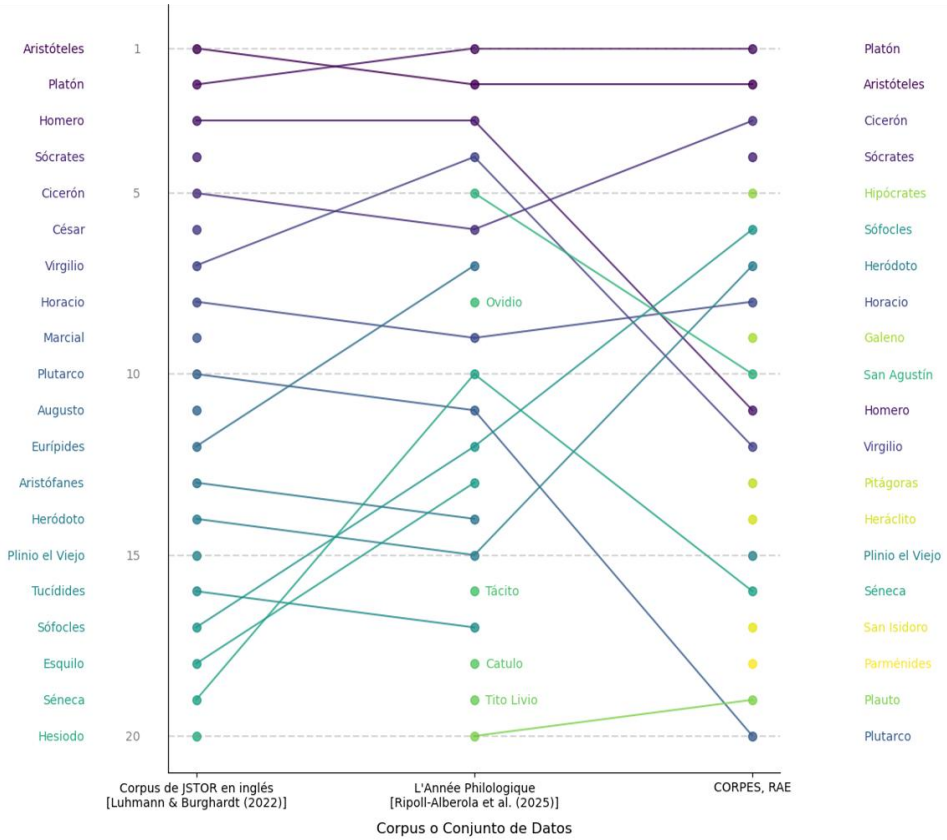


Figura 5. Comparación del ranking del top 20 autores canónicos por corpus.

Para obtener más información sobre en qué contextos se citan a autores canónicos, efectuamos modelado de temas (*topic modeling*, en inglés). Para determinar el número óptimo de temas usamos las visualizaciones del paquete ldatuning (Murzintcev, 2016; Fig. 6) y optamos por 24 temas.

De los 24 temas, descartamos 6 por ser temas basura. La selección se realiza por lectura atenta de las palabras más presentes en cada tópico, verificándose que generalmente los temas basuras tienen bajo valor de coherencia (Fig. 7). Dado que el vocabulario filosófico contiene palabras abstractas susceptibles de considerarse vacías (*stopwords*), optamos por etiquetar la mayor cantidad de temas. Esto genera diferencias cualitativas entre los temas seleccionados (por ejemplo, comprobando las Figs. 7 y 8, el tema 24, Sócrates_2, de baja prevalencia y baja coherencia, es poco significativo comparado con el tema 11, Tragedia, de mayor prevalencia y con un vocabulario muy diferenciado).

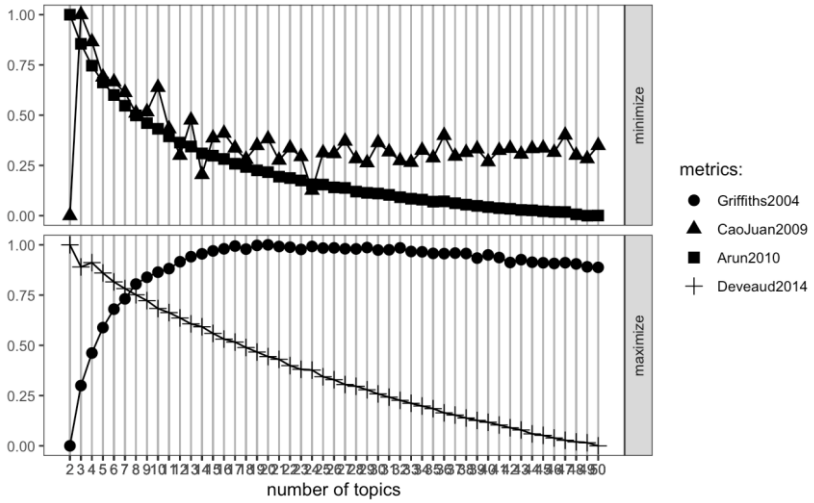


Figura 6. Métricas para hallar el número óptimo de temas.

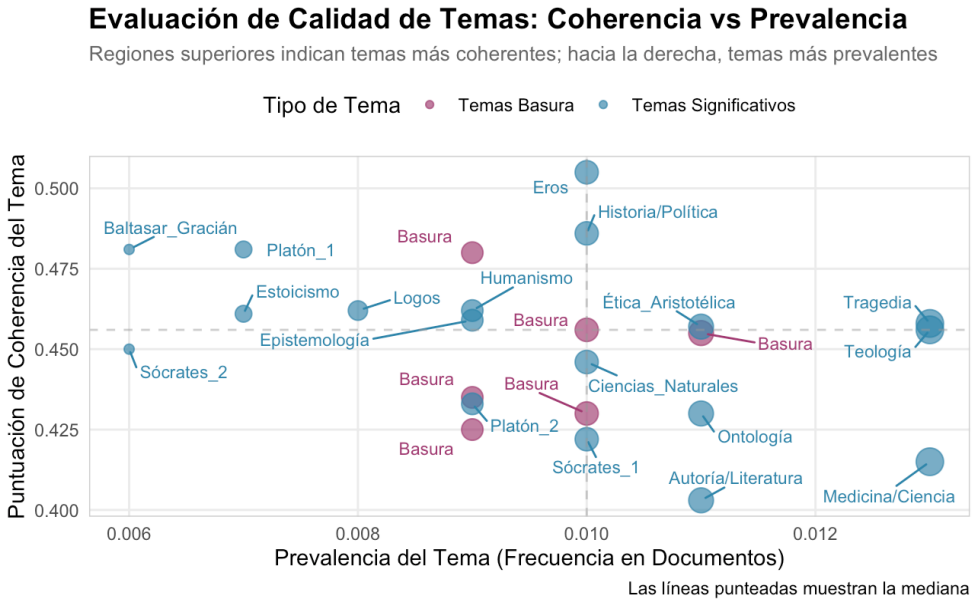


Figura 7. Coherencia y prevalencia de temas.

	topic_name	coherence_score	prevalence	top_words
Topic 1	Basura	0.480	0.009	sino, puede, sólo, texto, posible, momento, duda, aq...
Topic 2	Basura	0.456	0.010	mismo, tiempo, sentido, concepto, sociedad, precisa...
Topic 3	Platón_1	0.481	0.007	vez, platón, tal, lugar, discurso, hace, joven, tan, man...
Topic 4	Autoría/Literatura	0.403	0.011	siempre, autores, incluso, obras, cómo, aunque, part...
Topic 5	Basura	0.430	0.010	gran, poder, nuevo, embargo, aunque, libro, modo, n...
Topic 6	Basura	0.455	0.011	ser, decir, parménides, tan, filósofo, heráclito, parece...
Topic 7	Ética_Aristotélica	0.457	0.011	bien, aristóteles, ética, moral, política, virtud, justicia,...
Topic 8	Epistemología	0.459	0.009	conocimiento, teoría, idea, natural, experiencia, prim...
Topic 9	Medicina/Ciencia	0.415	0.013	siglo, medicina, galeno, hipócrates, primera, partir, te...
Topic 10	Teología	0.456	0.013	así, dios, agustín, san, ciudad, caso, propio, orden, ju...
Topic 11	Tragedia	0.458	0.013	cultura, arte, teatro, esquilo, homero, eurípides, trage...
Topic 12	Logos	0.462	0.008	lenguaje, misma, palabras, diálogo, platónico, diálog...
Topic 13	Ontología	0.430	0.011	mundo, realidad, cosas, sentido, real, ideal, lado, mie...
Topic 14	Basura	0.425	0.009	dos, saber, época, relación, ejemplo, cuanto, siglos, ...
Topic 15	Baltasar_Gracián	0.481	0.006	obra, gracián, además, obras, tradición, autor, virgilio...
Topic 16	Basura	0.435	0.009	aristóteles, años, tres, primeros, escritos, hecho, tale...
Topic 17	Humanismo	0.462	0.009	vida, hombre, naturaleza, social, humana, libertad, h...
Topic 18	Platón_2	0.433	0.009	platón, filosofía, pensamiento, platónica, ideas, cierto...
Topic 19	Sócrates_1	0.422	0.010	sócrates, griego, embargo, carácter, tradición, verdad...
Topic 20	Estoicismo	0.461	0.007	según, alma, cicerón, cuerpo, parte, fin, mediante, pl...
Topic 21	Ciencias_Naturales	0.446	0.010	dice, pitágoras, universo, nueva, tierra, música, movi...
Topic 22	Eros	0.505	0.010	amor, hombres, mujeres, mujer, solo, luego, mito, im...
Topic 23	Historia/Política	0.486	0.010	forma, pues, historia, cada, ello, término, presente, p...
Topic 24	Sócrates_2	0.450	0.006	platón, sócrates, muchas, aún, veces, muerte, conoce...

Figura 8. Vista de tabla de los temas.

En la Fig. 9 se prueba nuestra anterior hipótesis: en la academia hispanohablante, los autores clásicos se estudian predominantemente desde la filosofía. Temas como Epistemología (8) o Estoicismo[3] (20) están presentes junto a casi todos los autores principales. Otros temas son Medicina, Tragedia y Teología, asociados a Hipócrates, Sófocles y San Agustín respectivamente. El tema Baltasar Gracián se debe a que el corpus contiene muchos textos sobre este autor.

[3]Podría parecer que no tiene sentido que Estoicismo esté presente en autores alejados de esta doctrina, pero las palabras “cuerpo” y “enfermedad” tienen presencia en textos médicos, y Platón y Aristóteles también se ocupan de las “pasiones”.

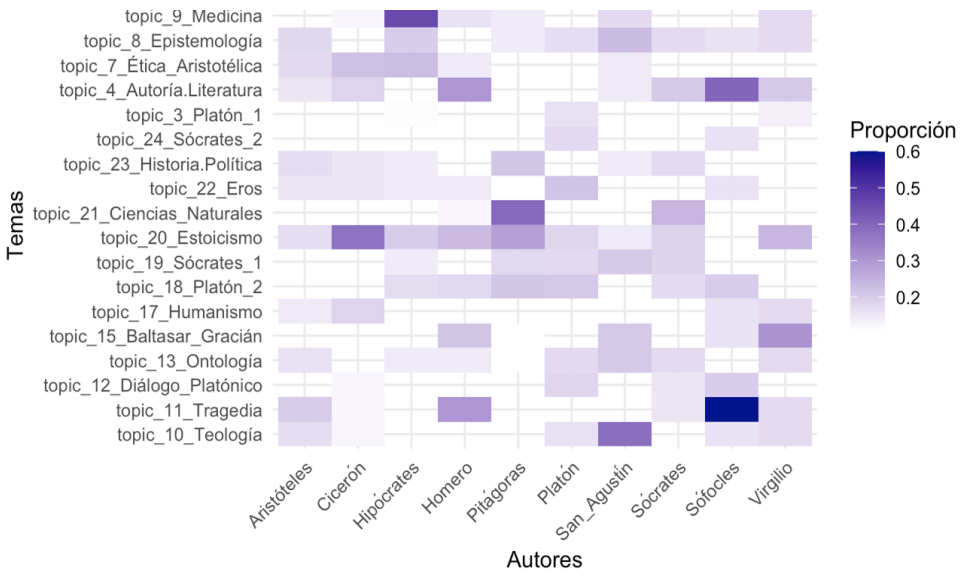


Figura 9. Relación entre temas y autores.

El análisis temático de Aristóteles y Platón (los autores más mencionados) confirma la coherencia de los temas y la prevalencia de la filosofía (Fig. 10).

Autor	Tema	Documentos	Porcentaje (%)
Aristóteles (166 documentos)			
Aristóteles	Tragedia	34	20.5
Aristóteles	Epistemología	30	18.1
Aristóteles	Ética Aristotélica	29	17.5
Aristóteles	Teología	28	16.9
Aristóteles	Estoicismo	28	16.9
Platón (130 documentos)			
Platón	Eros	28	21.5
Platón	Platón_2	27	20.8
Platón	Diálogo Platónico	24	18.5
Platón	Estoicismo	24	18.5
Platón	Ontología	23	17.7

Figura 10. Vista de tabla de temas co-ocurrentes con Aristóteles y Platón.

Finalmente, estudiamos la información mutua puntual (*Pointwise Mutual Information*, PMI) entre todas las columnas binarias del conjunto de datos (pares autor-autor, tema-tema y autor-tema). Las coocurrencias son más probables entre autores que entre autor-tema, sugiriendo que los procesos de canonización operan principalmente a través de redes intertextuales entre autores antes que por adhesión a marcos temáticos específicos. Simultáneamente, dado que la información mutua es consistentemente baja (<0.1, cuando es un valor entre 0 y 1, Fig. 11), este resultado sugiere que las referencias canónicas emergen de la relevancia intrínseca del autor y la obra antes que por mediación de o asociación

con otras referencias canónicas.

	pair	mutual_information
2327	mi_Esquilo_Eurípides	0.08304828
2533	mi_Eurípides_Sófocles	0.06859482
2376	mi_Esquilo_Sófocles	0.06339832
2868	mi_Galeno_Hipócrates	0.05353590
2498	mi_Eurípides_Jenofonte	0.04143559
1139	mi_Cato_el_Viejo_Marcial	0.03818232
3268	mi_Homero_Virgilio	0.03795050
3527	mi_Jenofonte_Tito_Livio	0.03719990
3740	mi_Juvenal_Marcial	0.03602518
4089	mi_Menandro_Plauto	0.03602518

Figura 11. Valores más altos de PMI.

Conclusiones

El estudio del CORPES arroja las siguientes conclusiones. La academia hispanohablante presenta un hipercanon similar al inglés: Platón y Aristóteles dominan ampliamente las menciones. No existe correlación clara entre los rankings español (CORPES), inglés (JSTOR) y bibliografía especializada (APh), sugiriendo procesos de canonización diferenciados según contexto lingüístico y disciplinario. La especificidad cultural hispanohablante se manifiesta en San Agustín (presencia católica), Galeno e Hipócrates (influencia médica) y Parménides y Heráclito (influencia de la filosofía presocrática). El desplazamiento de Homero del hipercanon evidencia que las aproximaciones a autores clásicos provienen principalmente de la filosofía, no de la filología clásica o historia antigua.

El modelado temático confirma este hallazgo: temas como Epistemología y Estoicismo aparecen transversalmente. Los análisis de información mutua sugieren que la canonización opera desde la autoridad aislada del autor o la obra, aunque las coocurrencias más persistentes son inter-autorales, no temáticas.

El proyecto MECANO continúa desarrollándose. Futuramente incorporaremos visualizaciones y análisis de redes para maximizar la información extraída del modelado temático.

Bibliografía

ÁLVAREZ-DÍAZ, J. A. (2012), «¿Bioética latinoamericana o bioética en Latinoamérica?», *Revista Latinoamericana de Bioética*, 12(1), pp. 10–27.

BERTI, M. (2024), «Digital Practice for Studying the Indirect Transmission of Classical Authors and Works», en F. FAVI & V. MASTELLARI (eds.), *Treasuries of Literature: Anthologies, Lexica, Scholia and the Indirect*

Tradition of Classical Texts in the Greek World, Berlin, pp. 189–206. https://doi.org/10.1515/9783111386010-010

DAMROSCH, D. (2006), «World Literature in a Postcanonical, Hypercanonical Age», en H. SAUSSY (ed.), *Comparative Literature in an Age of Globalization*, Johns Hopkins University Press, pp. 43–53.

LUHMANN, J., & BURGHARDT, M. (2022), «Digital humanities—A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape», *Journal of the Association for Information Science and Technology*, 73(2), pp. 148–171. https://doi.org/10.1002/asi.24533

MACÍAS LLANES, M. E. (2011), «Ciencias Sociales y Humanísticas en la formación médica», *Humanidades Médicas*, 11(1), pp. 18–44.

MARTÍNEZ-HERNÁEZ, Á.; GIL, E. P.; & COMELLES, J. M. (2015), «Genealogía de la Antropología Médica en España», *Disparidades. Revista de Antropología*, 70(1), Article 1. https://doi.org/10.3989/rdtp.2015.01.010

MURZINTCEV, N. (2016), *ldatuning: Tuning of the latent dirichlet allocation models parameters*. R package, v. 0.2.0.

PFEIFFER, R. (1968), *History of Classical Scholarship from the Beginnings to the End of the Hellenistic Age*, Oxford, Clarendon Press.

ROMANELLO, M. (2013), «Creating an Annotated Corpus for Extracting Canonical Citations from Classics-Related Texts by Using Active Annotation», en A. GELBUKH (ed.), *Computational Linguistics and Intelligent Text Processing. CICLing 2013*, Lecture Notes in Computer Science, 7816, Berlin, Springer, pp. 48–59. https://doi.org/10.1007/978-3-642-37247-6_6

RIPOLL-ALBEROLA, L.; D'ADDARIO, L.; BURGHARDT, M.; BERTI, M.; & DEPAUW, M. (2025), «Tracing Antiquity: References to Greco-Roman Authors in Modern Academic Discourse», en *Digital Humanities 2025: Book of Abstracts* (en prensa).

Cartografiar el infinito: una lectura interactiva de *El infinito en un junco* (2019) de Irene Vallejo a través de ArcGIS StoryMaps

Soledad Castaño Santos
Universidad de Basilea

Palabras clave: Visualización de datos; ArcGIS; StoryMaps; Irene Vallejo; cartografía literaria
Keywords: Data visualization; ArcGIS; StoryMaps; Irene Vallejo; literary cartography

La presente comunicación propone una lectura digital e interactiva de *El infinito en un junco* (2019) de Irene Vallejo. Este ensayo se ha convertido en una obra reconocida en el panorama literario internacional por su capacidad de interrelacionar la historia del libro en la Antigüedad con una suerte de autobiografía con auténtica maestría. El éxito editorial se confirma con múltiples reediciones en español, su traducción a más de cuarenta idiomas y su adaptación a otros formatos literarios como la novela gráfica en 2023. Pese a que esta investigación ha suscitado interés entre la crítica literaria y académica, son escasos los estudios de este ensayo que utilicen una perspectiva metodológica que combine el ámbito literario y las nuevas herramientas tecnológicas. De este modo, el presente proyecto pretende explorar el texto vallejiano desde las herramientas de geolocalización como ArcGIS y de visualización de datos como StoryMaps a través de la herramienta ArcGIS Online.

Nuestra propuesta surge con dos problemáticas esenciales: el género escogido y la transformación de la información literaria en datos cuantitativos. El ensayo de Vallejo abarca en su vasta extensión una gran diversidad de geografías y cronologías, lo que complica la visualización de datos conjunta. Es cierto que el lector advierte en su narración la historia del libro acompañada de lugares, eventos históricos y personajes históricos que contribuyeron a la preservación, persecución o reproducción de este valioso objeto material. Lugares que van desde Alejandría a la Atenas de Pericles, y de la espléndida ciudad helénica a la Roma Imperial. Además, la autora combina estos lugares históricos con geografías contemporáneas que fueron relevantes para la historia material del libro, por ejemplo, la Biblioteca de Sarajevo. De este modo, Irene Vallejo traza una red invisible de lugares importantes que configuran un mapa cultural disperso entre las páginas de su ensayo.

La nómina extensa de lugares, personajes y obras literarias recogidas en la obra requirieron de una clasificación previa y un modelo de datos para su posterior visualización. La heterogeneidad de las entidades geográficas, literarias y personales promovieron el análisis de los datos y la distribución de los mismos en diferentes tablas según su tipología: lugares, personas y obras. Esta tripartición, a su vez, genera subtipos, que permiten concretar significativamente el dato. En el caso de los lugares, obtenemos cinco categorías: ciudades o islas, estado o país, continente, ficticios y monumentos o ríos o mares. No obstante, en la entidad perteneciente a las personas solo diferenciamos entre personajes históricos o personajes mitológicos. Y por último, decidimos no establecer subtipos dentro de las obras literarias referenciadas. Este paso no solo exige una recogida de datos desde la obra literaria de Vallejo, sino también una interpretación de los mismos y asignar atributos a cada uno de ellos para producir nuevo conocimiento —latitud, longitud, etapa cronológica, frecuencia de aparición etc.

ENTIDADES LUGARES CIUDADES O ISLAS								
LUGAR	CIUDAD	PAÍS/ ESTADO	FRECUENCIA	VARIANTES	ÉPOCA	OBSERVACIONES	LATITUD	LONGITUD
Roma	Roma	Imperio Romano	79	Urbe, Ciudad Eterna, la ciudad de las 29 bibliotecas	Antigüedad		41.8933203	12.4829321
Atenas	Atenas	Estado de Atenas	54	La ciudad	Antigüedad		37.9755648	23.7348324
Alejandría	Alejandría	Reino de Egipto	42	La ciudad	Antigüedad		31.1991806	29.8951716

ENTIDADES PERSONAS								
NOMBRE	GÉNERO	FRECUENCIA	ÉPOCA	LUGAR DE NACIMIENTO	OCUPACIÓN	OBSERVACIONES	LATITUD	LONGITUD
Platón	Masculino	31	Antigüedad	Atenas	Filósofo		37.9755648	23.7348324

Tablas 1 y 2. Muestra de entidades pertenecientes al corpus extraídas de *El infinito en un junco*.

El enfoque metodológico del estudio se basa en la concepción del *distant reading* de Franco Moretti (2005, 2013), que propone un giro metodológico a la hora de estudiar los textos literarios desde una perspectiva cuantitativa, espacial y relacional. Gracias a este nuevo enfoque, detectamos patrones, redes y relaciones invisibles a simple vista y, especialmente, en textos extensos. El principal reto reside en la capacidad de adecuación de esta metodología al género ensayístico de Vallejo y transformar la información en conocimiento significativo e interactivo sin reducir su riqueza cultural.

Asimismo, incorporamos a nuestra investigación el concepto de cartografía literaria en el que se investiga desde un planteamiento interdisciplinar la disciplina geográfica y el análisis espacial en la literatura planteado por Piatti y Hurni (2011) y Piatti (2016). Esta concepción considera el espacio literario como construcciones narrativas cuantificables y susceptibles de interpretaciones subjetivas. En este sentido, el uso de herramientas digitales como ArcGIS StoryMaps no solo facilita la representación espacial de las entidades, sino también, vincularlos a otras características que permitan una lectura avanzada de los datos. En síntesis, la metodología empleada combina las teorías del

distant reading, las técnicas del análisis espacial y las herramientas de geovisualización.

La primera tarea de la investigación y una de las más complejas fue la selección y creación del corpus textual. Este conjunto de datos está distribuido en las tres categorías ya mencionadas: lugares, personas y obras literarias. Para conseguirlo empleamos herramientas de NLP (*Natural Language Processing*) para la extracción de entidades propias. En este caso, aplicamos la librería *SpaCy* en un *script* de Python en el que realizamos una primera división, que fue revisada de forma manual exhaustivamente. Las divisiones y caracterización de cada una de las entidades siguen los estudios de la cartografía literaria y facilitan la preservación de la riqueza narrativa de las entidades que emplea la autora.

La georreferenciación de los lugares se realizó mediante herramientas propias de QGIS con el establecimiento de coordenadas y otras fuentes fiables para la adjudicación de coordenadas en lugares históricos (GeoNames y Pleiades). Ante las complejidades de los espacios simbólicos, se adjudica una referencia espacial justificada en el campo de observaciones. Tras este proceso, las entidades se integraron en la herramienta QGIS para visualizar las entidades en el mapa con diferentes capas.

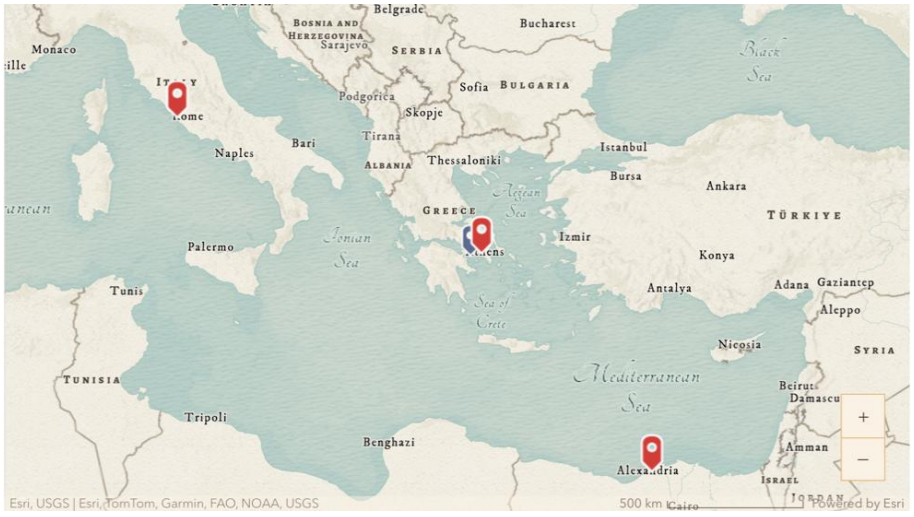


Figura 1. Prototipo de mapa de entidades en ArcGIS Online.

El siguiente proceso es la integración de los mapas con la información codificada a la herramienta de visualización ArcGIS StoryMaps, en la que creamos una narrativa digital sobre Irene Vallejo y su obra con una interfaz interactiva y atractiva para el usuario. La elección de esta herramienta para nuestro trabajo es deuda del proyecto europeo coordinado por María Luisa de Lázaro Torres (2021, 2023) sobre la biografía de autoras europeas en ArcGIS StoryMaps y la aplicación didáctica de estas herramientas digitales en su empleo en distintos niveles educativos.

Las metodologías implicadas en nuestro proyecto favorecen la representación innovadora de los datos y generan una lectura interactiva que aúna los saberes filológicos con el análisis espacial y la narrativa digital. Este

enfoque pretende ofrecer una nueva experiencia al lector y facilitar nuevas vías de investigación.

El resultado final es un StoryMap interactivo que describe las entidades incluidas en *El infinito en un junco* de un modo enriquecido a través de recursos visuales como los mapas interactivos. De esta visualización, obtenemos resultados significativos desde las perspectivas espacial y literaria. En primer lugar, la visualización espacial facilita la detección de los focos geográficos recurrentes como nodos centrales de la narración, gracias a la frecuencia de aparición, como es el caso de Alejandría, Atenas o Roma.

A nivel técnico, el proyecto demuestra la viabilidad de aplicar herramientas de georreferenciación al género ensayístico, pocas veces elegido en los trabajos del ámbito de las Humanidades Digitales, pero también revela sus retos y particularidades a los que el investigador debe atenerse. El funcionamiento de ArcGIS StoryMaps junto con el corpus textual es fundamental para preservar el contenido y a su vez proporcionar al lector una visualización atractiva e interactiva. Los usuarios pueden desplazarse por el mapa, escoger las capas de interés —lugares, personajes, obras—, consultar características y explorar patrones de similitud entre las entidades.

El StoryMap constituye, en nuestro proyecto, una herramienta de análisis, pero sobre todo, un recurso valioso desde el punto de vista pedagógico. Facilita la comprensión de la obra literaria por parte de un público general, y al tiempo, ofrece material de investigación a nivel espacial y literario para los especialistas. En este sentido, el estudio divulga de forma didáctica los datos en una herramienta como ArcGIS Online, que facilita el acceso abierto, siguiendo los principios FAIR. Asimismo, el estudio constituye una narrativa digital propia sobre la obra literaria que puede ser transformado en instrumento de estudio en el ámbito educativo. En el marco de la investigación, esta propuesta suscita nuevas formas de lectura de los datos espaciales, la interpretación de estos desde un enfoque interdisciplinar y promover nuevas investigaciones que sigan el mismo marco metodológico.

Este proyecto constata la potencialidad de las herramientas digitales para enriquecer las investigaciones críticas de obras literarias contemporáneas, como es el caso de *El infinito en un junco*. La visualización espacial no solo desvela patrones espaciales en la narración, sino que transforma la experiencia del usuario en una herramienta de interacción dinámica. La combinación de distintas metodologías —lectura distante, cartografía digital, utilización de GIS— ofrecen nuevas perspectivas en la investigación, en la docencia y en la difusión cultural.

Bibliografía

ESRI (s.f.a), *ArcGIS Online* [plataforma SIG en línea], recurso en línea: https://www.arcgis.com [Acceso: 14/10/2025].

ESRI (s.f.b), *ArcGIS StoryMaps* [herramienta de narrativa geográfica], recurso en línea: https://storymaps.arcgis.com [Acceso: 14/10/2025].

LÁZARO Y TORRES, M. L. DE; & MORALES YAGO, F. J. (2021), «Educar para el desarrollo sostenible empleando Story Maps creados en un MOOC», en L. VEGA CARO & A. VICO-BOSCH (coords.), *Igualdad y calidad educativa: oportunidades y desafíos de la enseñanza*, Barcelona, Octaedro, pp. 1696–1718.

LÁZARO Y TORRES, M. L. DE; BUZO SÁNCHEZ, I.; & DE MIGUEL GONZÁLEZ, R. (2023), «Cartoteca digital de biografías de literatos europeos en Story Maps», *Boletín de la Real Sociedad Geográfica*, 161, pp. 159–168.

MORETTI, F. (2005), *Graphs, Maps, Trees: Abstract Models for Literary History*, London, Verso.

MORETTI, F. (2013), *Distant Reading*, London, Verso.

PIATTI, B. (2016), «Mapping fiction: The theories, tools and potentials of literary cartography», en D. COOPER; C. DONALDSON; & P. MURRIETA-FLORES (eds.), *Literary Mapping in the Digital Age*, Abingdon/New York, Taylor & Francis, pp. 177–192.

PIATTI, B.; & HURNI, L. (2011), «Cartographies of fictional worlds», *The Cartographic Journal*, 48(4), pp. 218–223. https://doi.org/10.1179/174327711X13190991350051

QGIS.ORG (2024), *QGIS Geographic Information System* (v. 3.34) [software de código abierto], recurso en línea: https://qgis.org [Acceso: 14/10/2025].

VALLEJO, I. (2019), *El infinito en un junco: La invención de los libros en el mundo antiguo*, Madrid, Siruela.

Construcción del relato turístico durante el desarrollismo: un estudio de textos oficiales sobre Zaragoza ORTS con Voyant Tools

Andrea Todolí Moreno
Archivo de la Diputación de Alicante
Ángela Celis Sánchez
Universidad de Castilla-La Mancha

Palabras clave: Lectura Activa; Prompting; Chain-of-Thought; Marcación de textos; Desambiguación
Keywords: Active Reading; Prompting; Chain of Thought; Text Markup; Disambiguation

Zaragoza Orts y la construcción del Benidorm turístico: una historia a revisar desde las fuentes oficiales

Este trabajo aborda el desarrollo turístico de España, en concreto, en la ciudad de Benidorm, durante las décadas de 1950 y 1960 mediante el análisis de ciertos textos oficiales con la herramienta Voyant Tools. El objetivo es entender cómo las políticas públicas y las estrategias de promoción turística transformaron un municipio de carácter tradicional y pesquero en una insignia del turismo de masas, proceso que tuvo lugar en gran medida durante el período del desarrollismo español. La investigación se centra específicamente en el papel del alcalde Pedro Zaragoza Orts, quien durante su mandato de 1950 a 1966 implementó una serie de acciones y políticas destinadas a impulsar la notoriedad, urbanización y actividad turística del municipio.

El contexto histórico proporciona un fondo necesario para contextualizar la transformación. España, tras la Guerra Civil y en su fase de aislamiento internacional, experimentaba cambios políticos y económicos que facilitaban la apertura al turismo como motor clave de modernización. La manera en que esto sucedió, a partir de la visión y las políticas de Zaragoza Orts, y cómo esa narrativa se construyó en los discursos institucionales y mediáticos, constituyen un objeto de análisis fundamental.

Nuestra investigación se centra en comprender cómo los discursos generados en los textos oficiales, en concreto en el Boletín del Ayuntamiento de Benidorm, reflejaban y contribuyeron a la construcción de la imagen y percepción social del turismo en la ciudad. La relevancia reside en dilucidar las relaciones entre política, promoción, imagen y desarrollo económico-social, así como en evidenciar el impacto de estas políticas en la percepción pública de Benidorm

como destino turístico.

El trabajo se ha realizado aplicando la minería de textos con Voyant Tools, que permite analizar grandes volúmenes de datos textuales —que un análisis tradicional sería difícil de procesar— para extraer patrones y tendencias. El núcleo del trabajo consiste en demostrar que estas herramientas digitales pueden facilitar una comprensión más profunda y objetiva sobre la evolución discursiva en torno a las políticas turísticas, además de reflexionar sobre la medida en que estas políticas influyeron en la percepción y realidad del fenómeno turístico en Benidorm y en España en general.

Este problema de investigación, por tanto, busca llenar un vacío en los estudios de historia social y turística, integrando el análisis cualitativo y cuantitativo mediante técnicas de las humanidades digitales para comprender la narrativa oficial e institucional que ayudó a consolidar el turismo como una industria clave en la economía española, en un momento crucial de su historia.

Metodología

La metodología de esta investigación combina métodos tradicionales de análisis humanístico con técnicas avanzadas de minería de textos para entender en profundidad cómo se construyó discursivamente la política turística en Benidorm durante el período especificado. La estrategia metodológica puede dividirse en varias fases clave:

a) **Selección y construcción del corpus.** El primer paso consistió en la recopilación y selección de los textos relevantes. Para ello, se accedió a los archivos del Boletín del Ayuntamiento de Benidorm, que fue el órgano oficial de comunicación institucional durante la época de interés, aún disponible en la versión digitalizada en el Archivo Municipal. Se eligieron los números publicados entre noviembre de 1959 —fecha de inicio del análisis— y diciembre de 1966, coincidiendo con el período en que Zaragoza Orts gobernó y en el que se gestó la transformación turísticas y urbanísticas.

   Esta elección se basó en criterios de relevancia temática y temporalidad, asegurando un corpus representativo que reflejara las estrategias y discursos oficiales en relación con el turismo. La selección excluyó artículos no relacionados directamente con la promoción, urbanización, eventos turísticos o la percepción pública, logrando así una base de aproximadamente 160 artículos. Esta muestra proporcionó la suficiente variedad y volumen para un análisis riguroso, pero sin perder la especificidad temática.

b) **Preparación del corpus y herramientas digitales.** El siguiente paso fue digitalizar los documentos en formato adecuado para procesamiento con Voyant Tools, una plataforma web de análisis de textos que facilita la visualización y extracción de patrones. La limpieza del corpus (quitar errores de digitalización, normalización del texto, etc.)

fue realizada previamente para asegurar la calidad de los datos.

El uso de Voyant Tools permite explorar aspectos como la frecuencia de palabras, patrones de coocurrencia, tendencias temporales y relaciones semánticas, presentadas en visualizaciones interactivas. Además, se complementó el análisis con herramientas específicas, como nubes de palabras, diagramas de tendencias, gráficos de barras y tablas de concordancias, que facilitaron la interpretación de los datos y la detección de temas recurrentes y patrones discursivos.

c) **Análisis cuantitativo.** Este paso involucró técnicas de minería de textos para cuantificar y visualizar elementos de los discursos. Por ejemplo, se procedió a realizar análisis de frecuencia para identificar las palabras más utilizadas y su evolución a lo largo del tiempo. Se centró en términos clave como "Benidorm", "turismo", "España", "población" y otros relacionados, que reflejaban los temas principales.

También se analizaron relaciones semánticas mediante mapas de coocurrencia para entender qué términos estaban relacionados en los textos, lo cual ayudó a identificar conceptos vinculados, como "urbanización", "fama internacional", "eventos", "promoción", entre otros. Las tendencias temporales permitieron visualizar cómo evolucionaron estas ideas y cuáles fueron los focos del discurso a lo largo de los años, señalando momentos de mayor énfasis o cambio de narrativa.

d) **Análisis cualitativo.** Complementariamente, se realizó un análisis interpretativo del contenido extraído. La lectura en contexto de ciertos extractos, frases y palabras permitió identificar las ideas dominantes, las narrativas de promoción, las relacionadas con la moralidad, el progreso y la modernización, así como los discursos institucionales que buscaban modelar la percepción del pueblo y los visitantes.

Este análisis cualitativo ayudó a entender cómo las élites políticas y mediáticas construían una imagen positiva de Benidorm y cómo movilizaban ciertos discursos que justificaban las políticas urbanísticas y de promoción turística. Además, se interpretaron las visualizaciones para captar la dinámica de cambios discursivos en relación con eventos específicos, como festivales, visitas diplomáticas, campañas o cambios políticos internos.

e) **Validación y cotejo.** Finalmente, se cruzaron los resultados cuantitativos con la interpretación cualitativa para validar las conclusiones. La utilización de múltiples herramientas y fuentes permitió disminuir sesgos y reforzar la robustez del análisis, haciendo plausible la relación entre los patrones en los datos y las políticas públicas y su impacto social.

Resultados obtenidos

La aplicación del enfoque metodológico, combinando minería de textos y análisis humanístico, ha arrojado múltiples hallazgos que contribuyen al entendimiento del proceso de transformación de Benidorm y la narrativa que lo sustentó, destacando principalmente:

f) **Identificación de temas recurrentes y su evolución.** El análisis de frecuencia reveló que los términos "Benidorm" y "turismo" eran protagonistas en todos los textos, señalando una marcada centralidad del fenómeno turístico en la agenda institucional. La palabra "Benidorm" aparecía reiteradamente, llegando a ser la más frecuente, lo cual indica su papel como símbolo de identidad y marca de destino turístico. La presencia constante de "turismo" confirmaba que el principal eje discursivo giraba en torno a la industria turística, vinculada con la economía local, la urbanización y el posicionamiento internacional.

g) **Construcción de la identidad de Benidorm.** El análisis mostró que en los textos se utilizaban palabras relacionadas con la fama, la proyección internacional y el ocio, lo que contribuyó a construir una imagen de Benidorm como un destino atractivo y moderno. La presencia de términos como "fama", "internacional", "festival" y "promoción" evidencia una estrategia discursiva orientada a posicionar a la ciudad en el mapa turístico global.

h) **Impacto de las políticas de Zaragoza Orts.** Los patrones detectados en los datos aportan evidencia del impacto de las políticas del alcalde. La promoción activa de Benidorm como destino turístico se refleja en el aumento de los términos relacionados con campañas, festivales y eventos internacionales. La aparición de palabras referidas a la inversión, urbanización y crecimiento en los textos indica que el discurso oficial articulaba estas acciones como parte de una estrategia integral de desarrollo.

Conclusión

La aplicación de la minería de textos con Voyant Tools ha permitido identificar patrones recurrentes en los discursos del periodo, resaltando temas relacionados con la promoción turística, la modernización de Benidorm, y el impacto social de las políticas de Zaragoza Orts. Se evidencian términos asociados a la identidad de la ciudad, el turismo, la imagen internacional y la percepción social del cambio. Los resultados confirman que las estrategias de Zaragoza Orts tuvieron un impacto positivo en la transformación del municipio en un destino turístico, y que las herramientas digitales facilitaron la objetividad y profundidad del análisis social y político de la época.

Bibliografía

CAMINO PLAZA, L. (2021), «Voyant Tools en su aplicación al estudio de las novelas de Rosalía de Castro (1859-1881)», *Publicaciones de la Asociación Argentina de Humanidades Digitales (PublicAAHD)*, 2, pp. 109–120. https://doi.org/10.24215/27187470e025

CASERO-ALONSO, V.; CELIS, Á.; & LOZANO ZAHONERO, M. (2024), «Minería de textos», en G. FERNÁNDEZ-AVILÉS & J. M. MONTERO (eds.), *Fundamentos de ciencia de datos con R* (1.ª ed.), Madrid, McGraw-Hill Interamericana de España, pp. 653–670, recurso en línea: https://cdr-book.github.io/mineria-textos.html

FRADEJAS RUEDA, J. M. (2019, 4 de abril), «Alfonso X y la estilometría», *7 Partidas Digital*, recurso en línea: https://7partidas.hypotheses.org/1072

GUTIÉRREZ DE LA TORRE, S. (2019), «Análisis de corpus con Voyant Tools», *Programming Historian en español*, 3. https://doi.org/10.46430/phes0043

LESSA, B. (2020), «Análisis con Voyant Tools del primer número de la Revista de Humanidades Digitales», *Publicaciones de la Asociación Argentina de Humanidades Digitales (PublicAAHD)*, 1, pp. 118–121. https://doi.org/10.24215/27187470e012

LÓPEZ FERNÁNDEZ, P. S. (2021), «MORETTI, Franco. *Lectura distante*. Trad. Lilia Mosconi. FCE, Buenos Aires, 2015», *Revista de Humanidades Digitales*, 6, pp. 266–271. https://doi.org/10.5944/rhd.vol.6.2021.27380

SANZ CABRERIZO, A. (2021), «Para unas lecturas remediadas: Análisis cuantitativo y cualitativo de textos», *Revista de Humanidades Digitales*, 6, pp. 122–128. https://doi.org/10.5944/rhd.vol.6.2021.32297

SCHUMANN, F. R. (2022), «Using Voyant Tools for Data Mining Social Media Comments about a Destination: A Guam Study», *Journal of Hospitality & Tourism*, 20(2), pp. 106–118, recurso en línea: https://johat.org/wp-content/uploads/2022/12/6.pdf

Ciencia de datos para las Humanidades: estudio del lenguaje inclusivo con CRISP-DM

Ángela Celis
Gema Fernández-Avilés
Universidad de Castilla La Mancha

¿Por qué la ciencia de datos aplicada a las Humanidades?

Los humanistas trabajamos —también— con datos. A veces, esos datos son tangibles (textos, por ejemplo), a veces, intangibles (movimientos filosóficos, literarios o artísticos), pero datos, al fin y al cabo, que suponen un objeto de estudio que ha de ser abordado para su investigación desde un marco teórico concreto. Para los humanistas, el tratamiento de esos datos ha cambiado sustancialmente desde el advenimiento de las Tecnologías de la Información y la Comunicación (TIC) y, más concretamente, el uso de herramientas estadísticas o computacionales, que inspira lo que algunos autores como Berry (2011) han denominado "giro computacional" en nuestros estudios.

Una de las metodologías de ciencia de datos más utilizada es el CRISP-DM (*Cross-Industry Standard Process for Data Mining*), ampliamente utilizada en la industria para guiar el proceso de análisis de datos. En el contexto de las Humanidades, la aplicación de CRISP-DM puede ser particularmente útil para proyectos de investigación que involucren grandes conjuntos de datos. Por ejemplo, textos históricos u orales con el fin investigar patrones de migración o variaciones lingüísticas regionales o, como en la propuesta que presentamos, analizar el uso del lenguaje inclusivo en diferentes guías de distintas facultades de la Universidad de Castilla- La Mancha (UCLM) sobre la confección de los Trabajos Fin de Grado (TFG).

El presente trabajo tiene como objetivo principal demostrar la escalabilidad de aplicación de la metodología CRISP-DM en los estudios de Humanidades. Para ello, además de enumerar las fases de esta metodología, se irán ilustrando cada una de ellas con las preguntas de investigación humanísticas a propósito del análisis realizado.

CRISP-DM en las Humanidades: una aplicación detallada por fases

Fase I. Entendimiento de la situación: Identificación del problema de investigación

Esta fase inicial se centra en comprender a fondo los objetivos y requisitos del proyecto desde una perspectiva de negocio. En un contexto académico, esto

se traduce en definir el problema central de la investigación, establecer objetivos específicos y evaluar la situación actual para determinar el alcance y las limitaciones del proyecto. La omisión de este paso puede resultar en un considerable esfuerzo dedicado a responder preguntas irrelevantes.

En la investigación humanística, esta fase supone identificar el objeto del análisis de datos textuales, como analizar la frecuencia de palabras o frases en textos literarios o históricos para identificar tendencias y patrones culturales de una época determinada. En el caso que nos ocupa, se trataría de definir la pregunta de investigación, esto es, cómo y en qué medida se emplea el lenguaje inclusivo en la normativa de los TFG de la UCLM. Una vez establecida esta pregunta, pasamos a establecer los objetivos del análisis o, lo que es mismo, detectar patrones de uso, inconsistencias o ausencias de lenguaje inclusivo o evaluar si las normativas reflejan políticas institucionales de igualdad. A continuación, pasaríamos a delimitar el alcance, o circunscribir el análisis, en nuestro caso a documentos varias normativas de TFG de diversas facultades, enmarcándolo en el contexto legal, institucional y académico actual. Posteriormente pasaríamos a justificar la relevancia científica y social del estudio, apuntalándolo en el creciente debate en torno al lenguaje inclusivo y su papel en la producción académica. Además, debemos considerar las posibles limitaciones, por ejemplo, ambigüedades terminológicas, implicaciones ideológicas, marco legal de la Real Academia Española (RAE) o políticas universitarias, así como los costes y beneficios en términos de tiempo, acceso a documentos, impacto institucional y posibles repercusiones para la comunidad universitaria.

Fase II. Comprensión de los datos: Exploración de las Normativas Académicas

Esta fase implica la recolección de datos, el análisis exploratorio de datos (EDA) y la evaluación de la calidad de los datos. Es un paso crucial para formar las primeras percepciones e hipótesis. En el ámbito de las Humanidades, la recolección de datos puede implicar fuentes tradicionales como textos y documentos de bibliotecas y archivos o, cada vez más, el uso de bases de datos y corpus digitales. Para nuestra propuesta, esta fase ha consistido en la selección, localización y descarga de normativas oficiales de TFG en las distintas facultades de la UCLM. Dado que estos textos son de naturaleza normativa, formal y monogestionada, es decir, generados por entidades académicas sin réplica, presentan una estructura relativamente estable que facilita el análisis, aunque pueden variar considerablemente entre facultades.

Fase III. Preparación de los Datos: Selección del corpus normativo para el análisis inclusivo

En el contexto del análisis del uso del lenguaje inclusivo en la normativa de TFG de la UCLM, esta fase representa la transformación y depuración del corpus textual para hacerlo analizable. Para nuestro proyecto, se realizará mediante la

extracción de estos a través de la minería de textos o *text mining*, también conocida como minería de datos textuales. El software que vamos a utilizar para este propósito es R.

Esta etapa es clave, pues garantiza la calidad interpretativa y técnica del análisis. Los objetivos principales pasarían por asegurar textos limpios, normalizados y estructurados para su procesamiento; preparar el corpus para extraer indicadores significativos sobre el uso, o ausencia, de lenguaje inclusivo y reflexionar críticamente sobre cómo las transformaciones afectan la interpretación textual, especialmente en documentos normativos cargados de intención institucional. Ejemplo de tareas:

- Filtrado del corpus. Selección de normativas por facultad, año o versión.
- Conversión de formato. Unificación a formatos compatibles con análisis textual.
- Eliminación de ruido. Limpieza de encabezados, numeraciones irrelevantes, sellos institucionales, firmas.
- Normalización ortográfica. Corrección de errores, uniformidad en el uso de mayúsculas, acentos, etc.
- *Tokenización*. División del texto en palabras, frases o cláusulas. Especial atención a expresiones como "alumno/a", "los/las estudiantes", etc.
- Eliminación selectiva de *stopwords.* Eliminación de términos funcionales irrelevantes, preservando aquellos que puedan tener valor interpretativo.
- Anotación lingüística. Etiquetado de estructuras que evidencien inclusividad o sesgo: uso del masculino genérico, binarismos, fórmulas neutras.
- Codificación de variables. Asignación de categorías a fragmentos textuales: inclusivo, neutro, excluyente, ambiguo.

Para este trabajo hemos de tener en cuenta circunstancias especiales, como el hecho de que la limpieza de estos corpus no siempre es neutra debido a que implica una toma de decisiones a propósito de la eliminación o segmentación de términos, que pueden afectar al lenguaje. En ese sentido, una decisión clave podría se no eliminar palabras funcionales como como "todos", "los", "las", que en otros análisis serían consideradas *stopwords*. En este caso, son elementos centrales para detectar marcadores de género gramatical.

Fase IV. Modelado: Modelos y técnicas aplicables a conjuntos de datos de Humanidades

Esta fase implica aplicar técnicas estadísticas a los datos preparados para desarrollar modelos predictivos o descriptivos. El objetivo es identificar patrones y tendencias, transformando los datos en percepciones significativas o predicciones. En Humanidades, implica aplicar algoritmos para identificar temas, patrones lingüísticos o tonos emocionales en documentos. Técnicas como el

análisis de sentimientos, el modelado de temas y el análisis de redes son particularmente relevantes. Destacamos:

- Análisis de frecuencia léxica. Detectar términos asociados al género, por ejemplo, "alumno", "alumna", "estudiantes", y su distribución.
- Análisis de polaridad léxica. Explorar si las expresiones neutras tienen valoraciones implícitas.
- Análisis gramatical y sintáctico. Clasificar estructuras gramaticales por su grado de neutralidad o inclusión, por ejemplo, el uso del masculino genérico.
- Análisis de coocurrencias. Ver qué términos suelen aparecer juntos con expresiones inclusivas, por ejemplo, “persona que realiza el TFG".

El modelado no es un fin predictivo sino una lente interpretativa, pues permite revelar cómo el lenguaje regula simbólicamente la inclusión o exclusión de identidades en documentos institucionales. Así, el uso de modelos requiere transparencia metodológica, es decir, es fundamental explicar qué modelos se han usado, qué supuestos implican y cómo se interpretan sus salidas. Este enfoque podría suponer el planteamiento de un activismo académico basado en datos, al mostrar cómo ciertos modos de escritura institucional pueden perpetuar prácticas no inclusivas

Fase V. Evaluación: Interpretación y validación de los hallazgos

En esta fase consiste en analizar críticamente los resultados del modelado para comprobar si responden a las preguntas iniciales. Esto implica contrastar los hallazgos computacionales, por ejemplo, los patrones lingüísticos, con interpretaciones humanísticas. En ese sentido, es fundamental entender qué significan los datos dentro de su contexto histórico, cultural o político. Por ejemplo, si se detectara un uso dominante de lenguaje masculino en la normativa, habría que preguntarse si refleja una tradición institucional, una estrategia discursiva o una omisión voluntaria o involuntaria. Esta fase es iterativa, es decir, si los resultados no encajan, puede ser necesario volver a fases anteriores.

Fase VI. Despliegue: Comunicación del resultado del análisis

La fase final implica difundir los resultados y mostrar su relevancia, normalmente mediante publicaciones, seminarios y/o presentaciones en congresos. En el caso que presentamos, además de compartir el conocimiento académico, esta fase se podría plantear para dar visibilidad al impacto social del análisis, esto es, explicar cómo los hallazgos de la investigación pueden contribuir a políticas lingüísticas inclusivas o mejorar normativas universitarias.

Conclusión

El modelo CRISP-DM ofrece un marco claro, riguroso y flexible para aplicar la minería de datos en las Humanidades. Su estructura permite integrar técnicas computacionales con una interpretación crítica propia del enfoque humanístico. En el caso del análisis del lenguaje inclusivo en normativas académicas, este modelo no solo facilita la detección de patrones lingüísticos, sino que también permite reflexionar sobre sus implicaciones sociales, históricas y culturales. Entendemos, pues, que el modelo de análisis CRISP-DM no solo puede ser útil para estudios económico-empresariales, sino que puede resultar un modelo de referencia para generar conocimiento significativo y con impacto social desde las Humanidades Digitales.

Bibliografía

BERRY, D. M. (2011), *The Philosophy of Software: Code and Meditation in the Digital Age*, London, Palgrave Macmillan.

CASERO-ALONSO, V.; CELIS, Á.; & LOZANO ZAHONERO, M. (2024), «Minería de textos», en G. FERNÁNDEZ-AVILÉS & J. M. MONTERO (eds.), *Fundamentos de ciencia de datos con R* (1.ª ed.), McGraw-Hill Interamericana de España, pp. 653–670, recurso en línea: https://cdr-book.github.io/mineria-textos.html [Acceso: 14/10/2025].

ESPINOSA ZÚÑIGA, J. J. (2020), «Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública», *Ingeniería, Investigación y Tecnología*, 21(1), pp. 1–10.

FERNÁNDEZ AVILÉS, G.; SANCHÍS-MARCO, L.; RODRÍGUEZ, A.; CELIS, Á.; MONTERO, J. M.; & ARELLANO, I. (2024), «Enhancing Teaching Innovation in Business Administration and Management Through Interdisciplinary Coordination in Quantitative Subjects», en M. D. C. VALLS MARTÍNEZ & J. MONTERO (eds.), *Teaching Innovations in Economics: Towards a Sustainable World*, Cham, Springer. https://doi.org/10.1007/978-3-031-72549-4_2

La gestión y el procesamiento de datos del corpus Ameresco a través de la herramienta computacional Oralstats Aroca

Andrea Carcelén Guerrero
Universidad Católica de Valencia San Vicente Mártir –
University of Helsinki
Adrián Cabedo Nebot
Universitat de València

Palabras clave: Lingüística computacional; corpus lingüísticos; español coloquial
Keywords: Computational linguistics; linguistic corpora; colloquial Spanish

Los corpus orales como herramientas de análisis lingüístico en el ámbito digital

Tal y como indica el *Informe del Estado actual de los corpus en español, lenguas cooficiales y variantes del español* (Ministerio de Asuntos Económicos y Transformación Digital / Instituto Cervantes, 2023), los corpus son recursos indispensables para el desarrollo de las Tecnologías del Lenguaje y de modelos en el campo del procesamiento del Lenguaje Natural. Así, según este informe, para que el español ocupe un papel destacado dentro del mundo digital y de las nuevas tecnologías del lenguaje, se debe asegurar la disponibilidad de dichos corpus, así como garantizar unas condiciones mínimas homogeneizadas atendiendo al etiquetado, la representatividad y la fiabilidad de los datos recogidos. Trabajos previos señalan también la necesidad de que haya un estándar de codificación que permita el intercambio y la reusabilidad de los datos (Adolphs y Knight, 2010, Payrató, 1995). De esta manera, cualquier investigadora o investigador tiene a su alcance muestras de habla reales de diferentes géneros discursivos que pueden servir de base para el intercambio de datos lingüísticos desde cualquier disciplina.

En este contexto, el corpus Ameresco (Albelda y Estellés, en línea) se crea con el propósito de recoger muestras de habla espontánea para el análisis del español coloquial en sus diferentes variedades dialectales (Briz, 1998 [2001], 2016). Un corpus con aspiraciones panhispánicas que se hermana con los principios metodológicos establecidos por PRESEEA (Moreno Fernández, 2021) y el corpus Val.Es.Co. (Briz y Grupo Val.Es.Co., 1995, 2002).

Si bien, nuestra lengua cuenta con un amplio desarrollo en el campo de los corpus lingüísticos (veánse, entre otros, los trabajos panorámicos de Carcelén, 2024 o Llisterri, 2021), las dificultades metodológicas asociadas a la recopilación de corpus orales hacen que cuenten con una menor representación (Briz, 2012). Sin embargo, es destacable el compromiso desde los diferentes equipos de investigación por poner a disposición de la comunidad científica el material oral reunido en estos corpus, promoviendo así su reutilización y el intercambio de datos para el análisis lingüístico desde diversas perspectivas disciplinarias. Considerando el esfuerzo económico y el tiempo que requiere la elaboración de corpus orales, compartir los resultados con la comunidad de investigadores e investigadoras debería consolidarse como una práctica común y habitual.

El corpus Ameresco

Tras esta contextualización, esta comunicación tiene como objetivo presentar el corpus oral de conversación espontánea Ameresco (América y España Español Coloquial) y su metodología de construcción de manera que pueda servir como base replicable para construir otros corpus. Asimismo, se presenta la herramienta de explotación y visualización de datos, Oralstats Aroca (Cabedo y Carcelén).

El corpus Ameresco (Albelda y Estellés, en línea) surge por iniciativa de Briz en 2010 para el estudio de la conversación coloquial. En particular, su creación responde a la necesidad de contar con material de estudios para los fenómenos de la atenuación pragmática, así como de fenómenos pragmático-retóricos de la interacción espontánea. Es un corpus multidialectal que tiene carácter panhispánico, es decir, aspira a recoger materiales de las principales variedades diatópicas del español, si bien, es un corpus que aún no ha concluido, actualmente está compuesto por 202 conversaciones procedentes de 16 ciudades de 10 países de habla hispana (Argentina, Bolivia, Chile, Colombia, Cuba, Ecuador, España, Honduras, México y Panamá).

Ameresco está disponible en línea (https://corpusameresco.org/coloquial/web/) y cuenta con diversos modos de consulta. En primer lugar, permite un acercamiento sencillo desde la opción de repositorio de archivos, es decir, se puede acceder a todos los materiales que componen el corpus (archivos de audio anonimizados, transcripciones, tanto alineadas audio-texto en ELAN como en formato texto plano, las fichas técnicas y los archivos en Textgrid para su análisis prosódico en PRAAT). De esta manera, cualquier investigador o investigadora puede utilizar los datos y procesarlos según sus necesidades.

En segundo lugar, el corpus permite ser analizado por medio de un motor de búsqueda simple por intervención. Desde aquí, se pueden realizar consultas sobre formas lingüísticas exactas y filtrarlas por criterios sociolingüísticos (nivel de instrucción del hablante, sexo, edad y ciudad de procedencia). Esta opción de búsqueda nos permite acceder a las concordancias, además de facilitar la

descarga de los resultados, así como escuchar los fragmentos de audio en los que se encuentra la forma buscada.

Por último, la plataforma ofrece acceso a la búsqueda compleja a través de la herramienta Oralstats Aroca. A través de esta opción, los datos lingüísticos pueden analizarse atendiendo a diferentes variables, como son la frecuencia de palabras, las categorías de partes del discurso, bigramas y trigramas. Además de variables léxicas y morfológicas, incorpora variables prosódicas como el tono, la intensidad o la duración de las intervenciones.

Oralstats Aroca: una herramienta de consulta y descarga de datos de corpus orales

Como resultado, se expone cómo una herramienta informática, Oralstats Aroca, desarrollada por los lingüistas del proyecto Ameresco, cumple funciones tanto de gestión y procesamiento de datos como de análisis del corpus. Oralstats Aroca (Cabedo y Carcelén, 2022) se presenta como un entorno digital diseñado para transformar y examinar datos provenientes de cualquier tipo de interacción verbal, con un enfoque particular en conversaciones acompañadas de metadatos sociolingüísticos.

Este enfoque innovador permite integrar y cruzar toda la información recopilada en un análisis unificado, que puede ser tanto exploratorio como inferencial. Para ello, los datos se organizan dentro de un ecosistema de variabilidad compartida, compuesto por variables numéricas y categóricas generadas automáticamente. Estas últimas pueden seleccionarse de un conjunto predefinido (como sexo, edad, hablante) o diseñarse de forma personalizada según los objetivos específicos de la investigación (como estudios sobre atenuación, cortesía o descortesía, humor, ironía, entre otros).

La herramienta, desarrollada en R (R Core Team, 2021) y de acceso libre, permite analizar transcripciones sincronizadas con el audio mediante marcas temporales. Considera aspectos comunes en el análisis de corpus, como la frecuencia léxica, las categorías gramaticales y combinaciones léxicas (bigramas, trigramas), así como variables menos convencionales como el tono, la intensidad y la duración. De este modo, las transcripciones se enriquecen automáticamente con anotaciones de niveles adicionales como la morfosintaxis y la prosodia.

Oralstats se inscribe en el marco del uso de nuevas tecnologías y lenguajes de programación aplicados al análisis lingüístico. Su propósito es facilitar el acceso al corpus mediante una herramienta abierta que, además, permite a usuarios con conocimientos técnicos más avanzados modificar el código base y añadir nuevas capas de información.

Aunque en un diseño inicial la herramienta Oralstats estaba orientada al análisis fonético de transcripciones generadas por programas como ELAN o Praat, con el objetivo de identificar patrones verbales y paraverbales asociados a variables sociolingüísticas o a hablantes concretos, el uso en un corpus

conversacional como es Ameresco exigía modificaciones sustanciales. Estas se centraron especialmente en el módulo de visualización y consulta de la base de datos, con el fin de mejorar la experiencia del usuario. Así nació Oralstats Aroca, una versión modificada y de código abierto que prioriza la usabilidad por encima de las funciones estadísticas avanzadas (como mapas de calor o árboles de decisión.

Una de las ventajas más destacadas de Oralstats Aroca es su habilidad para generar visualizaciones de datos interactivas. La herramienta permite construir gráficos, tablas dinámicas y otros formatos visuales que facilitan la exploración del corpus. Estas representaciones no solo hacen más accesible la información, sino que también permiten a los usuarios interactuar con los datos para obtener detalles específicos y detectar patrones relevantes.

Bibliografía

ADOLPHS, S.; & KNIGHT, D. (2010), «Building a spoken corpus. What are the basics?», en A. O'KEEFFE & M. J. McCARTHY (eds.), *The Routledge Handbook of Corpus Linguistics*, London/New York, Routledge, pp. 38–52.

ALBELDA, M.; & ESTELLÉS, M. (coords.) (s. f.), *Corpus Ameresco*, Universitat de València, ISSN 2659-8337, recurso en línea: www.corpusameresco.com [Acceso: 14/10/2025].

BRIZ GÓMEZ, A. (1998 [2001]), *El español coloquial en la conversación. Esbozo de pragmagramática*, Barcelona, Ariel.

BRIZ GÓMEZ, A.; & GRUPO VAL.ES.CO. (2002), *Corpus de conversaciones coloquiales*, *Anejos Oralia*, Madrid, Arco/Libros.

CABEDO NEBOT, A.; & CARCELÉN GUERRERO, A. (2022), *Oralstats Aroca*, Valencia, recurso en línea: https://github.com/acabedo/aroca [Acceso: 14/10/2025].

CARCELÉN GUERRERO, A. (2024), «Novedades en los corpus digitales para el estudio del español oral», *Normas*, 14(1), pp. 241–260.

LLISTERRI, J. (2021), «Corpus para investigar sobre el componente fónico en español como LE/L2», en M. CRUZ & J. MUÑOZ (eds.), *e-Research y español LE/L2: investigar en la era digital*, London/New York, Routledge, pp. 164–196.

MORENO FERNÁNDEZ, F. (2021), *Metodología del Proyecto para el estudio sociolingüístico del español de España y de América (PRESEEA)*, Alcalá de Henares, Editorial Universidad de Alcalá.

PAYRATÓ, L. (1995), «Transcripción del discurso coloquial», en L. CORTÉS (ed.), *El español coloquial. Actas del I Simposio sobre Análisis del Discurso Oral*, Almería, Universidad de Almería, pp. 43–70.

SECRETARÍA GENERAL TÉCNICA DEL MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL; & INSTITUTO CERVANTES (2023), *Informe del estado actual de los corpus en español, lenguas cooficiales y variantes del español*.

AIMeRiC: análisis informatizado del metro y el ritmo del catalán

Pere Garau Borràs
Universitat de les Illes Balears

Palabras clave: Escansión automática; patrones rítmicos; métrica; lengua catalana
Keywords: Automatic scanning; rhythmic patterns; meter; Catalan language

La escansión es el conjunto de procedimientos que implica el análisis métrico de un verso, un método conocido en el análisis poético que pone de relieve los elementos prosódicos del verso para capturar su ritmo y, en última instancia, su metro. Las estrategias de escansión pueden diferir entre lenguas, debido a las características fonológicas y prosódicas locales, así como su tradición poética. Por eso, la tarea de escandir un verso requiere conocimientos precisos sobre la fonología de una lengua y su tradición métrica (Agirrezabal et al. 2016:3–5), conocimientos que un experto automático en escansión debe poseer y jerarquizar de manera sistemática en el análisis del verso. En catalán, la escansión implica la segmentación silábica de la cadena y la asignación de un nivel de prominencia para cada sílaba, como ejemplifica (1), con un verso de J. V. Foix. Para superar este ejercicio, el escansor debe conocer los procesos de silabificación i el sistema acentual de la lengua catalana. En otro estadio, el escansor debe conocer el sistema rítmico de acentos secundarios de la lengua, para poder realizar una escansión como la de (2), con un verso de Joan Maragall. También debe conocer los procesos de silabificación del catalán, para resolver contactos vocálicos como los que se dan en el ejemplo (3), verso de Joan Alcover; pero a su vez, debe saber cuándo no se aplican estas resoluciones, como demuestra el verso (4), de Jaume Bofill i Mates, y cuando deben violentar, como en el ejemplo (5) de Pere Serafí. Y más allá de todo esto, el escansor debe tener la capacidad de identificar estos cinco ejemplos como un mismo tipo de verso: el endecasílabo.

Em	plau	d'at-	zar	d'er-	rar	per	les	mu-	ra-	lles
0	1	0	1	0	1	0	0	0	1	0

on	tots	els	a-	met-	llers	ja	són	flo-	rits
0	1	0	1	0	1	0	1	0	1

E-	lla	se	de-	can-	ta‿i	de-	sa-	pa-	reix

1	0	1	0	1	0	0	1	0	1

La	vos-	tra	a-	mor	no	té	pi-	e-	tat
0	1	0	0	1	0	1	0	0	1

tro-	bar	ja-	mai	pie-	tat,	que	per	es-	tre-	na,
0	1	0	1	0	1	0	1	0	1	0

Con escasos ejemplos se esbozan los conocimientos básicos que el escansor de poesía debe poseer. Sintéticamente, debe tener conocimientos fonológicos sobre la silabificación y asignación del acento léxico y rítmico, y los conocimientos de la tradición métrica para aplicar soluciones prosódicas alternativas para satisfacer el metro. En conjunto, un programa destinado a la escansión automática de la poesía catalana debe contener un conjunto de reglas jerarquizadas que permitan obtener una representación abstracta i cuantificable del ritmo del verso a partir de una cadena textual.

En esta propuesta se abordan las especificidades de un programa diseñado para la escansión del verso catalán: AIMeRiC: Análisis Informatizado del Metro y el Ritmo del Catalán. El programa ha sido desarrollado en Python y permite obtener informaciones de ritmo y metro a partir del verso escrito. Así, esta propuesta supone la primera aproximación para la lengua catalana a la escansión automatizada y se sitúa al lago de numerosas experiencias en este campo, como bien recoge De Sisto et al. (2024) y podemos atestar para el inglés, español, griego clásico, latín, vasco, checo, neerlandés, finés, francés, alemán, italiano, occitano antiguo, portugués o ruso. Metodológicamente, el programa desarrollado consiste en un sistema de escansión basado en reglas, un autómata finito que concatena la aplicación de distintas operaciones jerarquizadas y ordenadas.

La herramienta desarrollada, pues, recibe un verso como input, del cual genera la escansión en formato de texto, el contorno acentual en formato binario numérico (0, para sílabas átonas, 1 para tónicas), detecta subdivisiones internas, licencias, y devuelve la segmentación en palabras y sílabas, con la información gramatical i prosódica de cada elemento. El programa permite una interacción básica con el usuario, a través de terminal para conseguir el escaneo de versos aislados. Aun así, el propósito principal del programa ha sido implementarlo como motor de un proyecto más amplio te etiquetaje y compilación de un corpus de endecasílabos catalanes. Así, en interacción con el módulo de etiquetaje, el programa registra en formato diccionario para su almacenaje en JSON, las diversas informaciones de escansión de cada verso.

Figura 1. Ejemplo de escansión de un verso con AIMERIC.

El flujo de trabajo del programa es parcialmente cíclico. En primer lugar, tras recibir la cadena textual, se procesa el texto para limpiarlo de elementos que puedan distorsionar la escansión: signos de puntuación, espacios extra, errores de codificación, caracteres no alfabéticos, etc. Preparado el texto, se procede con su etiquetaje gramatical, para el cual se usa el paquete SpaCy (Honnibal et al. 2020). En base a las etiquetes gramaticales se determinará, después, el nivel de tonicidad del acento léxico conjuntando la propuesta de Jiménez (2019:229–32) y (Haider 2023:232). La silabificación, usando el algoritmo de Liang (1983), adecuado por Valiente and Fuster (1993:254–56) al catalán y desarrollado por Ortolà (2023) en Python, es el paso previo a la asignación del acento léxico y secundario, que distingue cinco niveles de tonicidad: 0 (sílaba átona), 1 (sílaba desacentuada), 2 (acento secundario), 3 (acento primario de palabras bivalentes), 4 (acento primario), 5 (acento primario de verso). Seguidamente, se aplica la resilabificación, consistente en la resolución de los contactos vocálicos entre vocales adyacentes, según las reglas descritas por Vallverdú (2008) y, en caso necesario, la aplicación de licencias métricas. En penúltimo lugar, se asigna el acento prosódico a través de una reconfiguración contextual del acento léxico de cada sílaba, en qué determinados acentos pueden ser promovidos o degradados en la escala de tonicidad. Finalmente, se etiquetan los elementos métricos del verso: subdivisión, número de sílabas métricas y número de sílabas extramétricas, principalmente. Cabe señalar que actualmente, el programa se centra en el escaneo del endecasílabo, pero se prevé, en un cercano desarrollo, abastar todas las formas versales conocidas del catalán.

El programa desarrollado demuestra una alta precisión. Para su evaluación se ha acarado el escaneo automático y manual de la obra *Sol, i de dol* de J. V. Foix, un conjunto de 70 sonetos endecasilábicos, 980 versos. Para calcular el índice de precisión se han usado distintas aproximaciones. El cálculo de la desviación respeto las medianas de las frecuencias de tonicidad del contorno acentual de ambas escansiones reporta una precisión del 99,86%. El recuento total de escansiones idénticas entre ambos sistemas consigue un índice de precisión del 99,29%. Y el cálculo basado en la distancia de edición entre ambas escansiones, método usado en Agirrezabal et al. (2016), presenta una precisión del 98,67%. Principalmente, los errores detectados se relacionan con la acentuación de palabras o nombres propios de origen extranjero o con la

resolución del acento prosódico en colisión de más de tres palabras bivalentes. Su rendimiento, ejecutado en un ordenador con AMD© RYZEN™ 5-5500U con 16 GB de memoria RAM DDR4 registra una demora de 9 segundos en el etiquetaje del corpus de J. V. Foix de 980 versos, lo que deduce un tiempo medio de 0,009 segundos por verso.

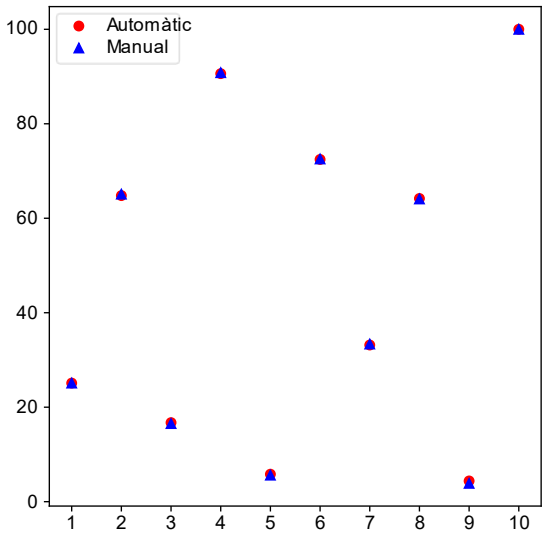


Figura 2. Comparación entre los contornos acentuales de la escansión automática y la manual de Sol, i de dol de J. V. Foix

En definitiva, nuestra propuesta supone la primera aproximación conocida a la escansión automática del verso catalán. Así pues, esta propuesta sitúa la investigación de la métrica catalana en el eje de las principales métricas europeas que cuentan con diversos sistemas para su escansión automática. A la vez, proyecta la posibilidad de configurar fácilmente corpus métricamente anotados para la investigación del verso en esta lengua. La herramienta ya ha sido usada exitosamente en la creación y análisis de un corpus de endecasílabos catalanes que cuenta actualmente con un total de 137.349 versos de 140 autores distintos.

Bibliografía

AGIRREZABAL, M.; ARRIETA, B.; ASTIGARRAGA, A.; & HULDEN, M. (2016), «ZeuScansion: A Tool for Scansion of English Poetry», *Journal of Language Modelling*, 4(1), pp. 3–28. https://doi.org/10.15398/JLM.V4I1.102

HAIDER, T. (2023), *A Computational Stylistics of Poetry: Distant Reading and Modeling of German and English Verse*, Stuttgart, Institut für Maschinelle Sprachverarbeitung (tesis doctoral).

HONNIBAL, M.; MONTANI, I.; VAN LANDEGHEM, S.; & BOYD, A. (2020), «spaCy: Industrial-Strength Natural Language Processing in Python», *Zenodo*. https://doi.org/10.5281/zenodo.1212303

JIMÉNEZ, J. (2019), «Poesia bastida amb materials planers: Optimització del ritme en els alexandrins d'Estellés», *Zeitschrift für Katalanistik: Revista d'Estudis Catalans*, 32, pp. 223–269.

LIANG, F. M. (1983), «Word Hy-Phen-a-Tion by Com-Pu-Ter», Stanford, Stanford University (tesis doctoral).

ORTOLÀ, J. (2023), «Hyphen-Ca».

DE SISTO, M.; HERNÁNDEZ-LORENZO, L.; DE LA ROSA, J.; ROS, S.; & GONZÁLEZ-BLANCO, E. (2024), «Understanding Poetry Using Natural Language Processing Tools: A Survey», *Digital Scholarship in the Humanities*, 39(2), pp. 500–521. https://doi.org/10.1093/llc/fqae001

VALIENTE, G.; & FUSTER, R. (1993), «Typesetting Catalan Texts with TeX», *TUGboat*, 14(3), pp. 252–259.

VALLVERDÚ, T. (2008), «Fenòmens en grups vocàlics», en J. SOLÀ; M.-R. LLORET; J. MASCARÓ; & M. PÉREZ SALDANYA (eds.), *Gramàtica del català contemporani*, vol. 1, Barcelona, Editorial Empúries, pp. 125–167.

VENY-MESQUIDA, J. R. (2024), *Primera lliçó sobre mètrica catalana*, Barcelona, Publicacions de l'Abadia de Montserrat.

La *Gramàtica essencial de la llengua catalana*: una gramática online

Maria Josep Cuenca
Institut d'Estudis Catalans – Universitat de València

Palabras clave: Gramática online; gramática normativa; Institut d'Estudis Catalans; glosario online
Keywords: Online grammar; normative grammar; Institute of Catalan Studies; online glossary

En 2016 se publicaron la *Gramàtica de la llengua catalana* (GIEC) y la *Ortografia catalana* (OIEC). Antes de la finalización efectiva de estas obras, en 2015, se puso en marcha el proyecto de la Gramàtica essencial de la llengua catalana, que dio lugar a tres recursos digitales: la *Gramàtica essencial de la llengua catalana* (GEIEC, 2018), versión https://geiec.iec.cat), la *Gramàtica bàsica i d'ús de la llengua catalana* (GBU, 2019) y el *Glossari de termes gramaticals* (https://cit.iec.cat/GTG, 2020).[1] La GBU y la GIEC fueron digitalizadas en junio de 2022 (https://gbu.iec.cat; https://giec.iec.cat).

En esta comunicación, explicaremos la estructura de la GEIEC, una gramática digital, no digitalizada, que sirve de punto de encuentro de las diferentes obras gramaticales de l'Institut d'Estudis Catalans.

Como primera fase, realizamos un prototipo de plataforma y procedimos a resumir y marcar los capítulos de la GIEC. Establecimos cuatro principios generales de adaptación basados en la simplificación y la complementación: a) resumen y reorganización de los contenidos, b) simplificación expositiva, a partir de la inserción de cuadros resumen y de un glosario interactivo, c) simplificación terminológica y d) simplificación de las marcas normativas y descriptivas.[2]

a) *Resumen y reorganización de los contenidos*. Eliminamos o resumimos informaciones no esenciales, como las correspondientes a los capítulos de conceptos básicos de la GIEC (capítulos 1, 6 y 13), que se integraron en otros puntos y sobre todo en el glosario, y pasamos de los 5 dígitos de la GIEC a 3 dígitos (capítulo, apartado y subapartado) reorganizando profundamente algunos contenidos.
b) *Simplificación expositiva y elementos de soporte.* Trabajamos la

[1] Sobre los antecedentes y el desarrollo del proyecto, cfr. Cuenca (2020: apartado 2). Cuenca (2024) contiene una descripción pormenorizada de los tres recursos.
[2] La GBU es una versión reducida de la GEIEC pensada para la publicación en papel y orientada a la resolución rápida de dudas normativas.

manera de presentar la información. Dimos una mayor visibilidad a la normativa en la exposición y gráficamente e ideamos un sistema que permite ampliar la información progresivamente.

c) *Simplificación terminológica.* Realizamos una selección y reducción terminológica y sistematizamos las definiciones del texto, que suelen encabezar cada apartado. Eliminamos términos gramaticales muy especializados, redujimos la sinonimia y desterminologizamos algunos términos[3].

d) *Simplificación de las marcas normativas y descriptivas.* Hicimos una revisión de las marcas normativas y descriptivas, que se redujeron para mostrar la priorización de opciones de forma más directa y comprensible para el público general.[4]

La GEIEC se concibió como gramática online, es decir, no es una gramática en papel digitalizada, a diferencia de al GIEC y la GBU. La plataforma GEIEC incluye diferentes componentes de consulta y navegación: sumario de capítulo con enlaces por apartado y subapartado, cuadros resumen, glosario, buscador por diferentes criterios, palabras clave, remisiones internas y remisiones externas a la GIEC y a la OIEC, en texto, y a la GIEC y a la GBU en el lateral derecho (estas dos gramáticas también incluyen esta funcionalidad a las otras dos gramáticas). La figura 1 muestra algunos de estos componentes.

Figura 1. Componentes básicos de la GEIEC (apartado 16.6)

[3] Sobre el trabajo terminológico, véase Cuenca (2019).

[4] En la GBU, el carácter normativo o descriptivo de las explicaciones está marcado, además, gráficamente, con iconos específicos, que permiten identificar el tipo de información que se consulta.

Los cuadros de la GEIEC. La GEIEC incluye 259 cuadros y 30 figuras, la mayoría inexistentes en la obra de origen, la GIEC. Funcionan como resumen y punto de entrada al contenido de cada apartado y esquematizan los fenómenos más representativos o frecuentes que se describen en el texto posterior. La mayoría contiene ejemplos representativos y comentarios de uso, que pueden ser suficientes para resolver una duda. Si se quiere ampliar la información, el signo (+) a la derecha de cada fila lleva directamente al apartado en el que se explica cada aspecto, como vemos en la figura 1. Cuando la explicación de un aspecto se desarrolla en otro apartado, aparece un enlace en forma de flecha (⇉), que permite abrirlo en una nueva ventana si se quiere ampliar la información. La flecha marca también la remisión interna dentro del texto.

Los archivos de sonido. Para complementar los aspectos fonéticos de la gramática, se incluyeron casi 2000 archivos de sonido (figura 2), que se marcan con un icono de audio (🔊) que da acceso a la reproducción de un fichero de sonido en diferentes variedades el catalán.[5]

Figura 2. Archivos de sonido en cuadro y en texto en la GEIEC

Marcas orientativas. La GEIEC contiene marcas de usos orientativas. Los pasajes que incluyen información normativa o relevante desde el punto de vista del uso se marcan situando un icono (🔴) al final del párrafo. Por lo general, estos párrafos contienen alguna etiqueta normativa (*no aceptable, se evita, es preferible*).

El glosario. El glosario de la GEIEC incluye 600 términos acompañados de definición, ejemplos y enlace al apartado donde se define (figura 3). Se pueden consultar alfabéticamente o por capítulos.

[5] Para más información sobre las características y la obtención de los ficheros de sonido, véase el apartado 3.3 de la presentación de la GEIEC (https://geiec.iec.cat/docs/presentacio_GEIECact2.rev.pdf)

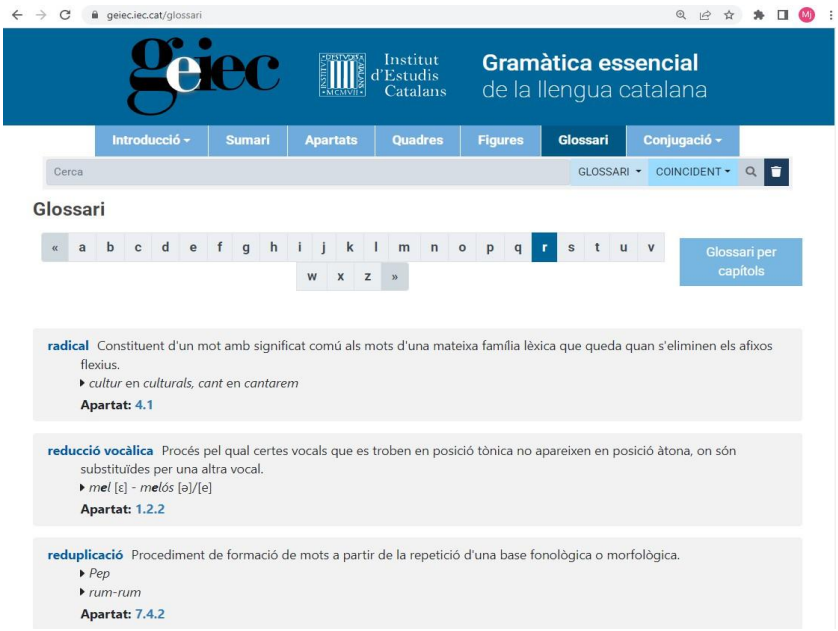


Figura 3. Glosario de la GEIEC

Los términos aparecen también en texto, dentro de las explicaciones, marcados en azul y enlazados con el glosario (figura 4).

Figura 4. Marcaje de términos y palabras clave

En la parte superior de cada pantalla, aparecen las palabras clave, una selección de los términos de cada capítulo, que funcionan como ventana emergente, lo que permite la consulta sin cambiar de pantalla, a diferencia de la consulta del glosario.

El *Glosario de términos gramaticales* A partir de las tareas realizadas para el glosario de la GEIEC, creamos el *Glosario de términos gramaticales*

(https://cit.iec.cat/GTG, 2020), que contiene 700 artículos[6]. La búsqueda por defecto (por entrada, "que contenga") nos ofrece todos los términos que contienen una palabra o una secuencia, como, por ejemplo, *demostratiu*:

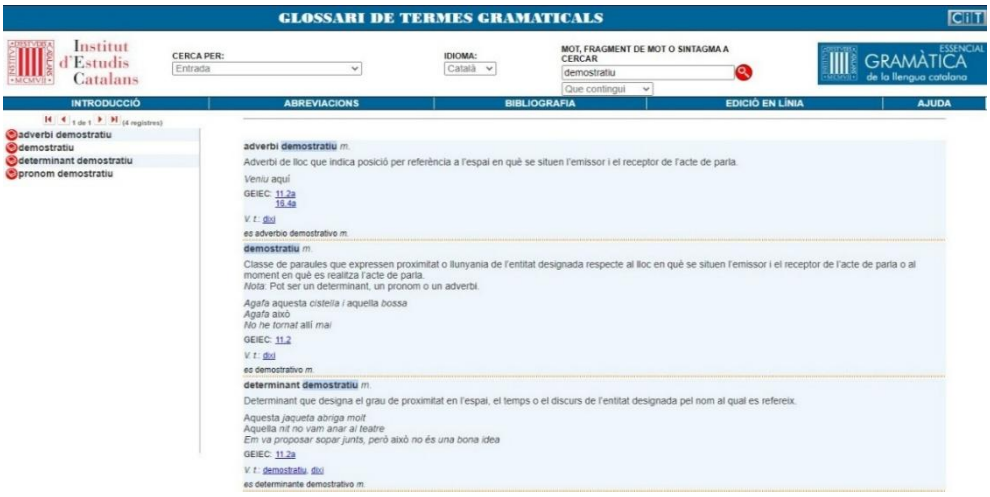


Figura 5. Búsqueda en GTG por entrada que contenga *demostratiu*

Se trata de una obra lexicográfica autónoma, a diferencia del glosario de la GEIEC, que es un elemento auxiliar y, por tanto, restringido a los términos definidos en la gramática (podía haber términos que no aparecieron porque no se usan en la obra). Cada artículo incluye definición, ejemplo y remisión a la GEIEC, así como informaciones complementarias diversas. La microestructura de un artículo del GTG se muestra en la figura 6.

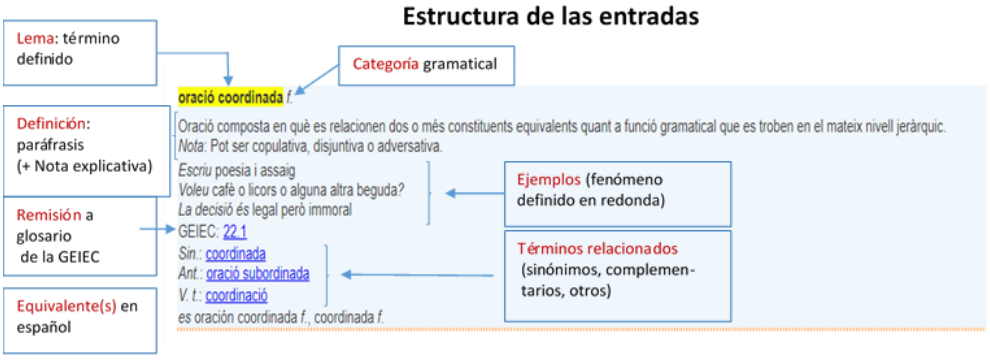


Figura 6. Microestructura de los artículos del GTG

El GTG incorpora nuevas entradas, términos de carácter general (por ejemplo, *enunciat*), que se utilizan en el texto de la GEIEC pero que no se tratan propiamente en ningún capítulo, por lo que quedaban fuera de los criterios de selección terminológica de aquella obra. También se han incrementado las informaciones adicionales en forma de nota que sigue la definición: sinónimos, términos relacionados conceptualmente, términos con los que la entrada mantiene relaciones de complementariedad o contraste y equivalentes en castellano, basándonos en el Glosario de términos gramaticales de la RAE

[6] Sobre el GTG y cómo se consulta, cfr. https://cit.iec.cat/GTG/default.asp?opcio=1.

(2019).

Sistemas de búsqueda. La búsqueda de un aspecto gramatical se puede realizar por cuadro y por glosario, entre otros sistemas. Por ejemplo, si buscamos *pronom relatiu* en cuadro, aparecen los cuadros que contienen este término, tanto en el título como dentro de cuadro. Desde ahí podemos acceder al cuadro solo o al apartado que lo incluye.

Figura 7. Búsqueda en cuadro en GEIEC (*pronom relatiu*)

Si buscamos *relatiu* en glosario (figura 8), obtenemos todos aquellos términos o definiciones que lo incluyen.

Figura 8. Búsqueda en el glosario de la GEIEC (*relatiu*)

Interrelación entre las obras. Las tres gramáticas se entienden como

obras complementarias que se ajustan a perfiles de usuario (o de búsqueda) distintos: especializado (GIEC), general (GEIEC) y de (alta) divulgación (GBU). El sistema de gramática digital de la GEIEC y la digitalización posterior de la GIEC y la GBU permite articular un sistema de interconexión múltiple. En el lateral derecho de la pantalla, se sitúan iconos de las otras dos gramáticas, que permiten navegar en los apartados equivalentes de la GIEC y la GBU para ampliar o reducir el contenido (como se puede ver en las figuras 1, 2 y 4). Así, la plataforma permite pasar de una obra a otra y también conecta con la ortografía, a través de enlaces. Igualmente, existe una conexión entre el GTG y la GEIEC (unidireccional).

En síntesis, en esta comunicación mostramos brevemente los componentes de la GEIEC, una gramática inherentemente digital. Las tareas de adaptación desde la GIEC se han basado en la simplificación del contenido y la estructura, pero sobre todo en la creación de instrumentos para la consulta en línea y la puesta en marcha de sistemas de interrelación entre las diferentes obras y recursos gramaticales.

Bibliografía

CUENCA, M. J. (2019), «La terminologia gramatical: de la *Gramàtica de la llengua catalana* a la *Gramàtica essencial de la llengua catalana*», en J. FELIU & M. TRIAS (eds.), *Gramàtica, esport i terminologia*, Barcelona, Institut d'Estudis Catalans, pp. 91–112.

CUENCA, M. J. (2020), «La *Gramàtica essencial de la llengua catalana* de l'Institut d'Estudis Catalans: característiques i aportacions», *Llengua & literatura*, 30, pp. 51–84, recurso en línea: https://publicacions.iec.cat/repository/pdf/00000286/00000046.pdf [Acceso: 14/10/2025].

CUENCA, M. J. (2024), «Eines normatives per a l'ensenyament de la gramàtica», en D. CASALS *et al.* (eds.), *Llegir, escriure i parlar a l'aula*, Barcelona, Institut d'Estudis Catalans, pp. 49–88, recurso en línea: https://publicacions.iec.cat/PopulaFitxa.do?moduleName=botiga&subModuleName=&idCatalogacio=42428 [Acceso: 14/10/2025].

INSTITUT D'ESTUDIS CATALANS (IEC) (2019), *Gramàtica bàsica i d'ús de la llengua catalana (GBU)*, Barcelona, IEC; recurso en línea: https://gbu.iec.cat [Acceso: 14/10/2025].

INSTITUT D'ESTUDIS CATALANS (IEC) (2018), *Gramàtica essencial de la llengua catalana (GEIEC)*, Barcelona, IEC; recurso en línea: https://geiec.iec.cat [Acceso: 14/10/2025].

INSTITUT D'ESTUDIS CATALANS (IEC) (2016), *Gramàtica de la llengua catalana (GIEC)*, Barcelona, IEC; recurso en línea: https://giec.iec.cat [Acceso: 14/10/2025].

INSTITUT D'ESTUDIS CATALANS (IEC) (2020), *Glossari de termes gramaticals (GTG)*, Barcelona, IEC; recurso en línea: https://cit.iec.cat/GTG [Acceso: 14/10/2025].

INSTITUT D'ESTUDIS CATALANS (IEC) (2016), *Ortografia catalana (OIEC)*, Barcelona, IEC; recurso en línea: https://oiec.iec.cat [Acceso: 14/10/2025].

REAL ACADEMIA ESPAÑOLA; & ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2019), *Glosario de términos gramaticales*, Madrid, Real Academia Española; Salamanca, Ediciones de la Universidad de Salamanca, recurso en línea: https://www.rae.es/gtg/ [Acceso: 14/10/2025].

El *Llibre dels feits* del rey Jaime I de Aragón: de la oralidad artística a la reproductibilidad técnica

Anna Saumell Almirall
Universitat Pompeu Fabra
Clàudia Raventós Albà
Institut Can Puig
Pere Poy Baena
Universitat de Barcelona

Palabras clave: Gamificación; Jaime I; literatura; humanidades digitales
Keywords: Gamification; Jaime I; literature; digital humanities

En el presente artículo se describe el proceso de adaptación al formato interactivo digital de un clásico de la literatura medieval: el *Llibre dels feits* de Jaume I. Dicha adaptación, en la actualidad, se encuentra en proceso de realización. A continuación, se exponen las bases teóricas, metodológicas y técnicas de las que parte esta conversión al formato interactivo.

En *La obra de arte en la época de su reproductibilidad técnica,* Walter Benjamin (2003) introduce varias claves para comprender el proceso de reproducción de obras de arte en formatos contemporáneos. En primer lugar, el concepto de reproductibilidad técnica se asocia a aquellas obras de arte que son susceptibles de ser adaptadas en contenido, forma y/o significado dentro de un modo de exhibición distinto al de la obra original. El ejemplo por excelencia, sin duda, son las adaptaciones cinematográficas de obras literarias. Benjamin asocia la reproductibilidad técnica a la pérdida de una característica intrínseca de toda obra de arte: el aura. Este término define todo aquello que concierne y rodea la obra de arte original en su formato y contexto de contemplación original. En una obra de arte "aurática", la autenticidad de la obra de arte se hace patente en forma de visualización ritual. Esta autenticidad, según Benjamin, se pierde al ser reproducida. Por ende, esas obras de arte originales permanecen inalcanzables ante la imposibilidad de ser reproducidas de forma fiel al contexto y formato originales.

No obstante, Benjamin es consciente de la transición de las obras de culto hacia lo que hoy llamamos cultura de masas. El valor de exhibición se añade a la obra reproducida como factor que marca su expansión hacia otros tipos de público espectador (Radetich, 2012). Este concepto, "valor de exhibición", resulta, quizás, más práctico de entender y usar en la actualidad, teniendo en cuenta el ecosistema multidisciplinar del arte y las experiencias artísticas de hoy, y cómo éstas suelen estar en diálogo con los medios de comunicación de masas (Mora,

2012). En el caso de la literatura, esto se traduce en forma de integración de elementos visuales, auditivos e incluso interactivos dentro de la obra, abriendo un debate sobre las posibilidades de la literatura en el siglo XXI (Scolari, 2013).

En el caso de Jaume I y el *Llibre dels feits*, nos encontramos con una obra literaria del siglo XIII escrita en primera persona, aunque no fue redactada por el propio rey. El escritor desconocido de esta obra, a quien el rey Jaume I habría encargado y dictado toda la obra, supuestamente fue un funcionario de la corte. No obstante, al adentrarnos en el texto es posible ver que el estilo de redacción, uso de diálogos e incluso expresiones concretas y manierismos lingüísticos cambian, aparecen y desaparecen a lo largo de las décadas que aborda el libro. Esto hace pensar en la posibilidad de que no fuese un solo escribano, sino varios, quienes redactaron el texto original a lo largo de la vida del monarca. Este estilo cambiante es, sin duda, uno de los puntos que dificulta la lectura para el lector contemporáneo.

A nivel práctico, una lectura de esta obra en el contexto actual requiere por parte del lector una renuncia a su sistema de valores actuales, así como un esfuerzo por imaginar un despliegue de interacciones socioculturales, políticas y morales de las que disponía la sociedad del siglo XIII, pero que actualmente quedan fuera del alcance de cualquier lector contemporáneo. Este esfuerzo se intensifica cuando, en lugar de tratarse de lectores adultos especializados, el lector modelo que debe abordar la obra es un grupo de adolescentes en un contexto de aula de secundaria.

El problema de investigación que originó el presente estudio surge en una clase de 2º de bachillerato (17-18 años), en un centro público del área metropolitana de Barcelona, al descubrir el elevado número de alumnos que, a menos de un mes de las Pruebas de Acceso a la Universidad (PAU), no había ni siquiera iniciado la lectura de la obra *Terra baixa* (1896) de Àngel Guimerà, considerada de carácter prescriptivo según el currículo de bachillerato y objeto de evaluación para el acceso a la universidad (Raventós, 2021).

Los datos expuestos por Raventós indican que casi la totalidad de los alumnos del nivel (88,2%) no finalizó la lectura de *Terra baixa* porque consideraron que obtenían una comprensión más efectiva del macrotexto a través de resúmenes, adaptaciones audiovisuales y materiales de apoyo (guías de lectura o glosarios de vocabulario), en lugar de mediante la lectura lineal de la obra original. Del mismo modo, más de la mitad de los alumnos (64,7%) consideró que realizar una lectura lineal de la obra, sin disponer de materiales complementarios para contribuir a su comprensión, no es suficiente para lograr una evaluación satisfactoria en su nivel académico.

A partir de ahí, una amplia mayoría de los alumnos coincidió en la necesidad de una interfaz gráfica, así como de un componente interactivo que favoreciera de forma visual la comprensión de los aspectos de la lectura considerados trascendentales para el logro de los objetivos académicos.

Llegados a este punto, la investigación se centró en la incorporación de mecánicas de *gameplay* o gamificación, propias de la novela visual, al texto literario original de Guimerà para convertirlo en una narración hipertextual. Raventós (2021)

pone de relieve una cuestión vinculada a la realización de cualquier actividad lectora: leer implica también entender el significado de lo que se está leyendo, y el contexto en que se realiza la lectura y las intenciones de cada lector condicionan su comprensión del texto.

De este modo, la actividad de lectura propuesta está guiada por los objetivos que pretendemos alcanzar a través de ella; no es lo mismo leer para ver si nos interesa seguir leyendo que leer con la intención de buscar una información muy concreta. Evidentemente, no nos perturbará de la misma manera percibir algunas lagunas en nuestra comprensión en un caso u otro, lo que, con toda probabilidad, nos llevará a emplear diferentes estrategias para compensar esas lagunas (Solé, 1992).

El inconveniente aparece cuando el contexto de lectura no se corresponde con el contexto histórico en el que el texto fue escrito y, en consecuencia, con el contexto en que su autor previó que se realizaría la lectura.

En lo que respecta al *Llibre dels feits* de Jaume I, alcanzar el máximo nivel de comprensión de la obra requiere una recreación del artefacto medieval desvanecido a lo largo de los siglos. En este sentido, consideramos que el despliegue visual, sonoro e interactivo del que disponen las novelas visuales, y la correspondiente adaptación de la obra original al formato de videojuego educativo serio, puede ser un buen aliciente para que los lectores no especializados adquieran un nivel de *engagement* equiparable al del lector modelo original y, en consecuencia, una comprensión panorámica del macrotexto.

La novela visual es un género de videojuegos caracterizado por su presentación de historias interactivas, en las cuales se ofrece a los jugadores la elección de diferentes opciones que afectan al desarrollo de la historia a medida que esta transcurre (Camingue et al., 2021). Tiene unas mecánicas de juego y unas interfaces gráficas de usuario simples, ya que toda la interacción se introduce a través del ratón del ordenador o, en el caso de los dispositivos móviles, tocando la pantalla. Adicionalmente, suelen existir opciones de aceleración y/o automatización de la lectura. La experiencia de juego derivada de estas mecánicas e interfaces hace que jugar una novela visual se asemeje a leer una novela digital presentada con imágenes y sonido, pero con el componente interactivo adicional que facilita al jugador un control absoluto del ritmo de la historia y un mayor grado de concentración o *engagement* (Pratama et al., 2018). En este sentido, uno de los puntos fuertes de las novelas visuales es la incentivación de la percepción visual y la imaginación gráfica del jugador, proporcionando visualizaciones en pantalla complementarias a lo que el jugador está leyendo. Esta característica tiene la capacidad de apoyar el logro de objetivos de aprendizaje, especialmente en casos en que se ofrece la lectura de un texto que no es coetáneo del lector, favoreciendo la inmersión en el texto y la atención a la diversidad, atendiendo a las capacidades lectoras de los distintos grupos de usuarios.

En el presente proyecto se aplica la metodología de desarrollo de productos audiovisuales denominada *Multimedia Development Life Cycle* (MDLC) (Verawardina, 2023). Este método establece un sistema de trabajo repartido en seis

etapas: conceptualización (*concept*), diseño, recopilación de materiales, ensamblaje, prueba (*testing*) y distribución. Es importante remarcar que esta metodología ha sido escogida como complemento al desarrollo ordinario de videojuegos en siete fases (planteamiento, preproducción, producción, prueba o *testing*, prelanzamiento, lanzamiento y postproducción) (Moore y Novak, 2010), debido al carácter académico y de investigación natural del proyecto.

El proyecto, tal como ha sido planteado, pretende abordar de manera experimental la problemática de la comprensión lectora en estudiantes de educación secundaria y bachillerato a través de la adaptación al formato interactivo de una obra canónica de la literatura medieval que, en su formato original, no tiene un macrotexto fácil de comprender para lectores no especializados. A través de la videoludificación de esta obra, se quiere plantear un nuevo sistema de evaluación de la comprensión lectora de los estudiantes; comparando la experiencia de juego con la experiencia de lectura tradicional, tanto a nivel de disfrute como de calidad educativa. A su vez, en caso de que los resultados de finalización del proyecto fueran positivos, se plantearía una sistematización de la adaptación de contenidos narrativos similares dentro de la industria del videojuego independiente español.

Este artículo se integra en el proyecto *Lecturas y relecturas del «Llibre dels feits» (siglo XIII) del rey Jaume I de Aragón. Transferencia didáctica al lenguaje videolúdico*, que la Sociedad Catalana de Historia Cultural desarrolla durante el quinquenio 2022–2026, bajo la dirección del profesor Pere Ignasi Poy (Universidad de Barcelona–IRCVM) y con el patrocinio del Ayuntamiento de Barcelona y el Institut Ramon Muntaner.

Bibliografía

BENJAMIN, W. (2003), *La obra de arte en la época de su reproductibilidad técnica*, trad. A. E. Weikert, México, Ítaca.

CAMINGUE, J.; CARSTENSDOTTIR, E.; & MELCER, E. F. (2021), «What is a visual novel?», *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY), pp. 1–18.

MOORE, M. E.; & NOVAK, J. (2010), *Game Industry Career Guide*, Delmar, Cengage Learning.

PRATAMA, D.; WARDANI, W. G. W.; & AKBAR, T. (2018), «The visual elements strength in visual novel game development as the main appeal», *Mudra Jurnal Seni Budaya*, 33(3), pp. 326–333.

RADETICH FILINICH, N. (2012), «El arte, la técnica y lo político: a propósito de *La obra de arte en la época de su reproductibilidad técnica*, de Walter Benjamin», *Argumentos (México, D.F.)*, 25(68), pp. 15–26, recurso en línea: http://www.scielo.org.mx/scielo.php?script=sci_arttextCpid=S0187-57952012000100002Clng=esCtlng=es [Acceso: 05/09/2024].

RAVENTÓS I ALBÀ, C. (2021), *Literatura 2.0 en el batxillerat. Adaptació didàctica de Terra baixa (1896), d'Àngel Guimerà, al llenguatge videolúdic*

(TFM Máster de Formación del Profesorado – Lengua y Literatura Catalanas), Universitat de Barcelona, junio.

SCOLARI, C. A. (2013), *Narrativas transmedia*, Barcelona, Deusto.

SOLÉ, I. (1992), *Estrategias en la lectura: Vol. 5. Materiales para la innovación educativa*, Graó, recurso en línea: https://www.uv.mx/rmipe/files/2016/08/Estrategias-de-lectura.pdf [Acceso: 17/08/2023].

Hacia una transcripción asistida de la documentación inquisitorial: lineamientos para el entrenamiento de modelos (siglos XV-XIX)

Bárbara Santiago Medina
Consorcio ARIANE – IR* Huma-Num

Palabras clave: Inquisición española; HTR; adquisición textual; principios FAIR; Paleografía, Diplomática
Keywords: Spanish Inquisition; HTR; textual acquisition; FAIR principles; Paleography; Diplomacy

Introducción

La documentación que contienen los expedientes de procesos de fe de la Inquisición española constituye una de los santos griales de la moderna historiografía en cuanto a aplicación de los actuales sistemas de reconocimiento textual automatizado se refiere. Sin embargo, las problemáticas que presenta hacen que, por el momento, sea muy difícil alcanzar unos resultados óptimos, acordes a las grandes expectativas que se depositan en ellos. El "expediente", definido por la Real Academia Española como el "conjunto de todos los papeles correspondientes a un asunto o negocio" y, asimismo, la "serie ordenada de actuaciones administrativas, y también de las judiciales en los actos de jurisdicción voluntaria", es, por definición, algo heterogéneo. De carácter eminentemente manuscrito, no se escritura de una sola vez, por una única persona, en un cierto lugar y en un mismo momento. Por otro lado, la actuación de la institución inquisitorial abarca desde el último cuarto del s. XV hasta el primero del s. XIX, lo que implica algo más de 350 años de historia. A lo que hay que sumar su marco geográfico, que no solo se extiende por varios reinos peninsulares, sino también a los territorios americanos y a algunas posesiones italianas (Córcega y Cerdeña). Distintas autorías, diferentes lugares, lenguas diversas con características propias y en constante evolución... Todo esto se encontrará la persona que se acerque a los procesos de fe del Santo Oficio español. A lo que hay que añadir el que quizás, suponga el principal problema para los sistemas de *Handwritten Text Recognition* (HTR): la multiplicidad de grafías presentes en ellos.

La documentación inquisitorial y sus problemas de aproximación e interpretación

Como ya se ha mencionado, el trabajar con la documentación de la Inquisición española no es sencillo. No lo es para quien se acerque a ella utilizando medios tradicionales y, mucho menos, para quien quiera hacerlo viéndola a través del prisma de las nuevas tecnologías. Además de los conocimientos históricos, para poder acceder a la información que contienen los textos, es necesario contar con las herramientas que proporcionan las Ciencias Historiográficas, las cuales, con demasiada frecuencia, siguen siendo vistas y presentadas como meras técnicas auxiliares para la Historia, la Filología, el Derecho... Los 350 años de devenir histórico del Santo Oficio, que cubren desde finales del medievo o los albores de la modernidad hasta la edad contemporánea, hacen que en sus documentos hayan quedado recogidos ejemplos escriturarios de dos de los grandes ciclos de la escritura latina: el gótico y el humanístico. Cada uno con su propia evolución, características y formas distintivas. Grafías que pueden variar en modos de ejecución y grado de cursividad, lo que afecta a su legibilidad. Porque el verdadero reto del investigador reside en si es capaz o no de leer unos textos que, como se ha adelantado, pueden estar escritos en góticas cursivas castellanas o aragonesas de finales del s. XV, pero también en humanísticas derechas o cursivas, lo que incluye la conocida como “escritura procesal”, la que no entendía “ni Satanás”, o las ya más cercanas a nuestra época como son las propias del s. XVIII o comienzos del s. XIX y que nos resultan más familiares. Y a las grafías se suman las lenguas. Cualquiera que haya consultado los archivos inquisitoriales se habrá percatado que entre las páginas de sus expedientes podrán encontrarse documentos en latín, castellano y otras lenguas vernáculas, hispanas o europeas, por no hablar del hebreo o el árabe. Incluso si se toman como ejemplo los escritos solo en castellano, sus textos son una fuente inagotable para estudios de la lengua, pues a través de ellos se puede ver la evolución de ésta durante más de tres siglos. El castellano de finales del s. XV no es el mismo que el del s. XVII, como tampoco el del s. XIX. Las construcciones y los vocablos varían, en especial cuando se está ante documentos que, en cierto modo, tratan de asentar testimonios orales, donde lo normativo desaparece para dar paso a todo tipo de expresiones personales, sin olvidar los localismos.

Y todo esto sin olvidar las problemáticas derivadas de escrituras procedentes de distintas manos, con mayor o menor pericia caligráfica, infinidad de tipologías documentales, diversos soportes, que pueden ir más allá del papel o el pergamino, tamaños variados, etc. Por no mencionar la presencia incluso de objetos o imágenes iconográficas. Y tampoco es posible dejar a un lado la manera en que no es infrecuente la aparición de textos en escrituras distintas a la latina, lo que supone una dificultad añadida a la hora de acercarse al contenido de los procesos.

Una vez conocidas algunas de las características formales y materiales que presentan los expedientes por procesos de fe y que pueden complicar la implementación de un sistema de reconocimiento textual, cabe preguntarse si será posible hacerlo en un futuro próximo y siempre de acuerdo a criterios científicos. Más aún cuando, todavía, las tasas de error por carácter (CER) de los actuales modelos genéricos son demasiado altas cuando se enfrentan a grafías como las que se encuentran en los documentos inquisitoriales, a la vez que las plataformas comerciales mantienen su opacidad en multitud de aspectos tales como los conjuntos de entrenamiento que utilizan. Partiendo de estas premisas, la presente comunicación formulará recomendaciones prácticas que permitan entrenar modelos HTR abiertos, trazables y específicos para la investigación con textos inquisitoriales, sea cual sea el objeto o alcance de la misma.

Dichas recomendaciones seguirán las siguientes pautas metodológicas:

- Delimitación y adquisición del corpus.
- Establecimiento de las normas de transcripción.
- Segmentación y diseño de página.
- Construcción de la “verdad fundamental”.
- Selección y entrenamiento del modelo.
- Evaluación iterativa y control de calidad.
- Difusión, preservación y ética de la infraestructura.

Conclusiones

Desconocer las dinámicas que vertebran las infraestructuras digitales, desde la captura del facsímil hasta la métrica que valida una red neuronal, equivale a renunciar al rigor epistemológico. Los expedientes inquisitoriales, por su complejidad intrínseca, obligan a adoptar flujos de trabajo que expongan cada decisión técnica: resolución, segmentación, normalización, parámetros de entrenamiento, métricas y licencias. Solo con esta transparencia es posible comparar resultados entre equipos, replicar experimentos y, sobre todo, construir modelos que sirvan a la historiografía, la lingüística y la ciencia documental sin reproducir sesgos ocultos en algoritmos cerrados.

Los lineamientos que se presentarán demostrarán que la construcción de un modelo HTR útil no depende de un milagro tecnológico, sino de la combinación disciplinada de Paleografía, Diplomática, Lingüística e ingeniería de datos, entre otros campos. Aplicados con coherencia, permitirían transformar reproducciones fotográficas obtenidas en las instituciones archivísticas en datos legibles por máquina sin obviar ni eliminar el contexto diplomático y archivístico que les da sentido. Así se potenciaría una investigación que permita aproximaciones interdisciplinares, pero sin que se sacrifique la metodología científica que debe ser inherente a cualquiera de los procesos que se describirán.

Bibliografía

BARROSO, J., e SÁNCHEZ DE BUSTOS, J. (1993). Propuestas de transcripción para textos del XV y Siglos de Oro. *Estado actual de los estudios sobre el Siglo de Oro: actas del II Congreso Internacional de Hispanistas del Siglo de Oro*, Vol. 1, 1993, ISBN 84-7481-740-4, págs. 161-178, 161-178. https://dialnet.unirioja.es/servlet/articulo?codigo=909124

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.

BISHOP, C. M. (2006). *Pattern recognition and machine learning*. Springer.

CHAGUÉ, A., y CLÉRICE, T. (2022, junio 23). Sharing HTR datasets with standardized metadata: The HTR-United initiative. *Documents anciens et reconnaissance automatique des écritures manuscrites*. https://hal.inria.fr/hal-03703989

CHAGUÉ, A., CLÉRICE, T., y ROMARY, L. (2021, noviembre 15). HTR-United: Mutualisons la vérité de terrain ! *DHNord2021 - Publier, partager, réutiliser les données de la recherche : les data papers et leurs enjeux*. https://hal.science/hal-03398740

GALENDE DÍAZ, J. C. (2001). El proceso inquisitorial a través de su documentación: Estudio diplomático. *Espacio, tiempo y forma. Serie IV. Historia Moderna*, 14, 491-518.

GILLE LEVENSON, M. (2023). Dataset for late medieval Castilian text recognition [Dataset]. HTR-United. https://doi.org/10.5281/zenodo.7386489

GRÜNING, T., LEIFERT, G., STRAUß, T., MICHAEL, J., y LABAHN, R. (2019). A Two-Stage Method for Text Line Detection in Historical Documents. *International Journal on Document Analysis and Recognition (IJDAR)*, 22(3), 285-302. https://doi.org/10.1007/s10032-019-00332-1

IIIF Consortium. (s. f.). *IIIF Resource Guides*. IIIF. Recuperado 14 de agosto de 2023, de https://iiif.io/guides/

KIESSLING, B., TISSOT, R., STOKES, P., y STOKL BEN EZRA, D. (2019). eScriptorium: An Open Source Platform for Historical Document Analysis. *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 19-19. https://doi.org/10.1109/ICDARW.2019.10032

PINCHE, A. (2022a). *Guide de transcription pour les manuscrits du Xe au XVe siècle*. https://hal.science/hal-03697382

PINCHE, A. (2022b). *HTR Models and genericity for Medieval Manuscripts*. https://hal.archives-ouvertes.fr/hal-03736532

SANTIAGO MEDINA, B. (2019). Contra Lutero y sus secuaces: El procedimiento inquisitorial en el siglo XVI y su reflejo documental. En C. BORREGUERO BELTRÁN y A. RETORTILLO ATIENZA (coords.), *La memoria de un*

hombre: El burgalés Francisco de Enzinas en el siglo XVI y su reflejo documental (pp. 203-218). Burgos: Universidad de Burgos.

SANTIAGO MEDINA, B. (2016). *La burocracia inquisitorial: Escrituras y documentos* (tesis doctoral). Universidad Complutense de Madrid. https://hdl.handle.net/20.500.14352/27359

SANTIAGO MEDINA, B. (2024). Los expedientes de procesos de fe de las inquisiciones americanas en el Archivo Histórico Nacional: Historia archivística y estudio diplomático. En J. C. GALENDE DÍAZ y N. ÁVILA SEOANE (coords.), *Memoria escrita de Ultramar en la Villa y Corte: Fondos documentales madrileños relativos a Hispanoamérica* (pp. 333-389). Madrid: Universidad Complutense de Madrid y Asociación de Amigos del Archivo Histórico Nacional.

STRÖBEL, P. B., CLEMATIDE, S., VOLK, M., y HODEL, T. (2022). Transformer-based HTR for Historical Documents (No. arXiv:2203.11008). *arXiv*. https://doi.org/10.48550/arXiv.2203.11008

Modelizar para compartir. CÁLICE y la ecología de datos abiertos en la historiografía de la censura inquisitorial

Mathilde Albisson
FNRS –
Université Catholique de Louvain

Palabras clave: Censura inquisitorial; modelización de datos; vocabularios controlados; interoperabilidad; patrimonio digital
Keywords: Inquisitorial censorship; data modeling; controlled vocabularies; interoperability; digital heritage

La presente contribución se encuentra vinculada temática y estructuralmente a las que asimismo presentan A. Echavarría y B. Santiago Medina. Las tres comunicaciones parten de la comprensión del espacio digital como un espacio de modelos y de modelización en el que se debe tener en cuenta las necesidades epistemológicas de los investigadores para la constitución de autoridad de las áreas del conocimiento. La transformación digital de la investigación histórica ha generado un paisaje fragmentado de recursos que, paradójicamente, dificulta el acceso al conocimiento que pretenden democratizar. Esta situación revela una contradicción fundamental entre la promesa democratizadora de las humanidades digitales y la realidad de silos de información que replican las barreras tradicionales del acceso archivístico.

El desarrollo de CÁLICE (Catálogo digital de los libros censurados por la Inquisición española en el siglo XVII) surge como respuesta a esta problemática, no solo como herramienta de investigación sino también como propuesta metodológica para una ecología de datos verdaderamente abierta. La base de datos, centrada en los libros censurados por la Inquisición española en la Edad Moderna, se concibe desde su origen como un ejercicio de modelización reflexiva donde las decisiones técnicas responden directamente a preguntas de investigación historiográfica, estableciendo un diálogo productivo entre las necesidades de los investigadores y las posibilidades del medio digital.

La complejidad inherente a la documentación censoria inquisitorial presenta desafíos para su representación digital y exige una modelización que trascienda la simple digitalización de fichas catalográficas para capturar la dinámica procesal de la censura. Así pues, a diferencia de los catálogos bibliográficos convencionales o los inventarios archivísticos lineales, CÁLICE genera una red multidimensional de relaciones: entre obras y autores, órdenes religiosas y calificadores, ediciones y expurgaciones, idiomas y prohibiciones, etc.

El contraste con iniciativas previas es revelador. Mientras el catálogo de J. Martínez de Bujanda ofrece una relación alfabética de las obras censuradas por la Inquisición, carece de la dimensión procesal que permite comprender los mecanismos y razones de la censura. Por su parte, el inventario de A. Paz y Meliá, aunque incluye documentos procesales, presenta una organización y distintas carencias que dificultan el análisis sistemático de la documentación y del fenómeno. CÁLICE integra ambas aproximaciones mediante una arquitectura de datos que permita tanto la consulta puntual como el análisis agregado, respondiendo así a las necesidades metodológicas de la nueva historiografía cultural.

I. Principios de modelización

La construcción de CÁLICE se fundamenta en principios metodológicos interrelacionados que configuran su arquitectura conceptual y técnica:

Modelización orientada por preguntas de investigación: Cada campo responde a interrogantes específicos. Por ej., "Periodo del proceso" permite investigar la temporalidad de la censura y sus coyunturas políticas y "Tipo de condena" trasciende la dicotomía prohibido/permitido para capturar la gradación de intervenciones censorias. Esta aproximación garantiza que la estructura de datos emerja de las necesidades epistémicas de la disciplina y no de imposiciones tecnológicas externas.

Granularidad progresiva: La base de datos opera en múltiples niveles de detalle, desde la identificación básica de obras hasta la documentación completa del proceso censorio. Esta estructura permite diferentes tipos de usuarios y preguntas: desde el investigador que busca verificar si una obra específica fue censurada hasta el especialista que analiza patrones en los mecanismos de calificación. La granularidad se implementa mediante un sistema de capas informativas que se activan según las necesidades de consulta.

Contextualización procesal: CÁLICE reconoce que la censura no es un acto puntual sino un proceso complejo. Por ello, integra no solo el resultado final (prohibición, expurgación, etc.) sino toda la cadena documental: denuncias, calificaciones, sentencia. Esta aproximación permite reconstruir las dinámicas institucionales del Santo Oficio y comprender la censura como práctica social y cultural.

II. Arquitectura técnica y decisiones de diseño

La implementación técnica de CÁLICE en SQL responde a consideraciones pragmáticas, pero incorpora elementos de reflexión sobre estándares e interoperabilidad:

Estructura relacional reflexiva: El modelo entidad-relación mapea las complejidades del proceso censorio mediante tablas interconectadas que representan: obras (con sus múltiples ediciones), autores (con su afiliación

religiosa), procesos (con sus fases documentales), agentes (denunciantes, calificadores), y resoluciones (con sus tipologías específicas).

Vocabularios controlados emergentes: El desarrollo de CÁLICE ha generado la necesidad de vocabularios controlados para campos como "Tipo de condena" o "Materias". Dichos campos se han refinado mediante el uso y ahora constituyen propuestas de estandarización para el campo. La experiencia demuestra cómo la práctica de catalogación genera conocimiento taxonómico valioso para la comunidad investigadora.

Interfaz como mediación epistémica: El diseño del buscador de CÁLICE no es neutral, sino que materializa decisiones sobre qué preguntas son relevantes y cómo deben formularse. La posibilidad de combinar criterios de búsqueda permite construcciones complejas de corpus que responden a hipótesis de investigación específicas. La interfaz se convierte así en instrumento metodológico.

III. Gestión de la complejidad documental

CÁLICE aborda varios desafíos específicos de la documentación inquisitorial:

Multiplicidad de ediciones: Una misma obra puede aparecer en múltiples ediciones con diferentes destinos censores. CÁLICE permite rastrear estas variaciones manteniendo la identidad conceptual de la obra mientras documenta las especificidades de cada edición.

Autoría problemática: Las obras heterodoxas frecuentemente circulaban bajo seudónimos o atribuciones falsas. La base de datos incorpora campos para autorías disputadas y permite enlaces entre identidades múltiples de un mismo autor.

Geografía de la censura: Los campos de lugar de edición e impresor permiten mapear la geografía de la producción y circulación de libros prohibidos, revelando redes de impresión heterodoxa o clandestina.

Resultados obtenidos

I. Cobertura y alcance actual

La primera fase de CÁLICE, actualmente operativa, incluye todos los libros censurados cuyo expediente procesal se conserva en archivos españoles. Esto representa:

Corpus documental: Integración de fondos del Archivo Histórico Nacional y de otras instituciones patrimoniales españolas;

Cobertura temporal: Foco en el s. XVII, período de máxima actividad censora y de consolidación de los mecanismos de control libresco;

Tipología documental: Desde simples denuncias hasta procesos complejos con múltiples calificaciones, permitiendo análisis diferenciados según la intensidad del escrutinio inquisitorial.

II. Patrones emergentes y hallazgos metodológicos

El uso de CÁLICE ha revelado patrones que permanecían ocultos en la documentación dispersa:

Temporalidades de la censura: El campo "Periodo del proceso" ha permitido identificar coyunturas específicas de intensificación censora correlacionadas con crisis políticas o religiosas.

Redes de denuncia: El análisis cruzado de denunciantes y obras revela redes de vigilancia intelectual que operaban en ámbitos específicos.

Geografías del disenso: La correlación entre lugares de impresión y tipos de condena muestra la especialización de ciertos centros editoriales en literatura heterodoxa y las rutas de penetración en territorios hispánicos.

III. Impacto en la investigación

CÁLICE ha facilitado nuevas aproximaciones metodológicas:

Construcción de corpus temáticos: Investigadores han podido construir corpus específicos que antes requerían años de trabajo archivístico.

Análisis comparativo: La estructura estandarizada permite comparaciones sistemáticas entre períodos o categorías de obras, abriendo nuevas preguntas sobre la evolución de los criterios censores.

Identificación de lagunas: La base de datos también revela ausencias significativas en la documentación, orientando futuras investigaciones archivísticas.

IV. Hacia una ecología de datos abiertos

El desarrollo de CÁLICE ha evidenciado tanto las posibilidades como las limitaciones de su arquitectura actual:

Tensiones entre especificidad y generalización: La modelización específica para la censura inquisitorial española dificulta la integración con bases de datos de censura de otros contextos (Roma, Portugal). Esto plantea la pregunta: ¿hasta qué punto la especificidad histórica debe determinar la arquitectura de datos?

Evolución de vocabularios: Los vocabularios controlados desarrollados para CÁLICE han evolucionado con el uso, revelando categorías inicialmente no previstas. Esto sugiere la necesidad de arquitecturas flexibles que permitan la evolución controlada de taxonomías.

V. Implicaciones para las políticas de ciencia abierta

El caso de CÁLICE ilustra principios fundamentales para una verdadera ciencia abierta:

Apertura como proceso, no como estado: La transición de una base de datos cerrada a una abierta requiere trabajo continuo de documentación,

estandarización y mantenimiento. Las políticas de financiación deben reconocer estos costos ocultos.

Comunidades de práctica: La estandarización efectiva emerge de comunidades de usuarios, no de imposiciones *top-down*. CÁLICE demuestra cómo los vocabularios controlados más útiles son aquellos refinados mediante uso real.

Sostenibilidad mediante federación: En lugar de grandes proyectos monolíticos, el futuro está en bases de datos especializadas pero interoperables, cada una mantenida por comunidades expertas pero accesible globalmente.

En conclusión, CÁLICE representa un modelo de desarrollo de bases de datos historiográficas que privilegia las preguntas de investigación sobre las soluciones tecnológicas predeterminadas. Su evolución desde una herramienta específica hacia una propuesta de estándares abiertos ilustra cómo la práctica investigadora puede generar infraestructuras de conocimiento verdaderamente comunitarias. La experiencia acumulada demuestra que la modelización de datos históricos no puede separarse de la reflexión epistemológica sobre qué constituye el objeto de estudio. En el caso de la censura inquisitorial, esto significa reconocer que no catalogamos simplemente "libros prohibidos" sino complejos procesos de control del conocimiento. Esta comprensión debe traducirse en arquitecturas de datos que capturen tal complejidad sin sacrificar usabilidad. El contraste con bases de datos cerradas refuerza la urgencia de adoptar principios de ciencia abierta como necesidad epistemológica. Solo mediante datos accesibles y reutilizables podremos construir el conocimiento acumulativo y colaborativo que las humanidades digitales prometen. CÁLICE, en su modestia y especificidad, apunta hacia ese horizonte de posibilidad.

Bibliografía

CODD, E. F. (1970), «A Relational Model of Data for Large Shared Data Banks», *Communications of the ACM*, 13(6), pp. 377–387, recurso en línea: https://doi.org/10.1145/362384.362685 [Acceso: 14/10/2025].

CODINA, L. (2003), «La web semántica: una visión crítica», *El profesional de la Información*, 12(2), pp. 149–152.

CANDELA ROMERO, G. (2019), *Publicación y enriquecimiento semántico de datos abiertos en bibliotecas digitales*, tesis doctoral, Universitat d'Alacant.

EIDE, Ø.; & ORE, C.-E. (2007), *Mapping from TEI to CIDOC-CRM: Will the New TEI Elements Make any Difference?*, recurso en línea: https://app.cristin.no/results/show.jsf?id=225146 [Acceso: 14/10/2025].

HUDON, M. (2009), *Guide pratique pour l'élaboration d'un thésaurus documentaire*, Montréal, ASTED.

HUDON, M. (2012), «ISO 25964: pour le développement, la gestion et l'interopérabilité des langages documentaires», *Documentation et bibliothèques*, 58(3), pp. 130–140.

LONGLEY, D. (2024), *RDF Dataset Canonicalization: A Standard RDF Dataset Canonicalization Algorithm*, W3C Recommendation, World Wide Web Consortium (W3C), recurso en línea: https://www.w3.org/TR/rdf-canon/ [Acceso: 14/10/2025].

MARTÍNEZ DE BUJANDA, J. (2016), *El Índice de libros prohibidos y expurgados de la Inquisición española (1551–1819): evolución y contenido*, Madrid, BAC.

PAZ Y MELIÁ, A. (1947), *Papeles de Inquisición: catálogo y extractos*, Madrid, Patronato del Archivo Histórico Nacional.

SZABADOS, A.-V.; & LETRICOT, R. (2012), «L'ontologie CIDOC CRM appliquée aux objets du patrimoine antique», *3e Journées d'Informatique et Archéologie de Paris – JIAP 2012*.

TEI y RDF aplicados a la documentación de los procesos inquisitoriales

Andrés Felipe Echavarría Peláez[1]
CNRS Consorcio ARIANE – IR* Huma-Num

Palabras clave: Inquisición española; editorialización; TEI; SKOS; principios FAIR; ciencia abierta.
Keywords: Spanish Inquisition; editorialization; TEI; SKOS; FAIR principles; open science.

Descripción del problema de investigación

La documentación procesal del Santo Oficio reúne miles de documentos producidos entre los siglos XV y XVIII. Su complejidad formal y procedimental hace que los repertorios archivísticos y las bases de datos historiográficas convencionales transformen los expedientes en inventarios lineales, incapaces de mostrar las dinámicas internas de cada causa. Las descripciones ISAD(G) (NEDA en el contexto español) o las bases "relacionales" simplifican este entramado a registros por expediente, lo que diluye los contenidos textuales de las piezas de archivo, las funciones documentales y las relaciones internas generadas por el propio tribunal de la fe en la instrucción de sus causas. El resultado es una pérdida de contexto que repercute en la interpretación de los actores, la secuencia cronológica y la terminología (tanto documental como jurídica). Estas limitaciones son un obstáculo para los estudios comparativos y dificultan el diálogo entre disciplinas que abordan la documentación Inquisitorial desde perspectivas transversales a la historiografía social, la historia del derecho, la lingüística y la diplomática.

Frente a lo que precede, se propone un modelo compuesto para la estructuración digital de los contenidos de los expedientes procesales, que integra un esquema de anotación TEI personalizado y un tesauro bilingüe estructurado en RDF/SKOS. El objetivo es describir la organización interna de los expedientes procesales y habilitar, a partir de datos interoperables, nuevas líneas de análisis. La noción de editorialización —entendida como el conjunto de dinámicas constituyentes del espacio digital— enmarca la propuesta metodológica y garantiza la trazabilidad de cada intervención sobre las fuentes.

[1] Ingeniero de investigación

Metodología

El dispositivo metodológico se desarrolla sobre la base de dos propuestas acopladas. Un tesauro documental y un esquema de anotación fundado en la (*Text Encoding Initiative*). Clasificación tiene en cuenta características de forma, fondo y función de las documentaciones procesales sin imponer un marco analítico cerrado o de tratamientos propios del lenguaje natural o el discurso. En el caso de la modelización presentada, la forma hade entenderse como el conjunto de características formales que permiten situar el documento dentro de un tipo diplomático (acta, mandato, sentencia, poder, etc.). Por fondo, la modelización busca realizar una doble asociación de cada una de las piezas documentales por tipo a los procedimientos específicos de los que pueden provenir. Al mismo tiempo, el fondo se asocia a la función de la documentación dentro del expedientes específicos dado que la rede de declaraciones que componen los expedientes procesales de la inquisición pueden hacer caer a los investigadores en interpretaciones documentales que desnaturalizan el origen de los documentos. Por último, y en este caso desde una perspectiva de *corpus* documental, la función, debe entenderse como el conjunto de colecciones que, si bien no ordenan jerárquicamente los tipos de documentos procesales dentro del expediente inquisitorial, permiten clasificar los contenidos por funciones generales dentro del marco jurisdiccional del tribunal de la fe.

a) Un esquema de anotación estructural basado en la TEI P5
   El esquema TEI define tres niveles jerárquicos: proceso, fase procesal y pieza documental. Además, este esquema está diseñado para contener procesos compuestos que compilen antiguas causas y habilite la integración de documentos producidos después de la conclusión de esta. El encabezado <teiHeader> almacena metadatos de procedencia y las taxonomías de referencia. La anotación interna marca entidades (personas, lugares, instituciones) y estructuras textuales relevantes sin imponer un protocolo de transcripción cerrado, lo que permite transcripciones diplomáticas o normalizadas, según los objetivos de los investigadores.
   Para facilitar la adopción del esquema de anotación, se ha optado por representar los identificadores de las entidades nombradas con valores literales —una etiqueta de texto concatenada— con el fin de reducir al mínimo la barrera de entrada tanto para las personas que anotan como para los programas que procesarán después los archivos TEI. Esta solución permite que los conjuntos de datos permanezcan rastreables, aunque se transformen los criterios de catalogación dado que estos literales pueden operar (conciertas limitaciones ligadas a lo homonimia) como claves que sirven de puente entre versiones.
b) Un tesauro bilingüe para el control terminológico
   El Tesauro de Tipologías Documentales de la Inquisición Española organiza Fases procesales, procedimientos inquisitoriales y tipologías

diplomáticas en castellano y francés. Se estructura en SKOS para permitir la alineación con otros vocabularios de la Web Semántica. Y se extiende a los espacios ISO-Thes y UNESKOS para permitir su evolución en conjunto o por partes (Micro tesauros). Cada concepto posee URI estable, etiquetas preferidas, variantes y relaciones jerárquicas (BT/NT). El tesauro se publica para ser usado y gestionado en la plataforma de gestión de vocabularios controlados Opentheso. Para efectos de usabilidad, hemos realizado una traducción al castellano de la plataforma Opentheso, así como la publicación progresiva de toda la documentación en español. Sin embargo, este trabajo se desprende de toda voluntad de monopolizar la gestión o evolución de las herramientas, razón por la cual se ha previsto también un sistema de versiones citables con DOIs independientes para la reutilización de su código en cuatro serializaciones (RDF/XML, Turtle, JSON, JSON- LD), lo que cumple la dimensión Interoperable.

Resultados obtenidos

Esquema TEI para la validación de la anotación de expedientes inquisitoriales: disponible en DTD, XSD, RNG, RNC y ODD; incluye documentación HTML generada automáticamente. Valida 100 % del prototipo y que requiere pruebas de cobertura en pruebas de unidad.

a) **Tesauro Inq_Thes:** 103 conceptos, 361 etiquetas multilingües, 905 relaciones jerárquicas[2], Versión pública DOI permanente (Nakala) y versión en GitLab[3].
b) **Proceso de Pedro de Cazalla anotado en XML/TEI:** El prototipo de proceso se fundamenta en la transcripción de Marcelino Menéndez Pelayo del proceso de fe del tribunal de Valladolid contra Pedro de Cazalla acusado de luteranismo. La anotación de este expediente se realizó de forma completamente manual, contando 4268 entidades nombradas de las cuales 1099 son expresiones referenciales (formas pronominales, descripciones definidas, sintagmas nominales anafóricos, etc.). El proceso se compone de 168 piezas documentales que se encuentran clasificadas por tipología (correspondiente con el tesauro) así como por su condición de extracto o de completud (así como de la posición y de la secuencia procesal dentro del expediente) en el expediente. Los metadatos en el encabezado cuentan con una serie de bases de datos de personas, organizaciones, establecimientos y eventos que cuenta más de tres centenares de entidades identificadas y alineadas por más de 300 enlaces externos a registros de autoridad o páginas de interés para la historia religiosa del cristianismo católico.

[2] https://opentheso.huma-num.fr/?idt=Inq_Thes
[3] https://gitlab.huma-num.fr/aechavarria/annexesthese

c) **Scripts de soporte:** Con el fin de asegurar la evolución coherente del esquema de anotación y su alineación continua con el vocabulario controlado, el repositorio incorpora cadenas de tratamiento automatizadas que convierten expresiones SKOS en taxonomías TEI internas y en módulos ODD destinados a normalizar las formas nominales de las tipologías documentales. Se incluyen, asimismo, flujos para la extracción de cuestionarios procesales, el cálculo del calendario litúrgico y la generación de identificadores persistentes. La totalidad de los scripts se distribuyen en Zenodo con DOI propio y licencia CC-BY-NC-SA-2.0-FR[4]. La documentación ODD resultante se publica de manera automática con una cadena de tratamiento que transforma el esquema en un conjunto de *Guidelines* HTML y los despliega en un sitio Git Pages[5], de modo que cualquier actualización validada en el repositorio se refleja inmediatamente en la guía en línea sin intervención manual.

Evolución controlada de la modelización

Para conservar la coherencia del esquema frente a la incorporación de nuevas tipologías documentales, el repositorio incorpora un conjunto comentado de tipologías diplomáticas generales. Dicho repertorio puede adoptarse de inmediato en la anotación XML/TEI de colecciones inéditas, mientras se valida la integración de las nuevas clases en el tesauro. Así pues, el tesauro, además de sus versiones en SKOS/RDF, JSON-LD, Turtle y RDF/XML parafacilitar la reutilización en cualquier entorno; también admite la proposición de términos adicionales y de cambios en su estructura mediante el flujo estándar de la plataforma Opentheso. Este espacio colaborativo permite el intercambio académico en torno a las versiones estables. En cualquier caso, los scripts de conversión a ODD y TEI regeneran de manera automática las actualizaciones correspondientes, siempre que los RDF modificados (en otros entornos) mantengan la conformidad con la recomendación SKOS.

Reproducibilidad

Los scripts de soporte y las rutinas de análisis documental que acompañan a la modelización constituyen una declaración política a favor de la ciencia abierta: están escritos íntegramente en lenguajes libres (Python, XSLT) y operan sobre formatos estandarizados (TEI-XML, RDF, CSV), de modo que cualquier persona —sin depender de cajas negras corporativas— puede reconstruir paso a paso el flujo que va desde las fuentes inquisitoriales hasta los conjuntos de datos derivados. Cada transformación está versionada, documentada y empaquetada con ejemplos mínimos que permiten lanzar los scripts tal cual, verificar las salidas y, si se desea, modificar las funciones para producir

[4] Creative Commons Attribution-NonCommercial-ShareAlike 2.0 France.
[5] https://aechavarria.gitpages.huma-num.fr/annexesthese/

interpretaciones basadas en nuevas preguntas de investigación. Esta arquitectura modular garantiza la rastreabilidad de todos los cálculos, refuerza la posibilidad de réplica en distintos entornos y pone en práctica la idea de que la reproducibilidad no es un lujo metodológico sino un derecho epistémico: solo cuando los procedimientos son legibles, ejecutables y susceptibles de ser cuestionados colectivamente puede la investigación escapar de los monopolios de software y servir al bien común.

Conclusiones

La propuesta de modelización para los expedientes inquisitoriales resuelve la representación fragmentaria de los procesos mediante un esquema que preserva su lógica interna y al mismo tiempo favorece la exploración computacional. La articulación TEI-RDF, el control terminológico en SKOS y la publicación conforme a los principios FAIR convierten los expedientes en datos de bien común, listos para nuevas líneas de investigación. La documentación detallada, los ejemplos anotados y los scripts distribuidos con DOI garantizan la autonomía de futuros usuarios y la sostenibilidad técnica de la modelización. Esta última se concibe como un flujo abierto: los datos se generan en función de preguntas de investigación concretas. Las cadenas de tratamiento propuestas permanecen flexibles para garantizar la actualización automática (aunque supervisada) en función de necesidades específicas sin necesidad de reelaborar modelos desde cero. Esta arquitectura orientada a las características endógenas de los objetos documentales limita el sesgo introducido por decisiones de modelización y mantiene el margen interpretativo necesario para que los investigadores puedan adaptar el flujo a sus propios objetivos analíticos.

Bibliografía

ECHAVARRÍA, A. (2021), «Éditorialisation des procès inquisitoriaux espagnols: Modélisation TEI, description et vocabulaires contrôlés», *thèse en préparation*, Université Paul-Valéry Montpellier 3, recurso en línea: https://theses.fr/s306204 [Acceso: 14/10/2025].

ECHAVARRÍA, A. (2025), «Proceso de fe de Pedro de Cazalla», *XML Standard, TEI prototype*, NAKALA (Huma-Num – CNRS), recurso en línea: https://doi.org/10.34847/nkl.aaeej9jp [Acceso: 14/10/2025].

Inteligencia Artificial aplicada al análisis de corpus en prensa musical histórica: metodologías computacionales para el estudio de la crítica musical en español

María Palacios Nieto
Universidad de Salamanca

Palabras clave: Musicología digital; inteligencia artificial; prensa histórica; procesamiento de lenguaje natural; crítica musical.
Keywords: Digital musicology; artificial intelligence; historical journalism; natural language processing; music criticism.

La prensa musical histórica en España constituye un corpus cultural de gran valor para comprender la vida musical desde finales del siglo XVIII. Sin embargo, la calidad deficiente de los OCR en hemerotecas digitales nacionales impide su explotación mediante técnicas actuales de procesamiento computacional, limitando su análisis sistemático desde las humanidades digitales. A pesar del trabajo pionero de instituciones como la Biblioteca Nacional de España en digitalización, es necesario un esfuerzo en la limpieza del "ruido" de OCR para procesar estos documentos eficientemente con tecnologías actuales.

Adicionalmente, esta propuesta reflexiona sobre las dificultades de los grandes modelos de lenguaje existentes, entrenados predominantemente con textos actuales, para analizar estos materiales históricos de forma precisa y sistemática. Se pone el foco también en el hecho de que gran parte del contenido histórico-musical se encuentra disperso entre diferentes fuentes (prensa periódica, revistas culturales y musicales) sin explorar conjuntamente mediante métodos computacionales.

Esta propuesta surge del proyecto "Leximus: léxico y ontología de la música en español" (PID2022-139589NB-C33), con sede en la Universidad de Salamanca, coordinado desde el Instituto Complutense de Ciencias Musicales junto con la Universidad de La Rioja. El grupo USAL está formado por expertos en musicología de distintos periódicos históricos, permitiendo la revisión precisa de los elementos explorados de manera computacional con IA. El proyecto en su sede USAL se centra en desarrollar metodologías de inteligencia artificial para la creación, análisis e interpretación de datasets de fuentes histórico- musicales a partir de la prensa.

La investigación se enfoca específicamente en el periodismo musical desde finales del siglo XVIII hasta la primera mitad del siglo XX, explorando las limitaciones de los Grandes Modelos de Lenguaje (LLMs) actuales y los

problemas del Reconocimiento Óptico de Caracteres (OCR) para procesar textos históricos en español, con el fin de estudiar la terminología musical especializada y los diferentes formatos de publicación específicos del período.

Metodología

Nuestra hipótesis principal de trabajo sostiene que los métodos de IA, adaptados a las particularidades del discurso musical histórico en español, permiten descubrir patrones de representación invisibles al análisis tradicional, especialmente en género, recepción crítica y evolución terminológica.

El enfoque metodológico combina así procesamiento de lenguaje natural con análisis de discurso tradicional, desarrollando metodologías efectivas de IA para textos periodísticos- musicales en español que presentan dificultades particulares en vocabulario especializado y convenciones lingüísticas evolutivas de la crítica musical a lo largo de casi 150 años. Esta perspectiva se alinea con la "analítica cultural computacional" de Berry y Fagerjord (2017) y el modelo de distant reading de Manovich (2020), aplicados al análisis léxico- discursivo de fuentes musicales, combinando PLN con lectura crítica musicológica según principios de lingüística computacional para textos históricos (Piotrowski, 2012).

Corpus de trabajo

El análisis aborda cinco publicaciones fundamentales representativas de distintos momentos del periodismo musical español:

1. **Diario de Madrid (1788-1825)**: Periódico que abordó sistemáticamente eventos musicales, especialmente operísticos. (Corpus con más de 33.000 noticias sobre música)
2. **La Iberia Musical y Literaria (1842-1846)**: Publicación pionera que representa la primera convergencia especializada de crítica musical y literaria en la prensa española. (Corpus de 159 números con más de 500.000 palabras)
3. **Crónica de la música (1878-1882)**: Revista que publica crónicas de teatro y conciertos de ópera y zarzuela, artículos sobre instrumentos, técnicas e historia musical. (Corpus de 223 números con más de un millón de palabras)
4. **Revista Musical de Bilbao**: Transformada en 1914 en Revista Musical Hispano- americana. Ambas fueron fundamentales en su época, donde escribieron principales musicógrafos e historiadores del momento. (Corpus de 78 números con más de un millón de palabras)

Proceso técnico

Con este corpus de millones de palabras, el proyecto se estructura en cuatro fases principales:

Fase de recolección de datos: Establecimiento de la base digital. Acceso a prensa masivamente a través de Scripts de descarga masiva-ética,

desarrollados por Rafael Vidal.

Mejora de OCR: Vertex AI permitió trabajo por lotes desde Google Cloud, facilitando el procesamiento de grandes corpus. El reconocimiento de Vertex AI para extracción de información y Gemini para lectura de OCR resultó especialmente positivo para la mejora de las palabras individuales. La mejora resulta clara, por ejemplo, en la búsqueda de "Falla" (compositor) en la revista Ondas, la hemeroteca de la BNE obtuvo 1.165 resultados, mientras que con el OCR mejorado en LexiMus se obtuvieron 2.717. El problema derivó en la detección de columnas, mejorado con la fragmentación de eScriptorium. Actualmente se procesa el corpus con el modelo CATMuS-Print [Large], disponible en Zenodo, optimizándolo para lectura más precisa de prensa en español.

Construcción y anotación de datasets: Transformación del contenido digitalizado en estructurado y legible por máquina. Datos y estadísticas analíticas publicados de forma preliminar en https://leximususal.github.io/principal/. Claude Code ha sido fundamental al permitir un trabajo directo en terminal y sin límite de corpus.

Selección de IA: Se testaron distintos modelos (Claude, ChatGPT, Deepseek) con corpus validados por expertos humanos, evaluando la precisión en extracción de entidades musicales y análisis discursivo. Claude Sonnet 3.7 ofreció el menor índice de errores de interpretación, motivando su elección como herramienta principal. Las pruebas se realizaron con corpus controlados, validados por humanos antes del procesamiento por IA. Actualmente se trabaja con Sonnet 4 y Opus 4 para evaluar mejoras significativas en extracción y análisis de datos.

Resultados esperados y avances preliminares

Los enfoques de análisis han mapeado patrones de influencia entre publicaciones, artistas y obras, revelando estructuras sociales e intelectuales que dieron forma al discurso musical. Mediante análisis con IA se identificaron y mapearon principales compositores citados, datos de género sobre presencia de hombres y mujeres en las fuentes (y actividades relacionadas), conexiones entre comunidades y estructuras de redes evolutivas.

Estos datos revelaron nuevas líneas de investigación, como la prevalencia de mujeres frente a hombres citadas en la prensa especializada del XIX, debido al peso mediático de las divas de ópera, prevalencia que desaparece (y se invierte) en la primera mitad del siglo XX, al caer el peso de la ópera frente a otras prácticas musicales como conciertos sinfónicos.

El proyecto se encuentra en fase de "laboratorio", explorando posibilidades de análisis computacional de este extenso corpus textual, aún incompleto y con errores de OCR (especialmente en división de columnas), pero donde ya se pueden extraer ideas principales y tendencias que solo pueden identificarse con análisis estadístico- computacional de miles de documentos y millones de

palabras.

La propuesta reflexiona sobre el análisis de patrones discursivos en la crítica musical española histórica, incluyendo la transformación y evolución de principales tipos musicales en cada época (ópera, música sinfónica, de cámara), patrones de recepción crítica, principales compositores nombrados, espacios citados, y relación entre crítica y desarrollo de vida musical, mediante análisis estadístico-computacional con IA (Claude Code). Se implementan metodologías de análisis computacional especializadas que identifican, cuantifican y analizan sistemáticamente patrones de representación y omisión en el discurso histórico de la prensa musical, con enfoque en disparidades de género, exclusiones basadas en clase y comunidades marginalizadas.

Consideramos que las herramientas interactivas de visualización para explorar patrones cronológicos y temáticos podrán permitir a investigadores, educadores y público general interactuar con este patrimonio cultural actualmente disperso, incluyendo visualizaciones de línea temporal que muestran la evolución de noticias musicales.

Impacto y proyección

El proyecto contribuye significativamente al avance tecnológico y acceso democratizado al patrimonio musical español, permitiendo a académicos y cualquier persona interesada a explorar y analizar publicaciones histórico-musicales de forma directa y sencilla, dentro del compromiso cívico y social de las Humanidades Digitales. La plataforma digital integrará datasets con gráficos de visualización y estadísticas, haciendo este patrimonio más accesible.

Se busca generar transferencia tecnológica directa hacia las principales instituciones patrimoniales públicas, estableciendo colaboración con la Hemeroteca Digital de la Biblioteca Nacional de España (BNE), donde se encuentran las principales revistas y recursos analizados en LexiMus. Las metodologías de IA desarrolladas se compartirán como recursos abiertos para que otras instituciones interesadas ofrezcan experiencias de acceso más dinámicas, intuitivas y contextualmente enriquecidas con fondos digitalizados.

Esta colaboración facilitaría búsquedas inteligentes que trasciendan limitaciones de consultas por palabras clave, permitiendo la generación automática de resúmenes, etiquetado temático y conexiones intertextuales que transformarían la interacción actual con revistas musicales digitalizadas. El proyecto contribuye directamente a objetivos de preservación y democratización del acceso al patrimonio musical histórico, estableciendo un modelo escalable de innovación tecnológica aplicada a la gestión que beneficiará tanto a prensa musical especializada como al conjunto de publicaciones periódicas históricas españolas.

Bibliografía

BERRY, D. M.; & FAGERJORD, A. (2017), *Digital Humanities: Knowledge and Critique in a Digital Age*, Polity Press.

CASCUDO, T.; & PALACIOS, M. (eds.) (2013), *Los señores de la crítica: periodismo musical e ideología del modernismo en Madrid*, Doble Jota.

DE BOLLA, P. (2023), *Explorations in the Digital History of Ideas*, Cambridge University Press.

MANOVICH, L. (2020), *Cultural Analytics*, MIT Press.

McCARTY, W. (2005), *Humanities Computing*, Palgrave Macmillan.

PAJÓN, A. (2025), «Repensando el discurso de la prensa desde la musicología», *El oído Pensante*, vol. 13, n.º 1.

PIOTROWSKI, M. (2012), *Natural Language Processing for Historical Texts*, Morgan & Claypool.

VAN DIJK, T. A. (2008), *Discourse and Power*, Palgrave Macmillan.

Bases de datos relacionales para el estudio del libro musical en València (1413–1641)

Ferran Escrivà Llorca
Universitat de València
Alex Wingate
Indiana University

La circulación, consumo y apropiación del libro musical en la Edad Moderna ibérica constituye una veta historiográfica apenas explorada en profundidad. A diferencia de la atención recibida por los procesos de producción o edición musical, la información relativa a la posesión, uso y transferencia de libros de música ha sido tradicionalmente fragmentaria, dispersa o tratada de forma colateral dentro de estudios más amplios. En el caso del Reino de Valencia, esta situación se agrava por la escasa sistematización de fuentes primarias —como inventarios postmortem o registros de almonedas— en formato accesible, reutilizable y analizable digitalmente. La presente propuesta tiene como objetivo central presentar un caso práctico de aplicación del Book List Framework (BLF), una estructura de datos interoperable desarrollada para estandarizar y facilitar el tratamiento digital de listados históricos de libros, en el ámbito de la música impresa y manuscrita valenciana entre los siglos XV y XVII.

El proyecto se enmarca en una línea de investigación más amplia dedicada al estudio de la cultura musical urbana en Valencia y otros territorios del Mediterráneo hispánico. Se parte de la premisa de que los libros de música no solo constituyen objetos materiales susceptibles de estudio bibliográfico o filológico, sino que son también instrumentos culturales capaces de vehicular identidades, competencias de alfabetización musical y prácticas devocionales, educativas o lúdicas. En este sentido, la hipótesis de trabajo asume que el análisis digital de los listados de libros musicales permitirá reconstruir con mayor precisión redes de circulación, pautas de consumo y jerarquías de valor simbólico en contextos urbanos concretos.

La investigación se apoya en una base de datos en construcción, denominada Valencia Music Books (VMB), que integra actualmente 41 inventarios redactados entre 1413 y 1641, mayoritariamente en el contexto de la ciudad de Valencia, y que será ampliada progresivamente. Estos inventarios provienen de fondos notariales, judiciales y eclesiásticos y presentan grados variables de detalle. Algunos listados incluyen descripciones mínimas (“un libro de canto”), mientras que otros especifican títulos, formatos o contenidos. La VMB tiene como finalidad registrar y normalizar estos datos, aplicando una estructura basada en el BLF, de modo que las entradas de inventario se vinculen, cuando

sea posible, a manifestaciones o expresiones concretas de obras musicales, siguiendo el modelo FRBR (Functional Requirements for Bibliographic Records) y permitiendo la comparación con otras bases como FBTEE o MEDIATE.

Metodología

La metodología empleada combina técnicas de edición digital, normalización bibliográfica, modelado de datos y georreferenciación. En primer lugar, los documentos son transcritos íntegramente y codificados semánticamente mediante una plataforma relacional abierta (Heurist). Cada listado se introduce como un documento individual dentro de una estructura que distingue entre las siguientes entidades: Caso > Documento > Entrada (registro de un libro). A su vez, las entradas se vinculan con registros bibliográficos estructurados en niveles jerárquicos (Obra, Expresión, Manifestación, Ítem), lo que permite incluir tanto obras identificadas (como tratados musicales o libros litúrgicos impresos) como referencias indeterminadas.

El modelo permite también identificar eventos de descripción múltiple (por ejemplo, una entrada que agrupa más de un ítem) y la aparición reiterada de un mismo ítem en listados distintos. A través de esta arquitectura, se posibilita tanto la exploración detallada de cada fuente como su análisis agregado, algo impensable con los métodos tradicionales de edición en papel.

En paralelo, se realiza una georreferenciación progresiva de los documentos y propietarios, lo que permitirá —en fases futuras del proyecto— generar mapas interactivos de concentración de libros musicales, redes de librerías, o rutas de circulación de repertorios entre comunidades religiosas, educativas y aristocráticas del Reino de Valencia. Todo ello será de utilidad tanto para la historia cultural como para la musicología urbana, la historia del consumo y los estudios sobre alfabetización musical.

Resultados preliminares

La implementación del BLF en VMB ha permitido ya algunos resultados significativos:

1. Sistematización de fuentes dispersas: se han identificado y estandarizado más de 350 entradas bibliográficas sobre libros musicales, procedentes de listados de diversa naturaleza (testamentos, visitas pastorales, inventarios conventuales, etc.). Esta tarea ha permitido detectar títulos recurrentes, formas de catalogación y unidades léxicas específicas empleadas para referirse a los libros musicales.
2. Prototipado de un sistema de descripción común: gracias al uso de FRBR y de eventos de descripción conectados, se ha logrado vincular diversas menciones fragmentarias a manifestaciones concretas (por ejemplo, ejemplares de “Obras de música” de Antonio de Cabezón o

libros de canto llano franciscano), lo que abre la vía a estudios de historia bibliográfica digital comparada.

3. Detección de contextos de uso y perfil del poseedor: la combinación de datos bibliográficos con metadatos contextuales (edad, profesión, género, adscripción institucional) ha permitido esbozar perfiles de usuarios de libros musicales (eclesiásticos, bachilleres, artesanos, etc.), revelando patrones sociales de acceso al capital simbólico musical.
4. Establecimiento de relaciones con otras bases de datos: el uso del BLF ha facilitado la interoperabilidad con otros proyectos como Books in Navarre o IBSO, haciendo posible un diálogo metodológico y técnico más allá de la escala local.
5. Reutilización y visualización de los datos: la base de datos está pensada para ser consultable en línea, con posibilidades de búsqueda cruzada, filtros por cronología o tipología de libros, y descarga parcial de datos, lo que facilitará su integración en estudios posteriores o su uso docente.

Discusión y relevancia en Humanidades Digitales

Desde el punto de vista de las Humanidades Digitales, la propuesta contribuye a varios niveles: a) propone una arquitectura de datos abierta y escalable, que puede ser aplicada a otros contextos geográficos o temáticos; b) demuestra la utilidad del modelado semántico para el estudio de fuentes documentales complejas y no normalizadas; c) se inscribe en el esfuerzo colectivo por superar el aislamiento de bases de datos patrimoniales mediante estructuras interoperables y reutilizables.

Además, la VMB se alinea con los objetivos de preservación y difusión del patrimonio cultural musical, en tanto que permite documentar formas de conocimiento musical que no siempre han dejado huella en las historias canónicas de la música. En este sentido, el proyecto no solo tiene valor documental o bibliográfico, sino también social, ya que ayuda a reconstruir el papel del libro musical en la configuración de comunidades, saberes y prácticas a menudo invisibilizadas.

Conclusión

La implementación del Book List Framework en la base de datos *Valencia Music Books* constituye un ejemplo de aplicación metodológica avanzada en el ámbito de la musicología histórica, la historia del libro y las humanidades digitales. Mediante la digitalización, estructuración y análisis contextualizado de listados históricos de libros musicales, se abren nuevas vías para entender la circulación de repertorios, la alfabetización musical urbana y las redes de consumo cultural en la Valencia de los siglos XV al XVII. Esta propuesta espera

contribuir, así, tanto a los estudios sobre patrimonio bibliográfico musical como al debate sobre estándares y estructuras de datos en humanidades digitales.

Bibliografía

BERGER, P. (1987), *Libro y lectura en la Valencia del Renacimiento* (A. Balanzá Pérez, trad.), València, Institució Alfons el Magnànim.

BURROWS, S. (2015), «Locating the Minister's Looted Books: From Provenance and Library History to the Digital Reconstruction of Print Culture», *Library & Information History*, vol. 31, n.º 1, pp. 1–17. https://doi.org/10.1179/1758348914Z.00000000071

BURROWS, S.; & NURMIKKO-FULLER, T. (2021), «Charting Cultural History through Historical Bibliometric Research. Methods; Concepts; Challenges; Results», en K. Schuster & S. E. Dunn (eds.), *Routledge International Handbook of Research Methods in Digital Humanities*, pp. 109–124, London–New York, Routledge (Routledge International Handbooks). https://doi.org/10.4324/9780429777028

GARCÍA AGUILAR, I.; & CAMPILLO PARDO, A. J. (2023), «Private Libraries in New Spain: A Project in Progress», en R. Jagersma, H. Blom, E. Chayes & A.-M. Hansen (eds.), *Private Libraries and their Documentation, 1665–1830*, pp. 225–245, Leiden–Boston, Brill. https://doi.org/10.1163/9789004542969_012

JAGERSMA, R.; BLOM, H.; & HANSEN, A.-M. (2023), «Goldmines or Minefields? Private Libraries and Their Documentation (1665–1830)», en R. Jagersma, H. Blom, E. Chayes & A.-M. Hansen (eds.), *Private Libraries and their Documentation, 1665–1830*, pp. 1–17, Leiden–Boston, Brill. https://doi.org/10.1163/9789004542969_002

LAHTI, L.; MARJANEN, J.; ROIVAINEN, H.; & TOLONEN, M. (2019), «Bibliographic Data Science and the History of the Book (c. 1500–1800)», *Cataloging & Classification Quarterly*, vol. 57, n.º 1, pp. 5–23. https://doi.org/10.1080/01639374.2018.1543747

SCHUSTER, K.; & DUNN, S. (eds.) (2020), *Routledge International Handbook of Research Methods in Digital Humanities*, 1.ª ed., London–New York, Routledge. https://doi.org/10.4324/9780429777028

WINGATE, A. E. (2024), «Metadata for Digital Collections», *Journal of Education for Library and Information Science*, vol. 65, n.º 1, pp. 108–109. https://doi.org/10.3138/jelis-2022-0077

WINGATE, A. E. (2024), «"No entiende en el Balor de los libros": the Value of Books for Women Owners in Seventeenth-Century Navarre», en E. Watson & J. Farrell-Jobst (eds.), *Gender and the Book Trades*, pp. 217–242, Leiden–Boston, Brill. https://doi.org/10.1163/9789004701656_014

Mapping Pliegos. Un catálogo colectivo y biblioteca digital de literatura de cordel

Juan Gomis Coloma
Universitat de València
Pilar Martínez Olmo
Biblioteca Tomás Navarro Tomás – CCHS-CSIC

Palabras clave: Lectura Activa; Prompting; Chain-of-Thought; Marcación de textos; Desambiguación
Keywords: Broadsheet literature; digital library; collective catalogue; loose leaflet; popular literature

El portal *Mapping Pliegos* (https://biblioteca.cchs.csic.es/MappingPliegos/) ofrece una base de datos y biblioteca digital de 18.817 pliegos sueltos publicados en España desde los inicios de la Edad Moderna hasta el siglo XX. El proyecto, iniciado en 2015, es fruto de la colaboración entre investigadores del campo de la literatura popular impresa, bibliotecarios y documentalistas, e instituciones que albergan colecciones de pliegos sueltos. Su objetivo principal es unificar en un solo catálogo los registros y reproducciones digitales de estos impresos, facilitando así el acceso de los investigadores y del público general a los mismos e impulsando la consulta y estudio de este rico y variopinto género editorial.

Esta comunicación tiene por objeto explicar las sucesivas fases de realización del proyecto, su estado actual y los retos de futuro.

La Cambridge University Library, la Biblioteca Tomás Navarro Tomás del CSIC y la British Library conformaron el núcleo inicial del proyecto. Las tres instituciones utilizan sistemas de gestión bibliotecaria basados en formato MARC con posibilidad de exportar los registros, lo que permitió contar con un corpus inicial de más de 5.000 registros.

La exportación puso de manifiesto que, aunque la base de descripción era semejante, cada institución había seleccionado ciertos campos MARC considerados más adecuados para su institución, pero no siempre coincidentes con los seleccionados por otra; esto hizo necesario realizar un mapeado para conseguir un registro único donde todos los campos de información relevante pudieran incorporarse de forma automática a la etiqueta seleccionada para el proyecto Mapping Pliegos.

Una vez creada la estructura básica de la base de datos y tras comprobar que las tres grandes colecciones podían ser incorporadas, se iniciaron dos líneas de trabajo paralelas: se establecieron nuevos contactos con instituciones y

colecciones privadas que pudieran guardar en sus fondos pliegos de cordel españoles para ofrecerles la posibilidad de incorporarse al proyecto y, al tiempo, se convocaron reuniones de trabajo con un equipo científico-técnico de expertos para enriquecer la descripción bibliográfica ofrecida por las bibliotecas con otros campos que permitieran ampliar las posibilidades de búsqueda y estudio de estas colecciones de pliegos.

Los primeros contactos se llevaron a cabo con dos grandes catálogos colectivos españoles: el Catálogo Colectivo del Patrimonio Bibliográfico Español (CCPB) y el catálogo de la Red de Bibliotecas Universitarias y Científicas (REBIUN). Gracias a la generosidad de sus responsables, se realizó una extracción directa de los registros que describían pliegos de sus catálogos colectivos. De forma paralela se contactó con otras instituciones como la Biblioteca de Cataluña, la Biblioteca Valenciana o la Biblioteca de Castilla y León, que exportaron en formato MARC las descripciones bibliográficas que describían los documentos considerados inicialmente pertinentes, y los miembros del equipo técnico de la BTNT realizaron el mapeado y la limpieza de los registros para incorporarlos a la base de datos creada como soporte del proyecto *Mapping Pliegos*.

El primer paso consistió en comprobar cuántos registros contaban con un campo MARC 856 (enlace externo) para el acceso a la copia digital de los ejemplares físicos que custodiaban las bibliotecas y si podía realizarse de forma gratuita, desde cualquier lugar del mundo y sin necesidad de identificación de usuario. Un número elevado de registros no contaban con este enlace, pero antes de validarlo o desestimarlo, se consultó con las bibliotecas para comprobar si existía una digitalización y faltaba incorporar el enlace.

Algunas bibliotecas se encontraban en una fase intermedia del proceso cuando sus registros fueron exportados o facilitados por un catálogo colectivo. Es decir, sí habían realizado la reproducción digital de sus ejemplares, pero no se había introducido su URL o enlace externo en el registro bibliográfico. En otros casos, la biblioteca no había realizado la digitalización o no estaba accesible de forma gratuita.

En el primer caso, los miembros del equipo técnico del proyecto incorporaron las URLs que faltaban para que los registros formaran parte de la base de datos. En el segundo caso, se marcaron los registros para incorporarlos en una segunda etapa, una vez la reproducción digital estuviera disponible de forma gratuita.

La revisión de la base de datos resultante de la extracción masiva de registros de los catálogos mencionados detectó descripciones de obras que no cumplían con los objetivos del proyecto, por no ser pliegos de cordel originales, sino estudios, reproducciones facsímiles o colecciones de textos que hubo que identificar y marcar como no pertinentes.

La segunda línea de trabajo se centró en el enriquecimiento de la base de datos con aquellos campos que los asesores científicos del proyecto consideraron imprescindibles para facilitar estudios posteriores. Estos campos

ya eran totalmente libres y no guardaban relación directa con los recogidos en la descripción bibliográfica seguida en las bibliotecas: materia, tema específico, formato del texto, género, descripción de las ilustraciones y precio en los casos en los que aparece impreso. Hubo que ampliar también la base de datos con los campos necesarios para enriquecer los registros de pliegos cantables del siglo XX, donde los textos de las canciones tenían referencias al cantante, a la grabación sonora, al autor de la letra y al compositor.

La lengua o lenguas en que está escrito el texto, el íncipit y una nota libre que permita la incorporación de cualquier otro dato, han ampliado la ficha bibliográfica hasta un total de 45 campos, mejorando la descripción de los registros y facilitando los estudios de futuros investigadores.

Un equipo de técnicos bibliotecarios y otro de jóvenes investigadores han colaborado en el enriquecimiento de la base de datos accediendo a la reproducción digital que la biblioteca depositaria ofrece a través de su propia biblioteca virtual. Este es uno de los valores más relevantes de *Mapping Pliegos*: la comprobación de la pertinencia de cada registro; la accesibilidad del enlace al documento y el estado de la digitalización en su totalidad. El portal facilita el acceso a la reproducción digital de cada biblioteca garantizando así el uso y el registro de datos estadísticos, así como el reconocimiento de la importancia del trabajo técnico realizado por cada institución.

Todo este proceso ha permitido validar un número muy elevado de registros, pero también identificar otras descripciones como no pertinentes, ejemplares que no coincidían con los objetivos de *Mapping Pliegos*. Algunas reproducciones accesibles al iniciar el proyecto han sido retiradas por sus instituciones o se les ha asignado un acceso restringido, y también documentos no digitalizados al completo. Todos estos registros han sido marcados para una segunda fase, cuando se ajusten a los objetivos del proyecto.

Como resultado de estas tareas técnicas, el proyecto *Mapping Pliegos* ofreció en el momento de su lanzamiento (diciembre de 2020) 6698 registros bibliográficos enriquecidos, y gracias al trabajo realizado desde 2021 y a la colaboración de nuevas instituciones la base de datos cuenta en 2025 con 18.817 registros que facilitan el acceso a otros tantos ejemplares de pliegos de cordel españoles impresos entre los siglos XVI y XX, en cuya descripción de ejemplares se facilitan datos sobre diferentes tiradas, imprentas, temas, etc.

Mapping Pliegos es un proyecto colaborativo que está abierto a la incorporación de nuevas colecciones en las condiciones ya especificadas. Así, una de sus principales ventajas radica en que los nuevos registros están disponibles desde el momento de su incorporación y es posible acceder a ellos por los campos de descripción bibliográfica básica: título, imprenta, palabras clave, etc.

El portal *Mapping Pliegos* se propone incrementar el número de bibliotecas participantes y ampliar su marco geográfico (con impresos publicados en Latinoamerica), actualizar de forma continua su información, incorporando estudios, bibliografía, enlaces de interés y nuevas propuestas con la vocación de

convertirse en una fuente de información y referencia para los investigadores e interesados en este vestigio de nuestra historia cultural.

Bibliografía

BOTREL, J.-F.; & GOMIS, J. (2019), «“Literatura de cordel” from a Transnational Perspective. New Horizons for an Old Field of Study», en M. Rospocher, J. Salman & H. Salmi (eds.), *Crossing Borders, Crossing Cultures. Popular Print in Europe (1450–1900)*, Berlin, De Gruyter.

FUMERTON, P. (2015), *Broadside Ballads in Early Modern England: Moving Media, Tactical Publics*, Philadelphia, University of Pennsylvania Press.

GOMIS, J.; & MARTÍNEZ OLMO, P. (2022), «Mapping Pliegos: un proyecto de humanidades digitales para el estudio de la literatura de cordel», en J. Gomis, E. Lara & C. Bonet (eds.), *Quimeras de lo insólito. Fascinación y horror en el mundo hispánico*, Valencia, Tirant Humanidades, pp. 275–297.

GOMIS, J. (s. f.), *Menudencias de imprenta. Producción y circulación de la literatura popular (Valencia, siglo XVIII)*, Valencia, Alfons el Magnànim.

PENA SUEIRO, N. (s. f.), «Estado de la cuestión sobre el estudio de las Relaciones de sucesos», *Pliegos de Bibliofilia*, n.º 1, pp. 43–77.

ECHAVARRÍA, A. (2025), «Schéma d’annotation TEI pour les procès pour cause de foi de l’Inquisition espagnole», XML Standard, TEI Schema, NAKALA (Huma-Num–CNRS). https://doi.org/10.34847/nkl.ab374s00

——— (2025, marzo), «Scripts de conversion de thésaurus RDF vers TEI ODD et TEI Taxonomy», Zenodo. https://doi.org/10.5281/zenodo.15044448

——— (2025), «Thésaurus de typologies documentaires des procès pour cause de foi de l’Inquisition espagnole», SKOS thesauri, NAKALA (Huma-Num–CNRS). https://doi.org/10.34847/nkl.2a4e2h52

——— (2025, marzo), «Thèse de doctorat – Scripts extraction d’information XML-TEI», Zenodo. https://doi.org/10.5281/zenodo.15035164

ISAAC, A.; & SUMMERS, E. (2009), «Manual de SKOS (Simple Knowledge Organization System, Sistema Simple de Organización del Conocimiento)», W3C. http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/

TEI-C (2025), *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, XML Standard, Technical Council of the TEI Consortium, 24 de enero de 2025. http://www.tei-c.org/Guidelines/P5/

VITALI-ROSATI, M. (2018), *On Editorialization: Structuring Space and Authority in the Digital Age* (M. Rasch, trad.), Theory on Demand 26, Amsterdam, Institute of Network Cultures.

VOGELER, G. (2023), «Charters Encoding Initiative project – Vocabulaire international de la diplomatique», SKOS terminology, Munich, 5 de enero de 2023. https://www.cei.lmu.de/VID/

Uso de técnicas avanzadas de imagen en la documentación de obras afectadas por la DANA: un caso de estudio en el proyecto ChemiNova

Carmen Cano Sola
Álvaro Solbes García
Ester Alba Pagán
Arabella León Muñoz
Esperanza Villuendas Ferri
Universitat de València

Palabras clave: Conservación, Cambio climático, Valencia, Técnicas analíticas de imagen
Keywords: Conservation, Climate Change, Valencia, Image Analytical Techniques

Introducción

El cambio climático es uno de los desafíos más importantes que concierne a la población del siglo XXI (Pasikowska-Schnass, 2024). Sus efectos se están viendo manifestados en fenómenos meteorológicos extremos como lo ocurrido en Valencia el 29 de octubre del 2024, con la DANA (Depresión Aislada de Niveles Altos). Esta catástrofe devastadora dejó víctimas mortales y causó daños materiales incalculables. Asimismo, su impacto en el patrimonio cultural y la gestión de los bienes tras el suceso también ha sido significante, y ha servido para afirmar la urgencia en la adaptación de la gestión y la conservación del patrimonio cultual. Tras esta realidad, la Universitat de València y el Vicerrectorado de Cultura ha tomado un papel importante en la gestión y la conservación de algunas de las colecciones que se han visto afectadas. Las acciones llevado a cabo por el equipo interdisciplinar son tanto de documentación, conservación y salvaguarda de las obras, como una oportunidad para desarrollar estrategias de conservación con colaboraciones de investigación tecnológica como lo es el proyecto Horizonte Europa ChemiNova. Este proyecto desarrolla y promueve el uso de herramientas no invasivas para la conservación del patrimonio. En el marco del proyecto se ha continuado trabajando con las obras rescatadas, de forma que algunas de ellas han pasado a formar parte del propio proyecto y de sus procesos. Mediante la adquisición de diversos tipos de datos tanto analíticos, espectrales y espaciales se pretende mejorar y facilitar los procesos de conservación, análisis y monitoreo de las obras

con la utilización de las novedosas herramientas tecnológicas que el proyecto está implementando.

Metodología

La investigación se centra en el análisis y el estudio de las obras afectadas por las inundaciones siguiendo la línea del proyecto ChemiNova, utilizando como caso de estudio una obra de la colección del ayuntamiento de Algemesí del artista Francesc Sempere Fernández de Messa.

Algunas de las colecciones intervenidas son: la colección del artista Antonio Barroso, que se encontraba almacenada en el Museu de la Festa d'Algemesí o la colección completa del Antonia Mir, en su propio museo de Catarroja. Así como colecciones con obras de artistas más reconocidos nacionalmente, como lo es la colección de Martínez Guerricabeitia, perteneciente a la colección de obras contemporáneas de la Universitat de València.

Como ejemplo de este proceso y gestión se va a exponer la obra pictórica, *Tendur, Maria esposa* de Fransec Sempere Fernández de Messa. Esta pieza ha sido seleccionada también para ser utilizada como caso de estudio en el proyecto ChemiNova.

FICHA TÉCNICA	
OBRA	*Tendur, Maria, esposa*
AUTOR	*Francesc Sempere Fernández de Messa*
CRONOLOGÍA	*1957*
PROVINCIA	*Valencia*
LOCALIDAD	*Algemesí*
DIMENSIONES	*81 x 60cm*
MATERIALES Y TÉCNICAS	*Óleo sobre lienzo, marco de madera policromada*

Tabla 1. Ficha técnica de la obra.

ChemiNova es un proyecto interdisciplinar formado por un consorcio de 12 instituciones de 8 países distintos que apuesta por el desarrollo de nuevas herramientas tecnológicas que ayuden mejorar la conservación, el análisis y el monitoreo de bienes culturales, más concretamente de aquellos que se ven amenazados por el cambio climático y los conflictos armados. La aplicación, constituida por diversas herramientas específicas incluirán tecnologías como la Realidad Aumentada, la Realidad Mixta y la Inteligencia Artificial. El proyecto gira en torno al modelo 3D enriquecido (e3D), un gemelo digital con distintas dimensiones de información ¾espacial, espectral, fotométrica, temporal, etc.¾ tomada con técnicas analíticas no invasivas. La información se añade al modelo en forma de capas o de texturas de modo que se podrá analizar visualmente y de forma muy fácil con todos estos datos.

El proceso de trabajo comienza con el rescate de las obras, su traslado a un lugar seguro y continúa con el estudio y registro de las pinturas con las técnicas no invasivas incluidas en el proyecto.

Uno de los primeros pasos fue la incursión a los lugares afectados para identificar las obras, rescatarlas y reubicarlas en espacios más seguros. En el caso que nos ocupa la obra fue trasladada, junto con las demás de la colección, al Laboratorio de Análisis y Diagnóstico de obra de arte de la Facultad de Geografía e Historia de la Universitat de València, donde se almacenaron en un ambiente controlado de humedad y temperatura.

Se realizó un primer registro en el momento del rescate para poder tener un control de las obras afectadas y del traslado de estas. El siguiente paso en el rescate y salvaguarda de las obras es la realización de un análisis y diagnóstico de cada una de las piezas para poder estimar los daños y valorar las consecuencias de la catástrofe. Para ello se utilizó el informe de conservación que se había creado específicamente para el proyecto ChemiNova. Los informes de conservación fueron creados por expertos de ChemiNova utilizando como base la metodología de trabajo y la documentación empleada en otros contextos por los autores y definiendo el léxico de las alteraciones según los Estándares Técnicos Europeos CEN/TC 346 “Conservation of Cultural Heritage”.

Para la creación del modelo y para poder realizar un análisis extenso de las obras se han ido realizando tomas de datos con distintas técnicas de análisis, todas ellas no invasivas. Cada una de las técnicas se realizó tanto en el anverso como en el reverso, para poder recopilar la mayor cantidad de datos posible y hacer un registro completo de los daños y la obra. A continuación, se describen las técnicas que se han realizado:

- Fotografía Multibanda (MBI): La fotografía multibanda nos permite observar la obra con longitudes de onda más allá del espectro visible iluminando con distintos tipos de luces y capturando longitudes de onda específicas. Las imágenes son capturadas con la cámara Nikon D850 modificada para ser sensible en los rangos 360 a 1000nm. Se han realizado la fluorescencia de Ultravioleta, la Ultravioleta reflejada, la reflectografía infrarroja, pero también fotografías RGB cambiando la dirección de la iluminación como la fotografía rasante y la transmitida (Dyer et al., 2013).
- Imagen Hiperespectral (HSI): La toma de datos hiperespectrales se ha realizado con una cámara portátil SPECIM IQ con la que obtenemos información tanto del espectro visible como del no visible (infrarrojo cercano), de los 400 a los 1000nm. Esta técnica nos permite capturar tanto imágenes como un espectro formado por las 204 reflectancias de distintas bandas de cada uno de los píxeles que forman la imagen (515x515) (Picollo et al., 2020).
- Microscopía óptica (MO): Para la obtención de detalles se realizaron fotografías con el microscopio portátil Dino-Lite (10x-70x-200x, 8 LED blancos polarizados).

- Reflectografía infrarroja (IRR). En la imagen multibanda se incluye una IRR, pero en este caso se realizó una reflectografía con un equipo específico que nos permite ver la obra en longitudes de onda más altas y que penetran más la superficie. La fotografía IRR se hizo con la cámara Apollo de Opus Instruments, la cual opera en rangos del espectro no visible, de 900 a 1700nm, con filtros de 1250, 1250-1510 y 1510 nm.
- Fotogrametría y modelos 3D: El equipo utilizado para la fotogrametría fue la cámara Sony a 7R IV con la lente FE 24-105mm F4 G OSS. También se realizaron modelos 3D con software gratis para dispositivos móviles como el RealityScan y la aplicación Scaniverse. Y por último se utilizó un escáner láser de mano EinScan Pro-HD, de modo que podemos comparar los modelos con diferentes técnicas y diferentes calidades.

Resultados

Tras la obtención de las técnicas descritas tenemos una recopilación extensa con información muy diversa de las obras. Todas las técnicas que se presentan a continuación se realizaron tanto por el anverso como por el reverso.

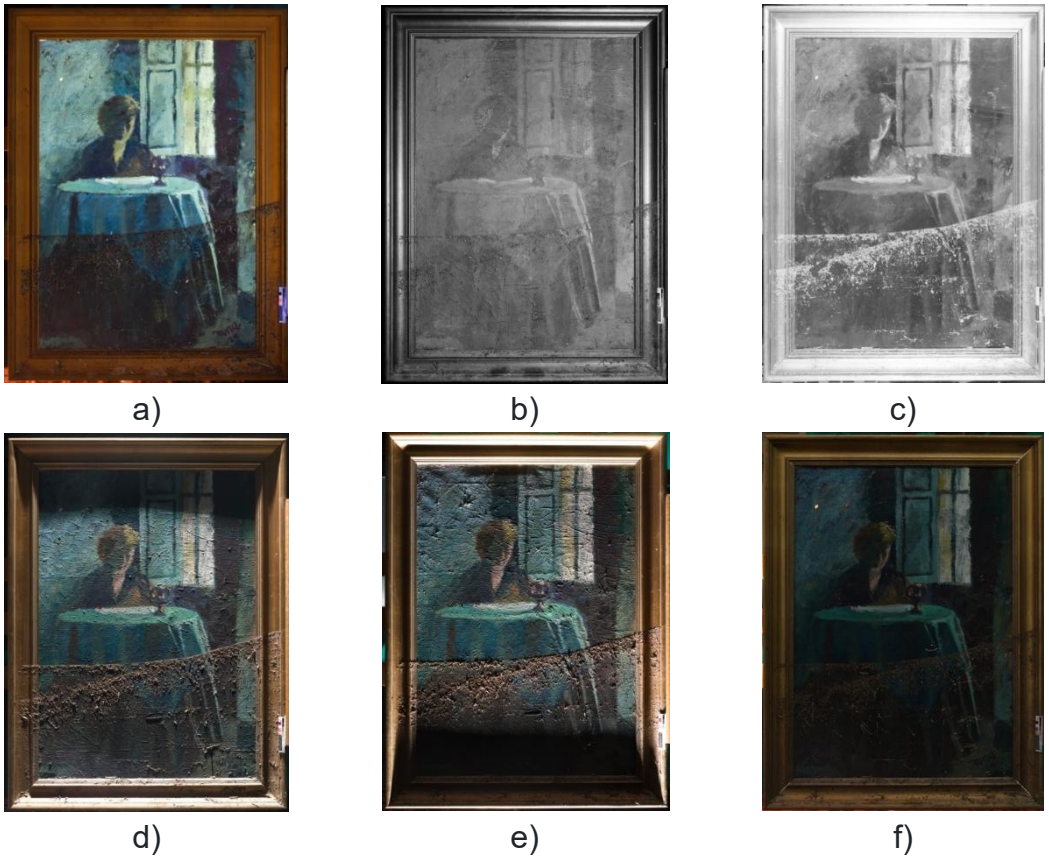
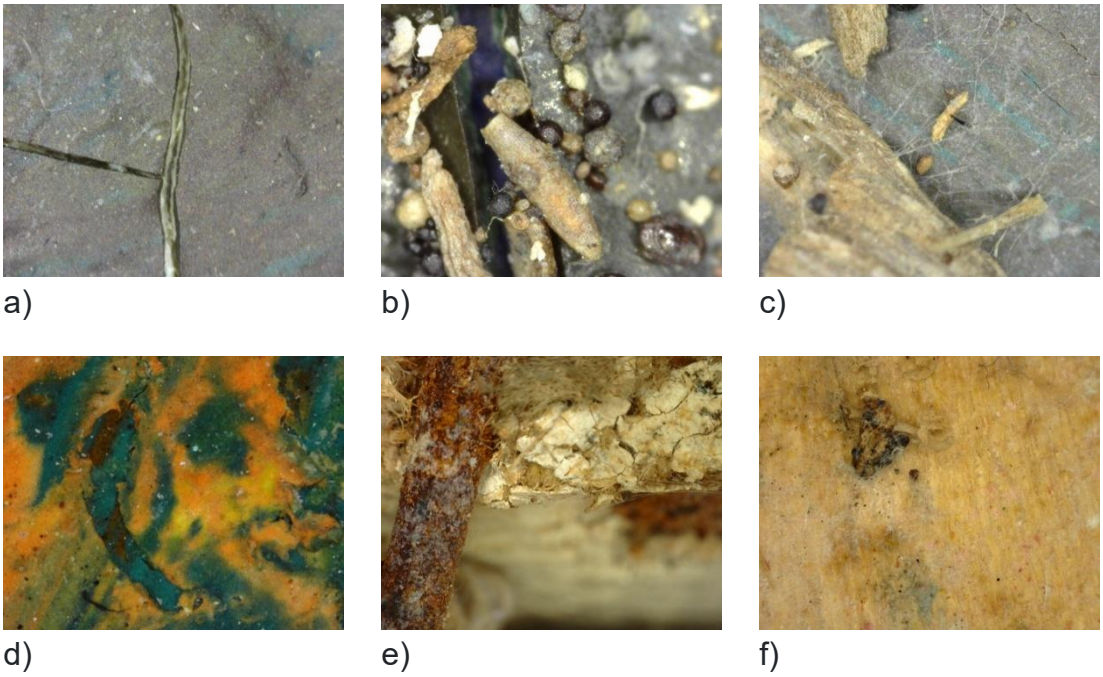


Figura 2. Fotografías multibanda. a) UVF; b) UVR; c) IRR; d) Rasante Inferior; e) Rasante Superior y f) Transmitida.

Figura 3. Fotografía de reflectancia infrarroja, cámara Apollo. Volteada 90º a la derecha.



a) b) c)

d) e) f)

Figura 4. Microscopías ópticas de detalle.

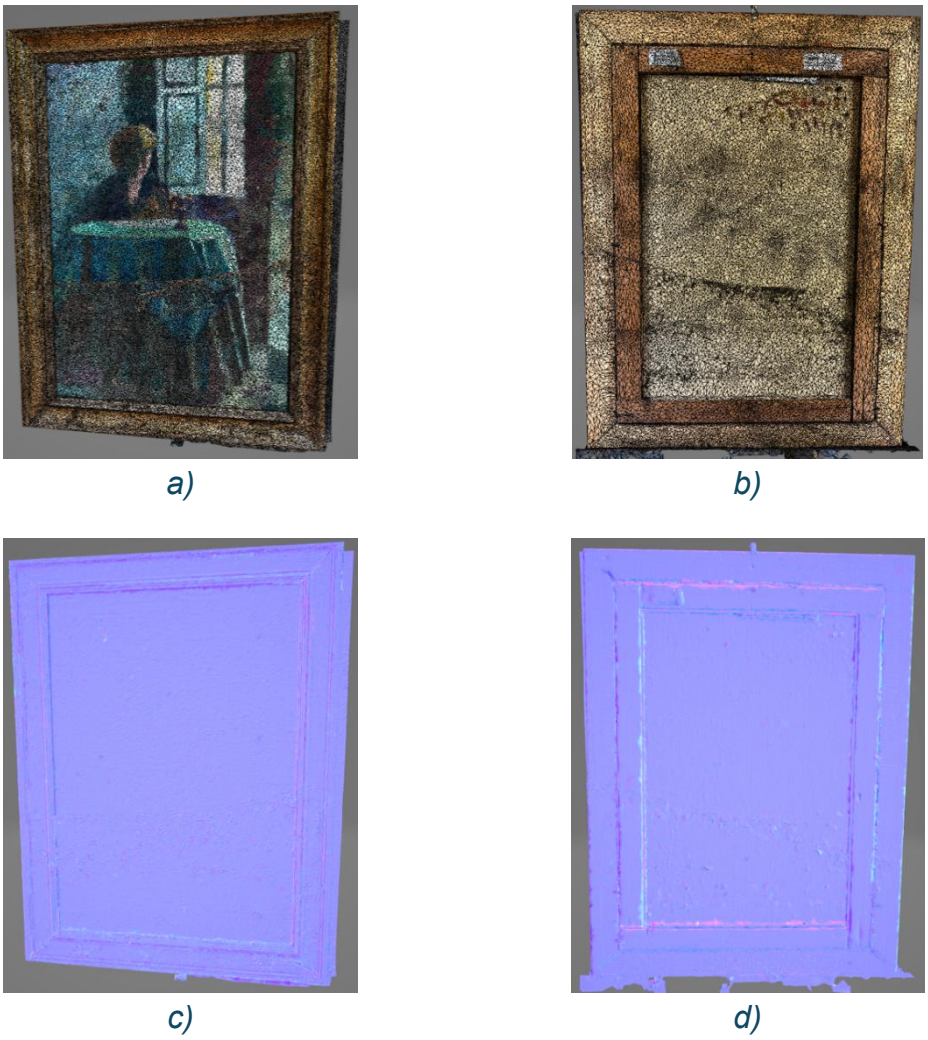


Figura 5. Modelos 3D realizados con software gratuito para smartphone. Malla a) anverso; b) reverso; Mapa de normales c) anverso; d) reverso.

Las imágenes multibanda han permitido caracterizar los daños con una mayor precisión, mientras que con la reflectografía infrarroja se ha determinado que la pintura, está realizada sobre otra obra pictórica anterior como se puede observar en la Figura 3. Los detalles microscópicos de la Figura 4 nos ofrecen información concreta sobre los restos terrosos (b) y vegetales (c) que cubren la obra y sobre la técnica (d) y el tipo de tejido utilizado por el autor, además de detalles sobre daños como micro-craqueladuras (a), oxidación (e) y microorganismos (f).

Los modelos 3D que se observan en la Figura 5 están en un proceso de mejora de la calidad para poder continuar con el solape de las demás técnicas en el modelo.

Conclusiones

Este artículo presenta parte del proceso llevado en el proyecto ChemiNova, parte del Programa Marco Horizonte Europa, en el rescate, análisis y conservación de obras afectadas por las inundaciones vividas en la zona sur de

Valencia el 29 de octubre del 2024. El trabajo presentado es el realizado hasta el mes de mayo del 2025, quedando aún la creación del modelo e3D y la implementación de los datos en las herramientas de ChemiNova que están actualmente en desarrollo.

ChemiNova ha recibido financiación del Programa Marco Horizonte Europa de la Unión Europea en virtud del acuerdo de subvención 101132442.

Bibliografía

DYER, J., VERRI, G., & CUPITT, J. (2013), *Multispectral Imaging in Reflectance and Photo-induced Luminescence Modes: A User Manual*, recurso en línea. http://www.charismaproject.eu/

MERCURI, F., CICERO, C., ORAZI, N., PAOLONI, S., MARINELLI, M., & ZAMMIT, U. (2015), «Infrared thermography applied to the study of cultural heritage», *International Journal of Thermophysics*, 36(5–6), pp. 1189–1194. https://doi.org/10.1007/s10765-014-1645-x

PASIKOWSKA-SCHNASS, M. (2024), *The Impact of Climate Change on Cultural Heritage*, recurso en línea.

PICOLLO, M., CUCCI, C., CASINI, A., & STEFANI, L. (2020), «Hyper-spectral imaging technique in the cultural heritage field: New possible scenarios», *Sensors*, 20(10), art. 2843. https://doi.org/10.3390/s20102843

Transformación semántica de museos virtuales locales

María Luisa Díez Platas
Susana Quirós Alpera
Luis Pedraza Gómara
Universidad Internacional de La Rioja
Paloma Centenera
Universidad CEU San Pablo

Palabras clave: Museo virtual, patrimonio digital, ontología, visualización semántica, buscador facetado
Keywords: Virtual Museum, digital heritage, ontology, semantic visualization, faceted search engine

Problema de investigación

En las últimas décadas, la digitalización del patrimonio ha promovido nuevas formas de representación cultural. Los museos virtuales, particularmente aquellos nativos digitales sin vinculación física emergen como plataformas accesibles, inclusivas y participativas para preservar la memoria colectiva. Su capacidad para superar barreras geográficas, temporales y presupuestarias los convierte en herramientas potentes para comunidades locales que desean mantener vivas sus tradiciones, historia y patrimonio natural y artístico. Sin embargo, esta proliferación digital ha estado acompañada de importantes carencias como la ausencia de organización semántica de los contenidos, plataformas rígidas, escasa interoperabilidad, interfaces poco intuitivas y baja adaptabilidad a contextos educativos o de investigación. El caso del Museo Virtual de Getafe, impulsado por la Asociación de Amigos del Museo y la ciudadanía local, representa una de estas iniciativas.

Esta propuesta está centrada en este tipo de museos virtuales, generalistas o específicos y locales, que están relacionados con poblaciones, comunidades o regiones, que buscan recuperar y difundir costumbres, tradiciones y la memoria cultural colectiva, así como el patrimonio cultural y natural. Estos museos suelen contar, además, con archivos documentales, hemerotecas y generan exposiciones para guiar al visitante.

La creación de un museo físico de estas características es costosa, debido a las infraestructuras necesarias, y sus impulsores enfrentan obstáculos administrativos y de presupuesto. Además, en algunos de estos museos, la participación ciudadana representa un papel clave.

El objetivo general persigue enriquecer la semántica de los museos virtuales nativos digitales y las formas de visualización y recuperación mediante el desarrollo de un modelo ontológico y una plataforma avanzada de visualización y recuperación de información.

La forma de recuperación de información es imprescindible, Bast et al. (2012) plantea un buscador semántico combinando texto completo y ontologías, consistente en un motor de búsqueda rápido y fácil de usar para la búsqueda semántica de texto completo. La búsqueda semántica de texto completo combina las capacidades de la búsqueda estándar de texto completo y la búsqueda ontológica.

Los objetivos específicos están orientados tanto a la parte semántica del modelado como a la parte técnica, permitiendo modos avanzados de visualización y recuperación de información.

Para ello, se considera necesaria una metodología que combina la conceptualización de dominio, la ingeniería ontológica, el diseño de experiencia de usuario (UX) y la evaluación empírica con usuarios reales y se está diseñando una plataforma de gestión y mantenimiento para este tipo de museos virtuales, que no sea meramente un espacio de exhibición, sino una herramienta semánticamente enriquecida tanto para usuarios interesados además de para la educación y la investigación.

Maestre, en su trabajo de 2023, expone el uso de una plataforma como Omeka para catalogar patrimonio mural y presentar visitas virtuales enriquecidas semánticamente. Por otra parte, Pérez (2011) analiza metadatos, empleando OAI-PMH y Omeka para exposiciones digitales. Así mismo, el trabajo de Gayoso-Cabada et al. (2016) propone una navegación facetada basada en folksonomías, que resulta de enorme utilidad en buscadores complejos. Adicionalmente, los autores Garijo y Poveda-Villalón (2020) establecen buenas prácticas para ontologías semánticas accesibles e interoperables en su trabajo.

El anteproyecto “UCO Museo Virtual”, auspiciado por la Universidad de Córdoba, sienta las bases, en 2012, para combinar patrimonio local, participación ciudadana y difusión científica de manera virtual. Una década después, en el mes de febrero de 2022, Díez Platas, en su presentación “Nueva Web del Museo Virtual de Getafe”, figura 1, manifiesta la voluntad de que la identidad histórica y cultural de Getafe se vea plasmada en “La Aldea Global” de Internet, trabajando con seriedad y rigor, en pos de un objetivo final, un museo físico para la ciudad de Getafe, basado en la colección de objetos digitales del museo virtual.

En el marco del proyecto de investigación “Ontología y sistema de visualización y recuperación de información avanzada para museos virtuales nativos digitales de ámbito local” (P25-002-I), financiado por la Universidad Internacional de La Rioja (UNIR), que se encuentra en sus primeras fases, se persigue el desarrollo de una plataforma modular destinada a museos virtuales de pequeño o mediano tamaño. Esta iniciativa tiene como objetivo superar las principales limitaciones observadas en este tipo de entornos digitales —

particularmente en los de ámbito local— mediante herramientas accesibles, de fácil uso y alto potencial semántico. La propuesta integra tecnologías de la web semántica, visualización contextual y sistemas de recuperación avanzada de información para mejorar tanto la experiencia del usuario como la utilidad educativa, patrimonial e investigadora de estas plataformas.

Metodología

Por tanto, en primer lugar, la primera parte de esta propuesta se ha desarrollado en la primera fase del proyecto mencionado y consiste en el desarrollo del modelo conceptual del dominio de estudio, que se plasmará posteriormente en una ontología. Esta ontología integrará información sobre obras, documentos, costumbres y exposiciones, y se organizará en un grafo de conocimiento. Todos los elementos serán presentados semánticamente relacionados y estarán disponibles en abierto.

La implementación técnica se llevará a cabo sobre Omeka S, una plataforma de gestión de colecciones digitales de código abierto que admite la integración de vocabularios externos y módulos personalizados

Para Rubén Alcaraz (Martínez, 2022), el gestor de contenidos Omeka S supone una herramienta potente para el desarrollo de museos virtuales y en éste se han superado algunas de las limitaciones de los gestores tradicionales. Este gestor tiene capacidad para gestionar múltiples exposiciones, integrar vocabularios de la web semántica y personalizar tanto la descripción de objetos como el diseño de la interfaz pública, lo que permite a los museos crear entornos virtuales interactivos y adaptados a sus necesidades específicas.

No obstante, y para salvar las limitaciones de esta plataforma en estructuración de la información y en visualización avanzada, se realizarán desarrollos que permitan para conectar la ontología con el sistema de contenidos, permitiendo la navegación basada en relaciones semánticas y la exploración mediante facetas (autor, época, tipo de documento, lugar de procedencia, etc.) algo que no es posible con Omeka.

Por tanto, se presentará un modelo de visualización atractivo y funcional, con un buscador facetado para facilitar búsquedas complejas, adaptadas a usuarios, educadores e investigadores.

Estado del proyecto

En el momento actual, se está elaborando el modelo conceptual como base para la implementación de la ontología de dominio que se pretende sea aplicable a distintos museos virtuales de distinta naturaleza. En la figura 1 se muestra la primera versión del modelo conceptual base para la ontología

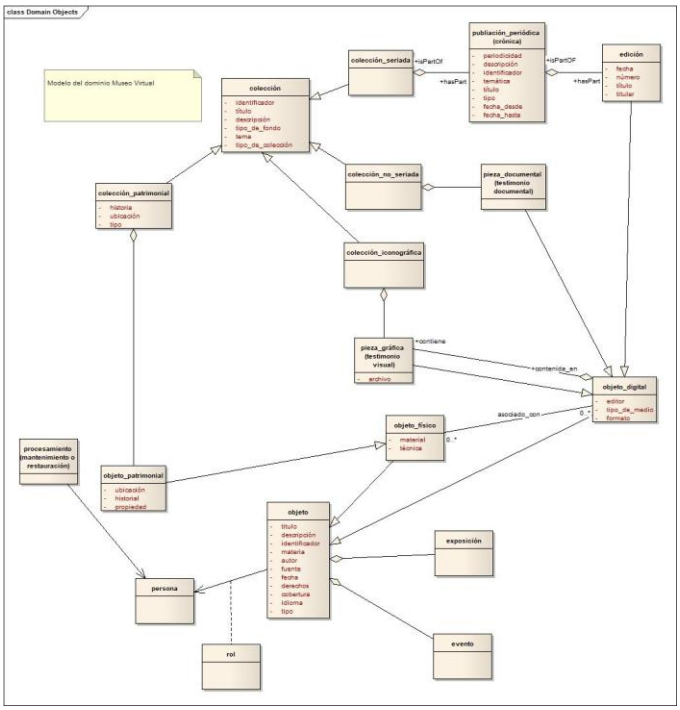


Figura 1. Modelo conceptual del dominio del Museo Virtual de Getafe.

Simultáneamente, se está trabajando en el Sistema de migración de los datos presente en el Museo Virtual de Getafe, sobre el que se implementará una prueba de concepto, y se generará el grafo de conocimiento basado en los conceptos de la ontología para su validación.
Se espera que en el momento de celebración de HDH2025 se hayan terminado estas fases y se pueda presentar un prototipo funcional.

Discusión e impacto

La propuesta y, en particular, el proyecto en su conjunto aborda no solo un problema tecnológico, sino que presentan un nuevo modelo para manejar el patrimonio digital local desde distintos puntos de vista.
Desde un punto de vista de investigación, el proyecto ofrece una metodología similar que puede ser utilizada para otros museos digitales o híbridos, considerando las situaciones locales particulares con recursos escasos.
Desde un aspecto técnico, también mostrará cómo combinar la web semántica y la ingeniería del conocimiento para desarrollar sistemas de gestión del patrimonio cultural que sean accesibles.
Desde un punto de vista educativo, el museo se convierte en un recurso educativo vivo capaz de adaptarse a diferentes niveles educativos y de promover el aprendizaje activo a través de rutinas de pensamiento, mapas mentales y trabajo por proyectos.

Bibliografía

ALCARAZ MARTÍNEZ, R. (2022), «Omeka S como alternativa para el desarrollo de colecciones digitales y proyectos de humanidades digitales», *BiD: textos universitarios de biblioteconomía y documentación*, n.º 48. https://doi.org/10.1344/BiD2022.48.06

BAST, H., BÄURLE, F., BUCHHOLD, B., & HAUSSMANN, E. (2012), «Broccoli: Semantic full-text search at your fingertips», *arXiv* (arXiv:1207.2615). https://arxiv.org/abs/1207.2615

DÍEZ PLATAS, M. L. (2022), «Nueva Web del Museo Virtual de Getafe», recurso en línea. https://museo.getafe.es/omeka/items/show/6429

GARIJO, D., & POVEDA VILLALÓN, M. (2020), «Best practices for implementing FAIR vocabularies and ontologies on the web», en G. Cota, M. D., & G. L. Pozzato (eds.), *Applications and Practices in Ontology Design, Extraction, and Reasoning*, pp. 39–54, IOS Press.

GAYOSO CABADA, J., RODRÍGUEZ CEREZO, D., & SIERRA, J. L. (2016), «Multilevel browsing of folksonomy-based digital collections», en W. Cellary, M. Mokbel, J. Wang, H. Wang, R. Zhou, & Y. Zhang (eds.), *Web Information Systems Engineering – WISE 2016, Lecture Notes in Computer Science*, vol. 10042, pp. 43–51, Springer.

RODRÍGUEZ ORTEGA, N., ROLDÁN GARCÍA, M. del M., DÍEZ PLATAS, M. L., & SALVACHÚA, M. (2018), *OntoExhibit: una ontología para modelar exposiciones como sistemas culturales complejos*, Universidad de Málaga.

MAESTRE, A. M. C. (2023), «Digital Humanities applied to the cataloguing and conservation of mural heritage», *Compendium: Journal of Comparative Studies*.

PÉREZ, T. S. (2011), «Exposiciones digitales y reutilización», *Métodos de Información*, 2(2), pp. 29–46.

Sistematización metodológica para el análisis multidimensional del subsistema lingüístico-textual del dominio de las exposiciones artísticas

Nuria Rodríguez-Ortega
María Ortiz Tello
Ángel Lumbreras Fernández
iArtHis_Lab –
Universidad de Málaga
María Luisa Díez-Platas
UNIR

Palabras clave: Ontologías culturales, exposiciones artísticas, análisis discursivo multiescala, grafos de conocimiento, sistemas culturales complejos
Keywords: Cultural ontologies, art exhibitions, multi-scale discursive analysis, knowledge graphs, complex cultural systems

1. Objetivos

Este trabajo tiene como objetivo demostrar el potencial de la ontología OntoExhibit[1] para generar conocimiento multidimensional sobre el ámbito expositivo a partir del análisis del subsistema lingüístico-textual que se configura en torno a las exposiciones artísticas. Este subsistema, entendido como una de las vías fundamentales de materialización de la dimensión semántico- discursiva del ecosistema expositivo, requiere una metodología específica que complemente los procesos de extracción y modelado ya descritos en Rodríguez Ortega et al. (2025). En este sentido, se propone una sistematización metodológica replicable y cuantificable para el análisis del discurso textual expositivo, centrada en su integración en un grafo de conocimiento.

Para ello, se aplica esta metodología sobre un corpus compuesto por 20 exposiciones celebradas en España entre 2010 y 2024, centradas en la figura de Picasso, y se presenta un flujo de análisis estructurado en tres niveles: arquitectura socio-textual, análisis semántico comparado y construcción narrativa del artista. De este modo, el estudio no solo avanza en la conceptualización del subsistema lingüístico-textual expositivo, sino que contribuye a consolidar un enfoque metodológico reproducible dentro del campo de las Humanidades Digitales.

[1] https://complexhibit-project.github.io/OntoExhibit/index-en.html

2. Justificación y antecedentes

Durante la última década, la consideración del dominio expositivo como red, ecología o ecosistema ha cobrado creciente relevancia tanto en el ámbito tradicional de los *exhibition studies* como en las investigaciones de Humanidades Digitales (Farquharson, 2006; Quiles, 2014; Proctor, 2017; Braden & Teekens, 2020, entre otros). En este contexto, el proyecto Complexhibit[2] [2] se articula sobre la premisa de que para comprender cabalmente estas estructuras socioculturales no basta con modelar los actores (artistas, comisarios, instituciones) y los objetos exhibidos; es imprescindible integrar también los discursos —materializados en catálogos, textos curatoriales, reseñas periodísticas, interacción en redes sociales o imágenes, entre otros— que se generan en torno a las exposiciones y circulan entre distintos agentes y comunidades.

El modelo ontológico OntoExhibit responde a esta concepción del ámbito expositivo al proponer una ontología en OWL2 que modela las exposiciones como sistemas socio-semánticos. Su principal innovación respecto de otras propuestas (como Linked Art[3], BasArt[4], MSDMP[5], etc.) reside en la integración explícita de los elementos semántico-discursivos en un grafo de conocimiento. De este modo, las entidades tradicionales (artistas, instituciones, objetos, eventos, etc.) se complementan con perspectivas externas y narrativas contextuales que posibilitan análisis culturales más complejos y enriquecidos (Rodríguez-Ortega et al., 2024; Rodríguez-Ortega, 2024; Rodríguez-Ortega et al., 2025).

El grafo OntoExhibit, consultable mediante SPARQL[6], combina información socioinstitucional — derivada de la base de datos Expofinder[7]— con datos lingüístico-textuales procesados mediante una metodología específica ya descrita en Rodríguez-Ortega et al. (2025). Esta metodología, basada en técnicas de PLN y mapeo semántico, permite extraer, normalizar e integrar información textual proveniente de distintos géneros (catálogos, textos curatoriales, reseñas de prensa) en forma de tripletas RDF validadas y alineadas con el modelo ontológico.

A modo de ejemplo, la figura 1 muestra cómo se modela una reseña crítica (clase exo:ArtReview), que se vincula con una exposición concreta y permite recuperar su contenido, autoría, fecha y valoración evaluativa dentro del grafo.

---

[2] Proyecto Complexhibit. Estudio e interpretación del dominio de las exposiciones artísticas como sistema cultural complejo mediante analítica de datos y procesamiento del lenguaje natural. PID2021-125037NB-I00 (Ministerio de Ciencia e Innovación). https://github.com/orgs/Complexhibit-Project/repositories

[3] https://linked.art/

[4] https://artlas.huma-num.fr/map/#/

[5] Modelo Semántico Digital del Museo del Prado: https://www.museodelprado.es/grafo-de-conocimiento/modelo-ontologico

[6] https://github.com/Complexhibit-Project/sparql?tab=readme-ov-file

[7] https://expofinder.iarthislab.eu/web/

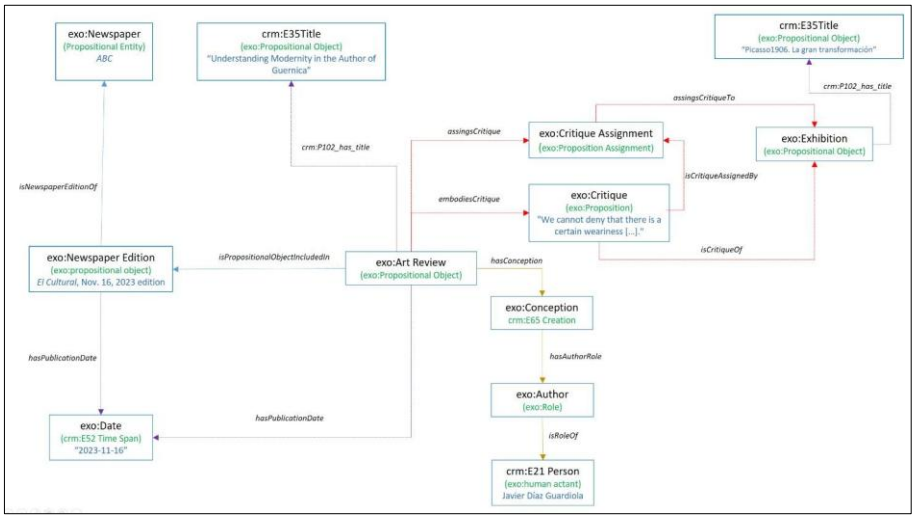


Figura 1. Modelado de la clase Art Review en la ontología OntoExhibit. Las clases-padre se muestran en verde, mientras que las instancias se representan en azul

Este proceso ha permitido generar un grafo cuya validación (SHACL) arroja una tasa de error inferior al 1 %. A partir de esta infraestructura, es posible realizar consultas complejas que interrelacionan actores, prácticas discursivas e imaginarios culturales articulados en torno al hecho expositivo.

3. Sistematización metodológica

Ahora bien, durante la fase subsiguiente de análisis —orientada al diseño e implementación de consultas SPARQL para extraer conocimiento multidimensional y significativo sobre el subsistema lingüístico-textual del ámbito expositivo—, se hizo evidente la necesidad de desarrollar una sistematización metodológica robusta. Esta necesidad dio lugar al diseño analítico que aquí se presenta.

Para extraer y analizar métricas clave a partir de un grafo RDF que incorpora textos vinculados al ámbito expositivo, proponemos un sistema metodológico que integra una variedad de métricas y estrategias analíticas complementarias entre sí. Este sistema se conforma a partir de una taxonomía de familias de consultas SPARQL: SYS (estructura del subsistema lingüístico-textual), SEM (análisis lingüístico-semántico comparado), ART (construcción discursivo-narrativa sobre artistas). Estas familias se integran en un flujo analítico único, que permite abordar de forma articulada las dimensiones estructural, léxico-semántica y narrativa del ecosistema discursivo. Cada familia incluye plantillas de consulta parametrizables; criterios de segmentación y agregación; métricas asociadas y procedimientos de análisis; formatos de salida orientados a la visualización y comparación.

Este apartado describe las preguntas de investigación expresadas como consultas SPARQL. Los métodos y procesos de análisis incluidos en el modelo se exponen en el Anexo I. El objetivo es explicitar la arquitectura analítica para

garantizar su transparencia, su reproducibilidad y evaluación por parte de la comunidad investigadora.

La sistematización metodológica parte de una pregunta general que ha guiado el diseño analítico:

¿Cómo se construye discursivamente Picasso en el entramado expositivo, y qué conocimiento emergente puede destilarse mediante el grafo Complexhibit que no sería accesible a través de raproximaciones tradicionales centradas en exposiciones o textos individuales?

1. Familia SYS. Arquitectura de subsistema lingüístico-textual [Cfr. Tabla 1. Anexo I]

La familia SYS tiene como objetivo caracterizar la infraestructura del subsistema lingüístico-textual generado en torno a las exposiciones. Para ello, se aplican consultas SPARQL orientadas al análisis estructural del corpus textual y de sus relaciones con actores, instituciones, géneros y cronologías. Las métricas generadas permiten medir la densidad textual por exposición (SYS-DENS), la distribución por géneros discursivos (SYS-TIPO), la cobertura mediática (SYS-MEDIA), la red de autorías (SYS-AUTHOR) y sus vínculos con instituciones (SYS-ACTOR+), así como la evolución temporal de la producción discursiva (SYS-TEMP).

Esta caracterización permite visualizar la circulación del discurso expositivo en términos cuantitativos y relacionales, identificar sesgos de cobertura, establecer un mapa jerárquico de las voces que sustentan el discurso, evaluar la centralidad o versatilidad de los autores, su vinculación o independencia respecto del aparato museístico, etc. En conjunto, la familia SYS proporciona el esqueleto socio-textual del sistema discursivo, sobre el cual se aplican posteriormente análisis semánticos (familia SEM) y narrativos (familia ART).

2. Familia SEM. Análisis lingüístico-semántico comparado [Cfr. Tabla 2. Anexo I]

La familia SEM se centra en el contenido lingüístico y semántico de los textos. A través de consultas SPARQL orientadas a la extracción léxica, la identificación de entidades semánticas y la detección de estructuras cognitivas, esta familia permite realizar comparaciones sistemáticas entre instituciones, autores y medios. SEM-TFIDF traza la proximidad léxica entre museos e instituciones, revelando afinidades o divergencias discursivas; SEM-CHI2 aísla las keywords que singularizan a cada entidad y a cada autor, señalando los rasgos diferenciales; SEM-AUTH agrupa firmas lingüísticas de críticos y comisarios, identificando agrupaciones de voz o posiciones marginales; SEM-METR identifica las metáforas y frames (marcos cognitivos) que estructuran la percepción colectiva de los artistas y cuantifica su distribución; finalmente, SEM-LEX perfila el repertorio léxico individual, midiendo amplitud y sesgos temáticos. Este nivel transforma el corpus en un mapa léxico-semántico multidimensional

que permite, por ejemplo, identificar qué museos privilegian ciertos vocabularios, qué críticos comparten marcos cognitivos, o cómo varía el sesgo valorativo entre autores institucionales e independientes. El conjunto, cruzado con la arquitectura textual (SYS) y la lectura narrativa específica de un artista (ART), permite entender quién dice qué, cómo lo dice y qué narrativa resultante se consolida.

3. Familia ART. Construcción narrativa de artistas [Cfr. Tabla 3. Anexo I]

La familia ART se orienta al análisis de las narrativas construidas en torno a artistas concretos. Las consultas de este bloque permiten destilar los temas curatoriales predominantes (ART- CURATION), medir la polaridad valorativa asociada al artista (ART-SENT), identificar periodos, estilos u obras sobrerrepresentadas (ART-TOPIC), y cartografiar los atributos adjetivales más frecuentes en su caracterización pública (ART-PERSONA). Asimismo, se rastrean metáforas y marcos conceptuales que estructuran su percepción (ART-META).

En su conjunto, este nivel analiza cómo se configura la imagen de un artista en el imaginario colectivo a través del discurso institucional y mediático. Permite observar desplazamientos en la narrativa —por ejemplo, de «genio revolucionario» a «figura problemática»— o contrastes entre la curaduría oficial y la crítica independiente.

4. Análisis integrado [Cfr. Tabla 4. Anexo I]

Los tres niveles analíticos descritos —estructural (SYS), semántico (SEM) y narrativo (ART)— se articulan en un marco integrado que permite abordar de manera global la construcción discursiva del ámbito expositivo. Esta integración no solo amplía la capacidad interpretativa del modelo, sino que posibilita la formulación de preguntas complejas que trascienden el análisis de capas aisladas. Gracias a esta arquitectura integrada, es posible explorar cuestiones como: ¿Qué redes curatoriales sostienen los marcos dominantes en torno a un artista? ¿Dónde se producen fisuras discursivas entre el relato institucional y la crítica independiente? ¿Cómo influye la estructura de comisariado en la propagación de determinadas categorías discursivas? ¿Qué comisarios actúan como «nodos bisagra», transfiriendo keywords entre instituciones? ¿Existen fenómenos de contagio semántico entre museos? ¿Varía la firma léxica de un autor en función de la institución con la que colabora?

4. Estudio de caso

Para demostrar la aplicabilidad de la metodología propuesta, estamos trabajando sobre un estudio de caso centrado en las exposiciones dedicadas a Picasso celebradas en España entre 2010 y 2024, y en el subsistema lingüístico-textual que dichas exposiciones han generado. La figura de Picasso constituye un terreno especialmente fértil debido a la diversidad de sus facetas —pintor, grabador, ceramista—, la elevada frecuencia expositiva y, especialmente, la

intensa recepción crítica y mediática que suscita. Estos factores configuran un corpus amplio y heterogéneo que cubre enfoques y lenguajes diversos. El estudio se articula en torno a un corpus textual plural que engloba los cuatro principales centros de producción discursiva sobre el artista en España: Museo Picasso Málaga (MPM), Museo Casa Natal Picasso (MCNP), Museo Picasso Barcelona (MPB) y Museo Reina Sofía (MNCARS). El corpus lingüístico-textual se compone de: 96 ensayos de catálogo; 57 textos curatoriales breves; 312 artículos de prensa y un volumen de 600.000 palabras.

Aunque los resultados se presentarán y discutirán durante la comunicación en el congreso, los primeros experimentos con el grafo confirman su capacidad para desagregar la narrativa picassiana en capas discursivas diferenciadas y revelar tensiones internas que no afloran mediante aproximaciones lineales.

a) Nivel global del ecosistema: Las métricas SYS-DENS y SEM-TFIDF permiten identificar un núcleo léxico persistente (‘genio’, ‘icono’, el binomio ‘tradición/vanguardia’, el eje ‘moderno/modernidad’) rodeado de satélites temáticos que ascienden y declinan según coyunturas históricas o curatoriales (‘misoginia’, ‘antifascismo’, ‘monstruosidad’). Los clústeres detectados por SEM-METR localizan dos ejes dominantes en la estructuración de los marcos dominantes (o frames): la innovación creativa y la figura de Picasso como icono político. Esta cartografía global sirve de referencia para leer desviaciones o desplazamientos locales.

b) Nivel institución: Al aislar los textos emanados por cada institución, el grafo evidencia huellas semánticas diferenciadas[8]:

- El Museo Picasso Málaga refuerza la categoría «biografía» y activa el frame de la raíz andaluza.
- El Museo Reina Sofía prioriza el eje «político», reactivando el frame Guernica/compromiso, con un léxico de alta carga ideológico-belicista (‘militante’, ‘trinchera’).
- El Museo Picasso Barcelona oscila entre las categorías de «vanguardia» y «exilio», con un vocabulario centrado en el «desarraigo».

Museo	% BIOGRAFÍA	% POLÍTICO	% VANGUARDIA	% EXILIO-DESARRAIGO
MPM (Málaga)	**42 %**	9 %	17 %	6 %
MNCARS (Reina Sofía)	11 %	**34 %**	15 %	10 %
MPB (Barcelona)	18 %	14 %	**26 %**	**22 %**
Media global	19 %	18 %	19 %	11 %

Valores ≥ 10 p.p. sobre la media aparecen en negrita

[8] Los planes estratégicos publicados por cada institución confirman la intención de «profundizar en la dimensión biográfica» (MPM), «reivindicar el compromiso político» (MNCARS) o «releer la vanguardia desde Barcelona» (MPB), consolidando la recurrencia de estos *frames*.

c) Curaduría vs. crítica especializada: La combinación de las métricas SEM-CHI2 (léxico distintivo) con ART-SENT (polaridad valorativa) permite localizar fisuras discursivas recurrentes entre el relato institucional y la crítica independiente. Un ejemplo especialmente revelador se observa en torno a la representación del periodo final de Picasso: mientras los textos institucionales ensalzan esta etapa con expresiones como 'etapa de felicidad', 'vitalidad', 'reinvención radical', etc., la crítica especializada tiende a problematizarla con términos como 'viejo', 'desfase ideológico', 'saturación', 'exaltación forzosa'[9].

Estos primeros indicios apuntan a que el grafo no solo permite identificar quién habla y cuánto, sino también cómo se transforma el significado de los términos al cruzar capas institucionales, autorales y mediáticas. Con la incorporación de otros análisis complementarios, aspiramos a mostrar durante nuestra presentación en el congreso el potencial del sistema metodológico propuesto para abordar el análisis discursivo del ámbito expositivo desde una perspectiva multiescala y comparativa.

5. Conclusiones y líneas futuras

Este trabajo ha presentado un marco metodológico para el análisis multidimensional del subsistema lingüístico-textual del ámbito expositivo, estructurado en tres niveles analíticos (estructural, semántico y narrativo) e implementado a través de consultas SPARQL sobre un grafo de conocimiento RDF basado en la ontología OntoExhibit. A partir de un estudio de caso centrado en la figura de Picasso, hemos demostrado de manera preliminar cómo esta arquitectura permite descomponer el ecosistema discursivo de manera escalable y comparativa, revelando patrones, tensiones y trayectorias semánticas que resultan inaccesibles mediante enfoques tradicionales. Consideramos que el modelo aquí descrito ofrece una vía robusta para avanzar hacia una cartografía discursiva compleja del hecho expositivo, articulando estructura, lenguaje y narrativa en un mismo entorno de análisis. Por su parte, el diseño es replicable y escalable: los ejes de análisis y el vocabulario controlado pueden adaptarse a otras figuras artísticas o a fenómenos expositivos distintos.

Para asegurar la transparencia, el conjunto completo de consultas SPARQL se depositarán en el GitHub del proyecto Complexhibit inmediatamente después de la presentación en el Congreso HDH 2025, acompañados de ejemplos comentados y una guía rápida de despliegue. Este compromiso abre la puerta a estudios comparativos, validación externa y extensión del grafo con nuevos corpus.

Como línea de desarrollo futuro, se plantea: (i) incorporar analítica visual dinámica que permita representar la evolución de los marcos dominantes (o frames) en el tiempo; (ii) vincular el grafo con métricas de audiencia y datos

[9] En los 29 ítems coincidentes (mismo evento o periodo 1963-73) la desviación media entre polaridades institucionales y periodísticas es > 0,9 DE, y los términos distintivos presentan $\chi^2 > 15$ ($p < 0,01$), lo que valida empíricamente la fisura discursiva.

procedentes de redes sociales, ampliando así la dimensión de recepción pública; y (iii) aplicar la metodología a exposiciones colectivas, donde la competencia entre relatos y la multiplicidad de voces hacen aún más compleja la configuración discursiva. Asimismo, la integración de una red de coapariciones de artistas en exposiciones —a nivel museográfico e institucional— se considera una vía prioritaria para futuras investigaciones. Lo ideal sería conectar esta dimensión con los grafos semántico-discursivos en análisis multiescala, aunque por cuestiones de alcance, este desarrollo queda reservado para trabajos posteriores.

Figuras, tablas, mapas o gráficas

Tabla 1. Familia SYS. Arquitectura de subsistema lingüístico-textual

Etiqueta	Propósito	SPARQL (extracto)	Salida
SYS-DENS	Medir la densidad textual por exposición	SELECT ?exhib (COUNT(?txt) AS ?n) WHERE { ?txt exo:relatesTo ?exhib } GROUP BY ?exhib	n.º textos ↔ exposición
SYS-TIPO	Distribuir texto por géneros (catálogo, prensa, web, etc.)	SELECT ?genre (COUNT(?txt) AS ?n) WHERE { ?txt a ?genre . } GROUP BY ?genre	matriz género textual-exposición
SYS-MEDIA	Medir el peso relativo de medios (periódicos, revistas, portales)	SELECT ?media (COUNT(?rev) AS ?n) WHERE { ?rev a exo:ArtReview ; exo:publishedIn ?media } GROUP BY ?media	ranking de cabeceras según producción texual
SYS-AUTHOR	Calcular el peso de los autores según su presencia global y por museo	SELECT ?author (COUNT(?txt) AS ?nTexts) (COUNT(DISTINCT ?museum) AS ?nMuseums) WHERE { ?txt dct:creator ?author ; exo:relatesTo ?exhib . ?exhib exo:hostedBy ?museum . } GROUP BY ?author ORDER BY DESC(?nTexts)	ranking de autores por n.º de textos. índice de alcance = nMuseums / nTexts (mide versatilidad de los autores).
SYS-MED-MUS	Analizar la conexión museo ↔ medio (cobertura)	SELECT ?museum ?media (COUNT(?rev) AS ?n)… GROUP BY ?museum ?media	tabla cobertura cruzada

SYS-ACTOR+	Obtener una red autor-institución con atributo de autor independiente/institucional (según los textos sean encargos de la institución organizadora o no)	CONSTRUCT { ?author exo:role ?status ; exo:actsOn ?museum } WHERE { ?txt dct:creator ?author ; exo:relatesTo ?exhib . ?exhib exo:hostedBy ?museum . OPTIONAL { ?txt exo:commissionedBy ?museum } BIND(IF(BOUND(?museum) && ?txt a exo:CuratorialText , "institucional" , "independiente") AS ?status) }	grafo bipartito + atributo
SYS-TEMP	Obtener una cronología de publicaciones	SELECT (YEAR(?date) AS ?y) (COUNT(?txt) AS ?n) WHERE { ?txt dct:date ?date } GROUP BY ?y	serie temporal normalizada

Tabla 2. Familia SEM. Análisis lingüístico-semántico comparado

Etiqueta	Propósito	SPARQL (extracto)	Métrica
SEM-TFIDF	Vectorizar vocabulario por museo o institución y medir proximidad léxica	SPARQL\nSELECT ?museum ?lemma (COUNT(?t) AS ?freq)\nWHERE {\n ?txt a exo:CuratorialText ; # o exo:ArtReview\n exo:hasLemma ?lemma ;\n exo:relatesTo ?exhib .\n ?exhib exo:hostedBy ?museum .\n}\nGROUP BY ?museum ?lemma (se exporta a CSV y se normaliza en Python → TF-IDF)	Matriz TF-IDF museo × lema. Similitud coseno entre museos (0–1). Clúster jerárquico (Ward) para agrupar instituciones con léxico afín. Dendrograma de proximidad.
SEM-CHI2	Detectar términos significativamente distintivos por museo/institución y por autor.	SPARQL\nSELECT ?group ?lemma (COUNT(?t) AS ?k)\nWHERE {\n VALUES ?groupType { \"museum\" \"author\" }\n { ?txt exo:relatesTo ?exhib . ?exhib exo:hostedBy ?group . }\n UNION\n { ?txt dct:creator ?group . FILTER(?groupType=\"author\") }\n ?txt exo:hasLemma ?lemma .\n} GROUP BY ?group ?lemma	Tabla de contingencia grupo × lema. χ^2 con Yates y p-value < 0.05. Top-20 keywords por grupo ordenadas por χ^2. Indicador std.residual para peso relativo.

SEM-AUTH	Comparar «firma» discursiva de autores y detectar comunidades de estilo	SPARQL\nSELECT ?author ?lemma (COUNT(?t) AS ?freq)\nWHERE { ?txt dct:creator ?author ; exo:hasLemma ?lemma }\nGROUP BY ?author ?lemma	Vectores TF-IDF autor × lema. Matriz de similitud coseno (autor↔autor). Louvain para comunidades. Betweenness de autores- puente.
SEM-METR	Identificar metáforas y frames dominantes y su distribución	SPARQL\nSELECT ?entity ?frame (COUNT(?occ) AS ?n)\nWHERE {\n ?occ exo:instantiatesFrame ?frame ;\n exo:appearsIn ?txt .\n { ?txt exo:relatesTo ?exhib . ?exhib exo:hostedBy ?entity } # museo\n UNION\n { ?txt dct:creator ?entity } # autor\n} GROUP BY ?entity ?frame	Frecuencia y % relativo de cada frame por entidad. Frame Dominance Index = freqFrame / ΣframesEntidad. PMI frame–categoría CS. Heat-map entidad × frame.
SEM-LEX	Perfilar el repertorio léxico individual con foco temático y valorativo	SPARQL\nSELECT ?author ?lemma ?cs (COUNT(?t) AS ?freq)\nWHERE {\n ?txt dct:creator ?author ;\n exo:hasLemma ?lemma .\n OPTIONAL { ?lemma exo:inCategory ?cs }\n} GROUP BY ?author ?lemma ?cs	Type–Token Ratio (TTR) y HD-D por autor. Distribución % de categorías CS por autor. Sesgo valorativo (media de exo:sentimentValue) por categoría. Radar-chart para firmas individuales.

Tabla 3. Familia ART. Construcción narrativa de artistas

Etiqueta	Propósito	Esquema SPARQL (extracto)	Output previsto
ART-CURATION	Extraer tópicos curatoriales	SELECT ?lemma (COUNT(?lemma) AS ?n) WHERE { ?txt a exo:CuratorialText ; exo:hasLemma ?lemma } GROUP BY ?lemma	Lista lemas + frecuencia → LDA
ART-META	Identificar metáforas conceptuales y frames en torno al artista	`SELECT ?txt WHERE { ?txt exo:containsPhrase ?m . FILTER regex(?m,"genio	Ranking de frames dominantes por museo, autor o medio.
ART-SENT	Medir polaridad valorativa	SELECT ?txt ?pol WHERE { ?txt a exo:ArtReview ; exo:sentimentPolarity ?pol }	Vector de sentimiento → promedios
ART-TOPIC	Detectar la sobrerrepresentación de periodos creativos, temas, obras.	SELECT ?period (COUNT(?ref) AS ?n) WHERE { ?txt exo:mentionsPeriod ?period } GROUP BY ?period	Ranking y porcentaje de menciones por museo/autor
ART-PERSONA	Analizar la construcción adjetival en torno al artista (“genio”, “monstruo”, “icono político”, “visionario”)	`SELECT ?adj (COUNT(?adj) AS ?n) WHERE { ?txt exo:hasAdjective ?adj . FILTER regex(?adj,"genio	Listado de locuciones adjetivales por institución, medio y autor

Tabla 4. Resumen integrado de los tres niveles analíticos

Aspecto clave	Nivel SYS – Arquitectura del subsistema lingüístico-textual	Nivel SEM – Análisis semántico comparado	Nivel ART – Construcción discursiva de artista
Objeto primario	La infraestructura y el flujo de los textos: volumen, géneros, medios, autores, instituciones y cronología.	El discurso en sí mismo: léxico, categorías semánticas y marcos cognitivos empleados en los textos, comparados entre museos, medios y autores.	La imagen narrativa del artista que emerge de ese discurso: temas, periodos, metáforas y valoración.

Preguntas guía	¿Cuánta producción textual genera cada exposición? ¿Qué medios cubren a qué museos? ¿Qué proporción de voces es independiente vs. institucional?	¿Qué categorías discursivas (“creatividad”, “político”, etc.) privilegia cada actor? ¿Cómo varía el vocabulario, las metáforas y los frames entre autores, medios y museos? ¿Qué firmas semánticas distinguen a autores instituciones vs. independientes?	¿Qué tópicos biográficos y estéticos se priorizan? ¿Con qué adjetivos se le representa y con qué polaridad? ¿Cómo fluctúa la atención a cada periodo creativo?
Consultas núcleo	SYS-DENS, SYS-TIPO, SYS-MEDIA, SYS-AUTHOR SYS-MED-MUS, SYS- ACTOR+, SYS-TEMP	SEM-TFIDF, SEM-CHI2, SEM- CS-MUS, SEM-CS-AUTH, SEM-METR, SEM-EVAL, SEM-SHIFT	ART-CURATION, ART- META, ART-SENT, ART-TOPIC, ART-PERSONA
Métricas / salidas	Cobertura textual Matriz museo × medio Grafo autor-institución con rol (independiente/institucional) Serie temporal de publicaciones	Palabras-clave distintivas Clústeres de autores por firma léxica Similaridad léxica Ranking de metáforas y frames Polaridad evaluativa por rol	Ranking de tópicos curatoriales Polaridad media por institución Léxico que construye la narrativa sobre un artista Evolución de la valoración del artista
Escala de análisis	Estructural y cuantitativa: cartografía el subsistema lingüístico-textual y las relaciones entre sus agentes.	Comparativa y lingüística: mide cómo se modela el discurso de manera global y sus diferencias internas.	Interpretativa y temática: destila cómo ese entramado discursivo configura la narrativa sobre el artista.
Relación entre niveles	Proporciona el esqueleto de datos (quién escribe, dónde, cuándo).	Proporciona el mapa semántico base (qué se dice, con qué palabras y en qué marcos).	Traduce esos patrones a quién es este artista en la narrativa colectiva.

Ejemplos de preguntas cruzadas

Pregunta	Niveles utilizados	Flujo analítico
¿Qué redes curatoriales sostienen los frames dominantes?	**SYS-ACTOR+, SEM-METR**	1) Extraer el grafo comisario ↔ museo (SYS-ACTOR+) para localizar curadores que trabajan en varias instituciones.2) Calcular la frecuencia de cada frame en los textos de esas instituciones (SEM- METR).3) Correlacionar presencia del comisario multi-sede con la difusión de frames como «laboratorio creativo» o «icono político».
¿Dónde se producen «fisuras discursivas» entre el relato institucional y la crítica independiente?	**SEM-CHI2, PIC-SENT**	1) Medir la divergencia léxica entre catálogos y reseñas (SEM-CHI2) para cada exposición.2) Calcular la diferencia de polaridad entre textos institucionales e independientes (PIC-SENT).3) Señalar exposiciones con alta desviación en ambos indicadores, revelando desajustes narrativos.
¿Cómo influye la estructura de co-comisariado en la dispersión de categorías discursivas?	**SYS-ACTOR+, SEM-CS-MUS**	1) Construir sub-grafos de coautoría de comisarios (SYS-ACTOR+) y detectar equipos compartidos.2) Generar la matriz museo × categoría CS (SEM-CS-MUS).3) Comparar el perfil semántico de museos conectados por co-comisarios y medir convergencia o divergencia de categorías.
¿Cómo varía la firma léxica de un autor al pasar de un medio a otro?	**SYS-MEDIA, SEM-AUTH**	1) Identificar los nodos del mismo autor en diferentes soportes (SYS- MEDIA).2) Vectorizar su vocabulario por soporte con TF-IDF (SEM-AUTH).3) Calcular la similitud coseno entre firmas y cuantificar cambios estilísticos entre formatos.
¿Qué comisarios actúan como «nodos bisagra» que transfieren keywords de un museo a otro?	**SYS-ACTOR+, SEM-CHI2**	1) Trazar las rutas temporales de cada comisario entre museos (SYS- ACTOR+).2) Generar series mensuales de keywords distintivas por museo (SEM-CHI2).3) Detectar comisarios cuyo desplazamiento coincide con la adopción rápida de nuevas keywords en la institución receptora.
¿Existe un efecto de contagio semántico desde las instituciones museísticas a los medios?	**SYS-TEMP, SEM-CHI2**	1) Recuperar la fecha de publicación de cada exo:WebText institucional (SYS-TEMP).2) Monitorizar la frecuencia de keywords homólogas en prensa antes y después del lanzamiento (SEM- CHI2).3) Evaluar si se produce una subida estadísticamente significativa que indique contagio semántico hacia los medios.

Bibliografía

BRADEN, L. E. A., & TEEKENS, T. (2020), «Historic networks and commemoration: Connections created through museum exhibitions», *Poetics: Journal of Empirical Research on Culture, the Media and the Arts*, 81, art. 101446. https://doi.org/10.1016/j.poetic.2020.101446

FARQUHARSON, A. (2006), «The ecology of exhibitions», en S. MACDONALD & P. BASU (eds.), *Exhibition Experiments*, Blackwell, pp. 13–30.

QUILES, D. R. (2014), «Exhibition as network, network as curator: Canonizing art from "Latin America"», *Artl@s Bulletin*, 3(1), pp. 62–78.

PROCTOR, N. (2017), «The museum as distributed network, a 21st-century model», *Museum-ID*. Disponible en: https://museum-id.com/museum-distributed-network-21st-century-model-nancy-proctor/ [Acceso: 22/06/2025].

RODRÍGUEZ-ORTEGA, N. (2024), «Contours of Knowledge: Epistemological Implications of Semantic Models in the Representation of the Art Exhibition Domain through the Lens of the OntoExhibit Ontology», *Život umjetnosti*, vol. 114, nº 1, pp. 122–147. https://doi.org/10.31664/zu.2024.114.06

RODRÍGUEZ ORTEGA, N., ROLDÁN GARCÍA, M.ª M., PLATAS DÍEZ, M.ª L., & SALVACHÚA, M. (2024), «OntoExhibit: An Ontology for Modeling the Field of Art Exhibitions and Its Semantic-Discursive Dimension», en J. KARAJGIKAR, A. JANCO, & J. OTIS (eds.), *Digital Humanities 2024: Book of Abstracts*. https://doi.org/10.5281/zenodo.13761079

RODRÍGUEZ-ORTEGA, N., DÍEZ PLATAS, M.ª L., ORTIZ TELLO, M., & LUMBRERAS FERNÁNDEZ, A. (2025), «Bridging Discourses: Integrating Text Catalogs and Art Reviews into Knowledge Graphs for Enriched Exhibition Analysis», *Digital Humanities Annual Conference 2025*, Lisboa, (14–18 junio 2025).

Novus ordo seclorum? Retos y posibilidades para la visualización de datos espaciotemporales en las humanidades digitales

Jorge Sebastián Lozano
Universitat de València

Palabras clave: Visualización de datos; Humanidades digitales; Heterogeneidad; Ambigüedad; Granularidad.
Keywords: Data visualization; Digital humanities; Heterogeneity; Ambiguity; Granularity.

Las humanidades, como campo del saber, se caracterizan por la acumulación y el entrecruzamiento de múltiples categorías conceptuales. Además, esos conceptos o herramientas propios de cada disciplina humanística no solamente son aplicados en distinta medida y con diferentes criterios, sino que a menudo son sometidos a crítica y deconstrucción por sus propios usuarios. Por ello, habitualmente la metodología de investigación acaba siendo uno de los campos en los que más profundos resultan los desacuerdos entre quienes supuestamente practican una misma disciplina.

Sin embargo, tiempo y espacio son dos categorías transversales a todos los estudios humanísticos. Cualquier juicio o afirmación humanística se encuadra, de manera implícita o explícita, dentro de unas circunstancias espacio temporales. No solo las propias de quien enuncia algo (que también, aunque eso no sea exclusivo de las humanidades), sino principalmente las referidas al asunto sobre el cual se pronuncia. Todo conocimiento humanístico implica ineludiblemente unas coordenadas de cronología y ubicación, que en muchos casos son determinantes para el conocimiento que se formula. Hasta cierto punto, todo saber humanístico es, además de otras cosas, un saber histórico y geográfico. Todo hecho humano está, ha estado, inextricablemente incardinado en una cuadrícula espacio temporal.

En este contexto, esta comunicación se interroga por las formas en que las humanidades digitales están suponiendo rupturas y continuidades en los estudios humanísticos, especialmente en lo referente a ese encuadramiento en el tiempo y el espacio. La cronología y la periodización, las influencias entre zonas culturales, los enfoques comparatistas, las inferencias causales acerca de acontecimientos coetáneos o sucesivos, el análisis y construcción de identidades o sujetos históricos y geográficos, las taxonomías y jerarquías tácitas o explícitas… son objetos permanentes de estudio en las disciplinas humanísticas,

sobre los que están impactando la disponibilidad y complejidad crecientes de las herramientas digitales. Particularmente, es necesario profundizar en cómo la visualización de datos (en cuanto campo especializado dentro de la ciencia de datos) invita a realizar análisis más cuantitativos dentro de disciplinas muy marcadamente cualitativas, como son las humanísticas. En el fondo subyace la pregunta de cómo las herramientas, datos y métodos digitales -principalmente, la visualización de datos- pueden alterar nuestra comprensión del pasado.

Para ello, se realiza una panorámica somera sobre las principales formas de visualización espacio temporal usadas tradicionalmente por las humanidades. Hablamos de herramientas o dispositivos como son los mapas, las rutas, los anales, las líneas de tiempo, las tablas de periodos… más otros en los que las dos coordenadas básicas sustentan niveles añadidos de información y/o interpretación: árboles genealógicos, redes de agentes participantes en un fenómeno concreto, gráficas sobre magnitud o distribución de determinadas variables cuantitativas… Todas estas formas de visualización tienen una historia propia, más o menos desarrollada, anterior a su actual eclosión digital como disciplina ya casi independiente. A ese respecto, resulta innegable el mayor desarrollo de la visualización espacial, respecto a la temporal: disciplinas auxiliares como la cronología o la cronografía son casi anécdotas cuando se comparan con el actual apogeo de la geografía, la geomática o la cartografía, precisamente gracias a sus potentes y ubicuas herramientas digitales. Por lo mismo, cabe aventurar que la investigación sobre visualizaciones cronológicas tiene todavía un gran camino por recorrer. Las diferentes concepciones antropológicas y filosóficas sobre el tiempo, como un hecho cíclico, lineal o caótico, junto con sus combinaciones, ofrecen posibles nuevas formas de abordarlo.

Junto a las consecuencias, conviene preguntarse también por los requisitos, las demandas que estos tipos de análisis imponen sobre el trabajo humanístico. El más evidente es la necesidad de contar con repositorios de datos digitales de calidad suficiente con los que trabajar - o, dada su frecuente inexistencia, la de construirlos. La normalización de la información y la interoperabilidad entre repositorios, o la extensión de una cultura de acceso abierto a los datos, son aspectos colaterales de esa necesidad. Otra demanda es un mayor conocimiento y adopción de métodos estadísticos y cuantitativos, algo a lo en las humanidades que hemos sido particularmente renuentes. De ningún modo se aboga aquí por un olvido o relegamiento de lo cualitativo, que está y seguirá estando en el centro mismo de cualquier empeño humanístico; más bien se trata de una ampliación, de una adición de análisis complementarios, que en algunos casos puedan enriquecer, matizar o someter a crítica el conocimiento heredado. A ese respecto, explorar nuevas narrativas, nuevas formas de relato o de razonamiento histórico, son retos particularmente importantes, y se abordan algunos casos recientes.

Por ejemplo, traslaciones digitales de formatos tan clásicos como la línea de tiempo han sido muy exitosas entre el público general, en sitios web de

museos tan importantes como el Prado, el Museo Nacional Centro de Arte Reina Sofía o el Metropolitan Museum. En un contexto de investigación, los grafos de redes están siendo fructíferamente utilizados para visualizar los ecosistemas comerciales del arte moderno y contemporáneo, gracias a proyectos como Artl@s o Exhibitium. Múltiples aplicaciones basadas en Sistemas de Información Geográfica incluyen mapas dinámicos, que añaden la dimensión temporal a la representación espacial. Tales sistemas están demostrando una gran utilidad para fomentar nuevas interpretaciones acerca de la extensión o la evolución de movimientos culturales o grupos humanos. Pero también hay esfuerzos más experimentales, en la línea de visualizar cantidades masivas de datos a través de gráficas interactivas, que permiten poner a prueba, con bases cuantitativas, ideas comúnmente aceptadas en las investigaciones previas. Estas iniciativas muestran especial interés como herramientas hermenéuticas, como invitaciones a repensar categorías y relatos dominantes. Las visualizaciones espaciales y temporales ayudan de manera extraordinaria a visibilizar puntos ciegos y ausencias en las narrativas de base textual, tan predominantes en las humanidades.

Todas estas realizaciones, analizadas de forma crítica, permiten explicar cómo las peculiaridades de la información humanística, de sus preguntas de investigación, la distinguen notablemente de otros campos de estudio que tradicionalmente ya usaban métodos y visualizaciones basados en estadísticas. No se trata de un camino sencillo ni unidireccional. De hecho, los datos espacio temporales propios de las humanidades adolecen normalmente de problemas que suponen otros tantos graves retos para las herramientas y métodos computacionales. Se trata de características como la heterogeneidad o falta de normalización en la terminología; la ambigüedad o incertidumbre en los datos disponibles o de los propios conceptos subyacentes, por una inevitable falta de conocimientos, o por dudas y desacuerdos entre los expertos; su muy variable nivel de granularidad, con datos a veces extraordinariamente bien estructurados, y otras veces irremediablemente genéricos. Cualquier intento de visualizar información humanística con sus coordenadas espaciales y temporales ha de tomar conciencia de tales dificultades, y valorar maneras innovadoras de incorporarlas, sin pretender aplicar soluciones imitadas de otras disciplinas, poco o nada sensibles a las especificidades del saber humanístico.

Todas estas consideraciones forman el núcleo del proyecto de I+D+i “ClioViz: Visualización avanzada de datos históricos a través de mapas multidimensionales y gráficos interactivos” (ref. PID2021-126777NB-I00, financiado/a por MICIU/AEI/10.13039/501100011033/ - FEDER Una manera de hacer Europa), del cual forma parte esta comunicación.

Bibliografía

BENITO-SANTOS, A., DORAN, M., ROCHA, A., WANDL-VOGT, E., EDMOND, J., & THERÓN, R. (2021), «Evaluating a taxonomy of textual uncertainty for collaborative visualisation in the digital humanities», *Information*, 12, art. 436. https://doi.org/10.3390/info12110436

DRUCKER, J. (2020), *Visualization and Interpretation: Humanistic Approaches to Display*, Cambridge, MIT Press.

LIPPINCOTT, K., ECO, U., GOMBRICH, E. H., et al. (2000), *El tiempo a través del tiempo*, Barcelona, Grijalbo.

GREENWALD, D. S. (2021), *Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art*, Princeton, Princeton University Press.

MARTÍN-RODILLA, P., & GONZÁLEZ-PÉREZ, C. (2019), «Conceptualization and non-relational implementation of ontological and epistemic vagueness of information in Digital Humanities», *Informatics*, 6(2), art. 20. https://doi.org/10.3390/informatics6020020

MARTÍN-RODILLA, P., MARTÍN PEREIRA-FARIÑA, P., & GONZÁLEZ-PÉREZ, C. (2019), «Qualifying and quantifying uncertainty in Digital Humanities: A fuzzy-logic approach», en *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'19)*, Association for Computing Machinery, New York, pp. 788–794. https://doi.org/10.1145/3362789.3362833

MORETTI, F. (2005), *Graphs, Maps, Trees: Abstract Models for a Literary History*, Londres, Verso.

PORTALÉS, C., CASANOVA-SALAS, P., SEVILLA, J., SEBASTIÁN, J., LEÓN, A., & SAMPER, J. J. (2022), «Increasing access to cultural heritage objects from multiple museums through semantically-aware maps», *ISPRS International Journal of Geo-Information*, 11, art. 266. https://doi.org/10.3390/ijgi11040266

ROSENBERG, D., & GRAFTON, A. (2012), *Cartographies of Time: A History of the Timeline*, Princeton Architectural Press.

SEBASTIÁN LOZANO, J. (2021), «Mapping art history in the digital era», *The Art Bulletin*, 103(3), pp. 6–16. https://doi.org/10.1080/00043079.2021.1882819

Analíticas y visualización de datos a partir de modelos heterogéneos de discurso

César González-Pérez
Beatriz Calderón-Cerrato
Incipit CSIC
Martín Pereira-Fariña
Instituto de Humanidades (iHUS) –
Universidade de Santiago de Compostela

Palabras clave: Análisis de discurso; analíticas; visualización de datos; argumentación; ontologías
Keywords: Discourse analysis; analytics; data visualization; argumentation; ontologies

Problema

Los últimos años han visto un crecimiento enorme de las tecnologías de procesamiento del lenguaje natural. Disponemos de muchísimas maneras de analizar textos computacionalmente. Sin embargo, casi todos estos enfoques analíticos poseen una serie de características que los limitan fuertemente.

En primer lugar, muchos enfoques actuales de procesado analítico de textos, como los grandes modelos de lenguaje (LLMs), se basan en el uso de grandes volúmenes de textos, prefiriendo cantidad a calidad. Es difícil encontrar en la literatura avances en técnicas de análisis automático orientados a textos pequeños o fragmentos de discursos. En segundo lugar, los enfoques actuales se centran en el análisis de aspectos léxicos o gramaticales, siendo más difícil alcanzar niveles semánticos de los textos, y mucho más difícil llegar a los aspectos pragmáticos. Esto es esperable, ya que, desde un punto de vista computacional, tratar símbolos individuales o secuencias de símbolos es una tarea viable, mientras que tratar la conexión de estos símbolos con las entidades del mundo, o el contexto comunicativo y social de los textos, es muy difícil o imposible. En tercer lugar, los enfoques actuales utilizan el texto como input principal del procesado. En algunos casos, este texto ha sido anotado previamente por un humano, aunque esto, debido a lo tedioso de la tarea manual, limita severamente el volumen de texto que se puede analizar.

De este modo, la necesidad de grandes volúmenes de texto impide una anotación humana detallada y que pueda añadir riqueza. A su vez, la escasa anotación humana hace que los textos se reduzcan a meras secuencias de

palabras, con poca o nula información sobre su contexto comunicativo y social. Estos tres factores limitantes, así pues, se refuerzan mutuamente.

En esta situación, la visión que proponemos en este trabajo es la siguiente. Aunque el procesado masivo de enormes cantidades de textos es útil y valiosa, también es necesario avanzar en técnicas de análisis profundo y detallado de textos que puedan abordar los aspectos semántico y, sobre todo, pragmático de los discursos. Hacer esto reduce el volumen de texto que es factible procesar, y requiere que el análisis computacional, en vez de basarse de forma exclusiva en el texto como materia prima, utiliza modelos discursivos creados por humanos. Así pues, proponemos en este trabajo un nuevo tipo de procesado analíticos de textos, en el cual un analista humano comienza construyendo uno o más *modelos discursivos* de los textos a analizar. Un modelo discursivo es una representación del texto fuente que enfatiza aspectos determinados del texto. Qué aspectos se enfatizan depende del tipo de modelo. Por ejemplo, un modelo ontológico describe de qué habla el texto, y enfatiza las entidades, propiedades y relaciones que se mencionan en el texto. De forma similar, un modelo argumental describe cómo el texto justifica lo que dice, y enfatiza las proposiciones y relaciones argumentales (inferencias, conflictos, refraseos) que existen en el texto. Un modelo agencial, por último, enfatiza las creencias, deseos e intenciones que aparecen plasmadas en el texto. Sean del tipo que sean, los modelos discursivos se deben anclar sólidamente en el texto, y al mismo tiempo añadir riqueza y expresividad mediante la contribución del analista humano que un ordenador actual nunca podría obtener. Esta contribución humana puede referirse a los planos léxico o gramatical, pero también semántico o pragmático.

Una vez que el analista humano ha creado los modelos, estos constituyen la materia prima para el procesado analítico. Diferentes algoritmos son ejecutados sobre los modelos, produciendo textos estructurados y visualizaciones gráficas de estos modelos y, por tanto, de los textos subyacentes.

Metodología

La metodología empleada en este trabajo se ha basado en el paradigma de *design science* [5], de modo que el enfoque que proponemos, junto con las herramientas software que lo acompañan, fueron desarrollados al mismo tiempo que eran utilizados en distintos proyectos de análisis de discurso. Al comienzo, las versiones más iniciales del enfoque y las herramientas fueron adoptadas en proyectos diversos, como el análisis de la comunicación científica sobre la pandemia de covid en 2020 (https://pti-esciencia.csic.es/project/covid19-en-espanol-investigacion-interdisciplinar-sobre-terminologia-tematicas-y-comunicacion-de-la-ciencia/), o los fenómenos discursivos en torno al patrimonio cultural disonante del franquismo [1], [4]. Las experiencias obtenidas de estos primeros usos permitieron mejorar rápidamente el enfoque y las herramientas software, que siguieron siendo utilizadas a lo largo de 3 años y siendo

evolucionadas a la par, hasta alcanzar el estado actual. A lo largo de este tiempo, las versiones intermedias de estos productos también fueron utilizadas en numerosas actividades de formación, talleres y asesorías sobre análisis de discurso, de modo que analistas ajenos al desarrollo pudieron probarlas y emitir valoraciones y comentarios sobre ellas.

Resultados

El resultado principal de este trabajo es IAT/ML (www.iatml.org), una metodología multifocal para análisis de discurso [2], [3], y una herramienta software que la acompaña, LogosLink (www.iatml.org/logoslink). IAT/ML y LogosLink soportan los tres tipos de análisis discursivo esbozados arriba: ontológico, argumental y agencial.

LogosLink proporciona más de 60 analíticas automáticas que operan sobre los modelos discursivos confeccionados por analistas humanos. La lista completa, y su documentación, pueden encontrarse en el Manual de Usuario (http://www.iatml.org/LogosLinkDoc/UsersManual/en/Apps/Desktop/Analytics/). En este trabajo presentamos un subconjunto seleccionado: analíticas de denotaciones, analíticas de adhesión a posturas, y analíticas de razonamiento de posturas.

Analítica de denotaciones

Una denotación, en IAT/ML, es una conexión que el analista recoge entre un término en el texto y un concepto en la ontología. Por ejemplo, en el texto “la ermita es el único edificio que sobrevivió a aquel desastre”, el término “la ermita” se refiere la Ermita de Santa Catalina en Mansilla de la Sierra (La Rioja), mientras que “aquel desastre” se refiere a la inundación del pueblo viejo de Mansilla como consecuencia de la construcción de un embalse. Tanto la Ermita de Santa Catalina como la inundación son elementos en el modelo ontológico de este discurso. La analítica emplea los modelos ontológicos y argumentales para recoger datos; una vez que el analista ha registrado las denotaciones que estime oportunas, la analítica de denotaciones permite visualizar cuántas menciones existen de cada concepto, teniendo en cuenta, además, que los conceptos se relacionan entre sí mediante relaciones de subsunción, instanciación, o asociación semántica (Figura 4).

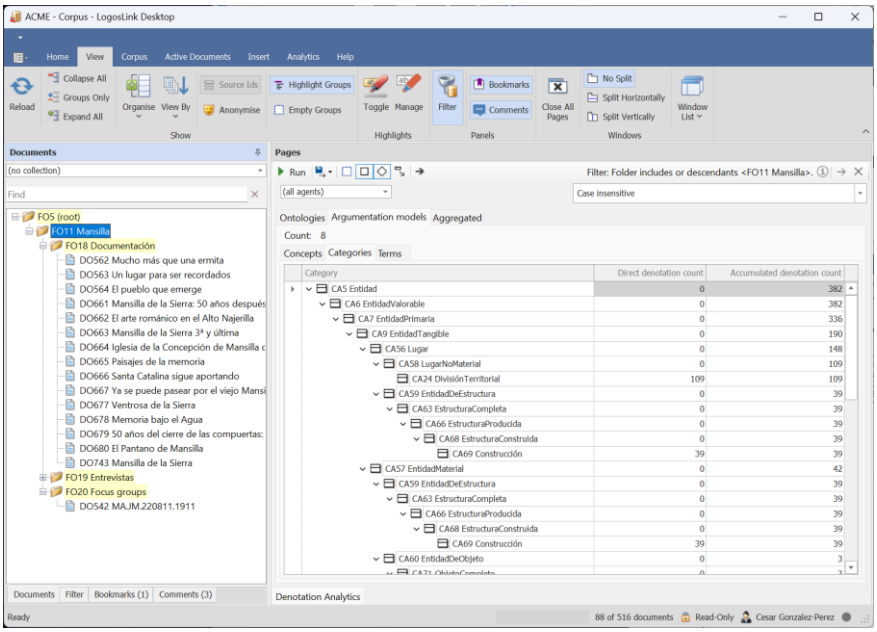


Figura 4. Jerarquía de subsunción de categorías en la analítica de denotaciones para 88 documentos sobre el pueblo inundado de Mansilla de la Sierra (La Rioja). Para cada categoría, se muestra el número de denotaciones directas y acumuladas.

Además, la analítica permite visualizar qué términos son utilizados por cada hablante, o grupo de hablantes, para referirse a cada concepto, de modo que es fácil encontrar usos cargados o selectivos de ciertos términos. Esta analítica se puede llevar a cabo sobre un texto individual o, de forma más interesante, sobre un conjunto de textos.

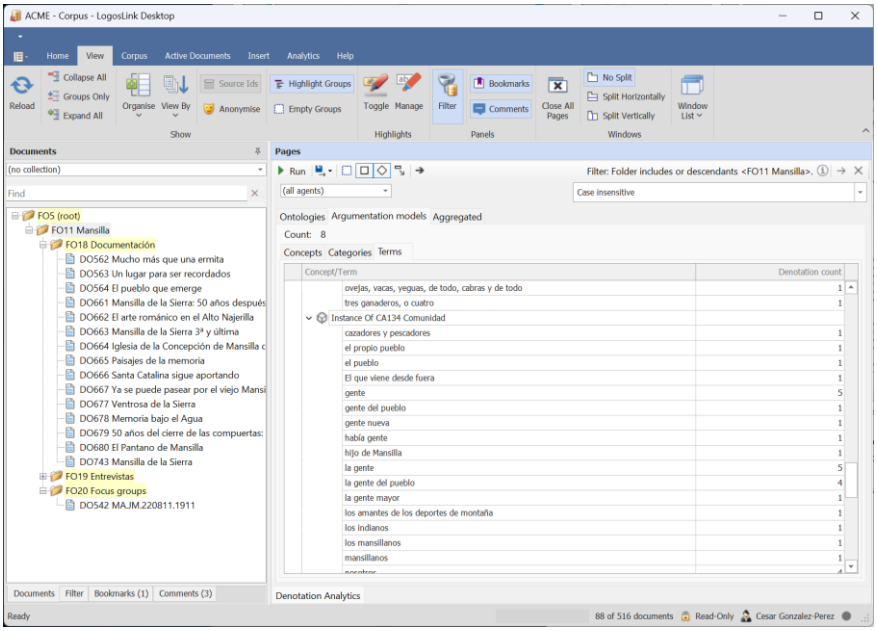


Figura 2. Terminología utilizada para denotar cada concepto en la analítica de denotaciones para 88 documentos sobre el pueblo inundado de Mansilla de la Sierra (La Rioja).

Analítica de adhesión a posturas

Una postura en IAT/ML es una afirmación que es defendida por algunos agentes y atacada por otros dentro de un grupo social. La analítica de adhesión a posturas utiliza los modelos argumentales y agenciales y los integra para ofrecer resultados agregados. Por ejemplo, el *Pont del Diable*, un puente histórico de gran valor patrimonial que conecta los ayuntamientos de Martorell y Castellbisbal en la provincia de Barcelona, a menudo se asocia con Martorell, aunque existe cierta reclamación por parte de Castellbisbal también. Al analizar los discursos de distintos agentes (vecinos de ambos lugares, técnicos, gestores, profesores, visitantes), la analítica de adhesión a posturas muestra en qué grado cada uno de ellos se opone o apoya la postura "el puente se debe asociar igualmente tanto a Martorell como a Castellbisbal". En la **¡Error! No se encuentra el origen de la referencia.**, la postura en cuestión es la etiquetada como PO10, en la parte inferior. Existen adhesiones en todos los grados; por ejemplo, el personal de los museos de Martorell, en color anaranjado, apoyan fuertemente la postura. Los investigadores, en rojo, están divididos, habiendo algunos que se oponen moderadamente a la postura, aunque otros la apoyan. Los visitantes, en verde, se oponen moderadamente o son neutros. Los profesores de los institutos de educación secundaria locales muestran desde la oposición fuerte hasta el apoyo fuerte. La figura también muestra el grado de controversia (*contentiousness*) de la postura, a la derecha en la tabla superior, expresado como una proporción respecto al máximo teórico.

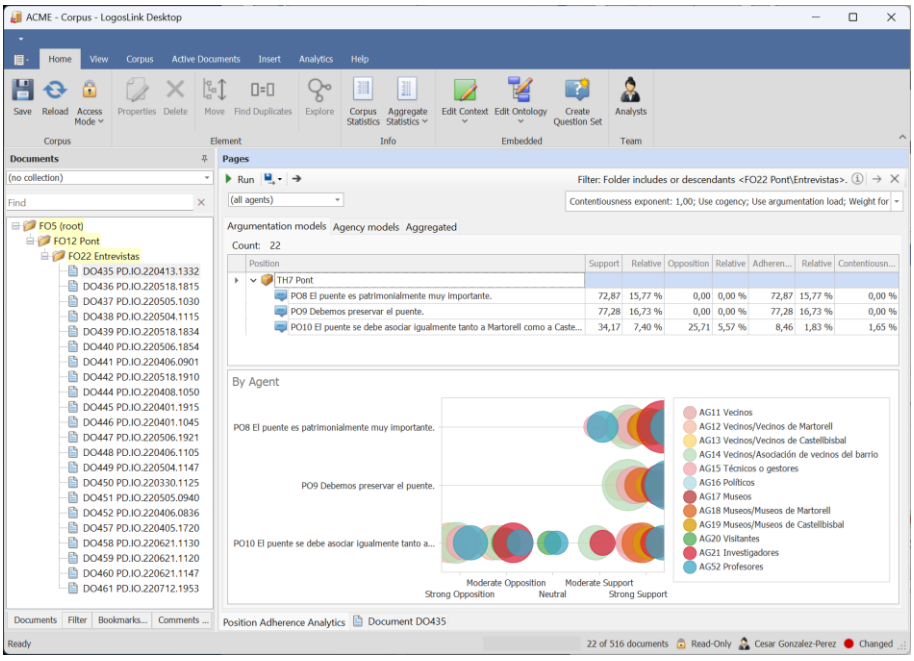


Figura 3. Resultados de la analítica de adhesión a posturas para el caso del Pont del Diable (Martorell/Castellbisbal, Barcelona).

Analítica de razonamiento de posturas

La analítica de razonamiento de posturas permite obtener una representación estructurada de las cadenas inferenciales y apoyos lingüísticos que los distintos hablantes emplean para defender o atacar cada postura. Esta analítica se basa en los modelos argumentales y agenciales que existan, y presenta resultados agregados. Siguiendo con el ejemplo del apartado anterior, la Figura 5 muestra qué evidencias contiene cada uno de los distintos textos analizados para apoyar u oponerse a cada postura. Por ejemplo, la figura muestra cómo el primer documento apoya fuertemente la postura mediante dos inferencias, y cómo la misma postura es atacada en el segundo documento mediante otras inferencias. Esta analítica es muy útil para comprender cómo distintos agentes, a través de distintas manifestaciones discursivas, han apoyado o atacado un conjunto de posturas, y en qué medida cada apoyo o ataque está fundamentado argumentalmente o mediante evidencias textuales.

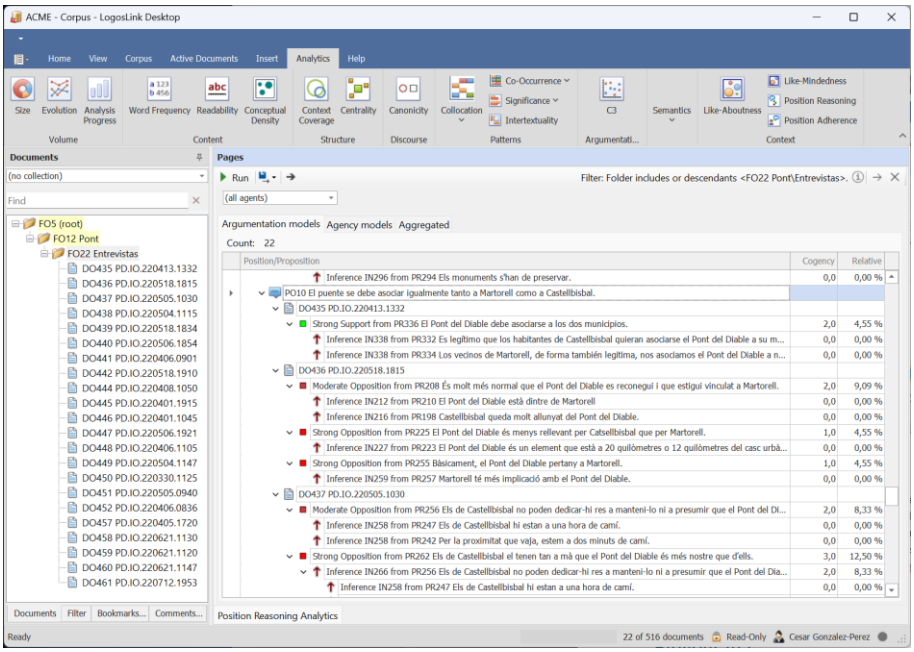


Figura 5. Resultados de la analítica de razonamiento de posturas para el caso del *Pont del Diable* (Martorell/Castellbisbal, Barcelona). La fuerza (*cogency*) absoluta y relativa de cada afirmación se muestra en las columnas de la derecha.

LogosLink cuenta con muchas otras analíticas, que permiten obtener información detallada sobre el contenido, estructura, patrones, semántica y contexto social de los textos analizados. De forma conjunta, estas analíticas proporcionan al especialista las herramientas necesarias para conseguir una mejor comprensión de los textos y poder desarrollar valoraciones, diagnósticos sociales y planes de acción más fundamentados.

Bibliografía

CALDERÓN-CERRATO, B.; GONZÁLEZ-PÉREZ, C.; & PEREIRA-FARIÑA, M. (2022), «Empirical evaluation of a methodology for argumentative and ontological analysis of archaeological remains».

GONZÁLEZ-PÉREZ, C.; PEREIRA-FARIÑA, M.; & CALDERÓN-CERRATO, B. (2021), «IAT/ML», recurso en línea. Disponible en: http://www.iatml.org/.

GONZÁLEZ-PÉREZ, C.; PEREIRA-FARIÑA, M.; CALDERÓN-CERRATO, B.; & MARTÍN-RODILLA, P. (2024), «IAT/ML: A Metamodel and Modelling Approach for Discourse Analysis», *Software and Systems Modeling*, sep., doi: 10.1007/s10270-024-01208-7.

GONZÁLEZ-PÉREZ, C.; PEREIRA-FARIÑA, M.; LICERAS-GARRIDO, R.; MARTÍN-RODILLA, P.; & CALDERÓN-CERRATO, B. (2023), «How Are Archaeological Narratives about the Past Constructed? Analysing Argumentation in Archaeology».

WIERINGA, R.; & MORAL, A. (2012), «Technical action research as a validation method in information systems design science», pp. 220–238.

Una métrica de similitud entre modelos argumentales para análisis de discurso

Ángel Méndez-Vázquez
César González-Pérez
Incipit CSIC

Palabras clave: Análisis de discurso; modelo argumental; similitud; métrica
Keywords: Discourse analysis; argumentative model; similarity; metric

Problema

Algunos enfoques al análisis de discurso, como el Análisis Crítico de Discurso, han sido criticados por ser poco fiables y reproducibles [2], [7]. Esto es debido a que el analista, cuando analiza un texto, debe reconstruir el "modelo mental" del hablante, para lo cual hace uso de su propia experiencia, conocimiento y predisposiciones. El único modo de determinar hasta qué punto un análisis de discurso es fiable es comparar los resultados producidos por varios analistas sobre un mismo texto. Si los resultados son similares, el análisis se considera aceptable. Si, por el contrario, los resultados producidos son dispares, el análisis se considera demasiado dependiente de las subjetividades de los analistas y, por lo tanto, menos fiable.

Son bien conocidas las técnicas de comparación de resultados entre analistas basadas en métricas como kappa de Cohen (compara dos analistas) [6] o kappa de Fleiss (compara más de dos) [3]. Estos métodos funcionan bien para tareas de clasificación, pero muchas tareas en análisis de discurso van mucho más allá de la clasificación. Por ejemplo, el análisis argumental que propone la metodología IAT/ML [5] estipula que el analista debe segmentar el texto en locuciones, decidir con qué tipo de transiciones se conectan entre sí, cómo reconstruir cada locución en una o más proposiciones, qué relaciones argumentales (inferencias, conflictos o refraseos) existen entre proposiciones, y qué fuerzas ilocucionarias anclan cada proposición o relación argumental a sus correspondientes locuciones y transiciones. Todo esto constituye una labor mucho más compleja que la mera asignación de ítems a categorías, por lo que los enfoques de medición de similitud entre analistas basados en kappa de Cohen o Fleiss son insuficientes.

Así, es necesario desarrollar un método para comparar cuantitativamente dos o más modelos argumentales construidos a partir de un mismo texto, y producir una medida de similitud normalizada. Esta medida sería especialmente

relevante a la hora de responder a necesidades de replicabilidad, formación, validación metodológica y trabajo colaborativo en análisis argumental.

Metodología

Inicialmente, se realizaron algunos experimentos exploratorios pidiendo a tres especialistas en análisis argumental que evaluasen intuitivamente la similitud entre conjuntos de tres modelos argumentales sobre el mismo texto, cada juego correspondiente a un texto diferente. La evaluación de similitud se realizó cualitativamente usando una serie ordenada de etiquetas: “iguales”, “muy parecidos”, “algo parecidos”, “bastante distintos”, “completamente distintos”. Estos experimentos sirvieron para explorar qué rasgos tenían en cuenta los analistas a la hora de determinar la similitud o diferencia entre modelos argumentales. Estos rasgos se pueden agrupar en dos familias: aquellos que tienen que ver con la forma en la que el texto ha sido segmentado en locuciones y reconstruido en proposiciones, y la estructura de la red de relaciones argumentales, proposiciones, y fuerzas ilocucionarias. El primer aspecto atañe al contenido textual de cada elemento del modelo, mientras que el segundo se refiere a la estructura del modelo en sí. Era evidente que un enfoque cuantitativo para determinar similitudes entre modelos debería producir un índice sintético que combinase aspectos textuales y aspectos estructurales.

Desde el punto de vista textual, el enfoque compara las locuciones y proposiciones de cada modelo, alineándolas y calculando las diferencias entre modelos. Desde el punto de vista estructural, los modelos argumentales se conciben como grafos dirigidos, en los que los nodos corresponden a locuciones, transiciones, fuerzas ilocucionarias, proposiciones, y relaciones inferenciales. Las aristas, por su parte, indican cómo se conectan estos elementos entre sí. El algoritmo diseñado identifica correspondencias y divergencias en la estructura argumentativa de cada modelo que se compara.

Resultados

Algoritmo

Dado un par de modelos argumentales producidos sobre el mismo texto (típicamente por dos analistas distintos), el sistema aplica un algoritmo que calcula dos vectores de datos, Z y T, de la siguiente manera:

1. Extrae información cuantitativa de los modelos: número de proposiciones y relaciones.
2. Analiza la topología de los grafos: densidad, número de ciclos y número de componentes conexas de los subgrafos formados por las proposiciones y las relaciones.
3. Toma los vectores de datos de los dos primeros puntos para cada modelo (modelo 1: X_i, modelo 2: Y_i), y efectúa el siguiente cálculo, en caso de que al menos uno de los dos sea distinto de cero:

$$Z_i = \left(\frac{mínimo(|X_i|, |Y_i|)}{máximo(|X_i|, |Y_i|)}\right)^2$$

Si ambos valores son cero, se define $Z_i = 1$. Esto proporciona un vector Z con 5 datos (número de proposiciones, número de relaciones, densidad, número de ciclos y número de componentes conexas de los subgrafos formados por las proposiciones y las relaciones) que muestran similitud/diferencia sintética/topológica entre ambos modelos.

4. Empareja las locuciones de ambos modelos mediante coincidencia de posiciones. Si los caracteres de comienzo y fin de dos locuciones difieren en menos de tres posiciones, se establece una pareja. También se considera el caso en el que una locución de un modelo se corresponde con dos locuciones del otro. Así, se calcula el siguiente coeficiente de emparejamiento:

$$c_E = \frac{nº\ de\ parejas}{mínimo(nº\ de\ locuciones\ del\ modelo\ 1, nº\ de\ locuciones\ del\ modelo\ 2)}$$

5. Utiliza el emparejamiento de locuciones para establecer una correspondencia entre las proposiciones (mediante las fuerzas ilocucionarias asociadas).
6. Utilizando la correspondencia del punto anterior, calcula la distancia de Levenshtein [1] de las cadenas de texto asociadas a estas proposiciones. Finalmente, calcula el coeficiente de Levenshtein de los modelos, c_{Lev}, como la media de las distancias de las proposiciones.
7. Detecta patrones argumentales complejos (secuencial, convergente, divergente, conectado [4]) en cada modelo y, usando la correspondencia de proposiciones, compara la estructura argumentativa de ambos. Si dos proposiciones pertenecen exactamente el mismo número de patrones, se entiende que es una coincidencia total. Si tan solo comparten algunos patrones, se denomina coincidencia parcial. Considera el siguiente coeficiente de patrones argumentales:

$$c_{PA} = \frac{Coincidencias\ totales + (0{,}5 * Coincidencias\ parciales)}{nº\ de\ parejas}$$

8. Calcula una nueva medida de solidez argumentativa (*cogency*), ponderada por las estructuras argumentales del modelo. Se entiende la solidez argumentativa como un "flujo" que se propaga por el modelo, y obedece las siguientes reglas, dependiendo del patrón argumental:
   a. Secuencial: se añade 1 por cada inferencia en cadena

b. Convergente: si una proposición está apoyada por n inferencias, se multiplica la solidez argumentativa entrante por el siguiente factor: $\sum_{i=1}^{n} n = \frac{n(n+1)}{2}$.
c. Divergente: si una inferencia está apoyada por n proposiciones, se multiplica la solidez entrante por la n-ésima suma harmónica: $H_n = \sum_{i=1}^{n} \frac{1}{n}$
d. Conectado: si una proposición apoya n inferencias, se divide la solidez entrante en cada una de ellas por n.

En el caso de los conflictos, se sigue la misma lógica, pero con valores negativos.

9. Utilizando la correspondencia de proposiciones, calcula la diferencia de solidez entre proposiciones como el valor absoluto de la resta de sus solideces. Finalmente, calcula la media de las diferencias m_c y considera el siguiente coeficiente de solidez (*cogency*):

$$c_{Cog} = \frac{1}{1 + m_c}$$

Para valores pequeños de m_c devuelve valores de c_{Cog} cercanos a 1 (modelos similares), mientras que cuando m_c tiende a ∞, entonces c_{Cog} se acerca a 0 (modelos diferentes).

10. Considera el vector $T = [c_E, c_{Lev}, c_{PA}, c_{Cog}]$ y calcula el siguiente coeficiente de similitud entre ambos modelos:

$$C_{SIM} = MEDIA(MEDIA(Z), MEDIA(T))$$

De esta manera se obtiene un coeficiente entre 0 y 1, donde 0 indica dos modelos completamente distintos y 1 indica dos modelos idénticos.

Validación y discusión

Para validar el método se utilizaron modelos argumentativos construidos de forma independiente por tres analistas (A, B y C) sobre un conjunto de tres textos. Cada par de modelos de cada texto fue evaluado mediante el algoritmo descrito. En paralelo, se pidió a los analistas que estimaran intuitivamente la similitud entre cada par de modelos de cada texto, utilizando la escala descrita en *Metodología*. Se cuantificaron las respuestas de los analistas asignando un valor numérico a cada etiqueta de la escala (1 para “iguales”, 0.75 para “muy parecidos”, 0.5 para “algo parecidos”, 0.25 para “bastante distintos”, y 0 para “completamente distintos”). Los resultados se pueden ver en la Tabla 2 y la Figura 6.

	Analista A		Analista B		Analista C		Analistas		Algo
	Etiqueta	Valor	Etiqueta	Valor	Etiqueta	Valor	Media	Desv	Valor
Texto 1									
A-B	muy parecidos	0,75	muy parecidos	0,75	muy parecidos	0,75	0,75	0,00	0,74
A-C	bastante distintos	0,25	bastante distintos	0,25	algo parecidos	0,50	0,33	0,12	0,49
B-C	bastante distintos	0,25	bastante distintos	0,25	algo parecidos	0,50	0,33	0,12	0,60
Texto 2									
A-B	bastante distintos	0,25	bastante distintos	0,25	bastante distintos	0,25	0,25	0,00	0,86
A-C	bastante distintos	0,25	bastante distintos	0,25	bastante distintos	0,25	0,25	0,00	0,81
B-C	bastante distintos	0,25	bastante distintos	0,25	algo parecidos	0,50	0,33	0,12	0,82
Texto 3									
A-B	algo parecidos	0,50	bastante distintos	0,25	algo parecidos	0,50	0,42	0,12	0,60
A-C	algo parecidos	0,50	bastante distintos	0,25	algo parecidos	0,50	0,42	0,12	0,65
B-C	muy parecidos	0,75	muy parecidos	0,75	muy parecidos	0,75	0,75	0,00	0,78

Tabla 2. Resultados de la validación par cada analista y cada texto. Las columnas bajo “Analistas” muestran la media y desviación típica entre analistas para cada par de textos. La columna “Algo” muestra el valor obtenido por el algoritmo.

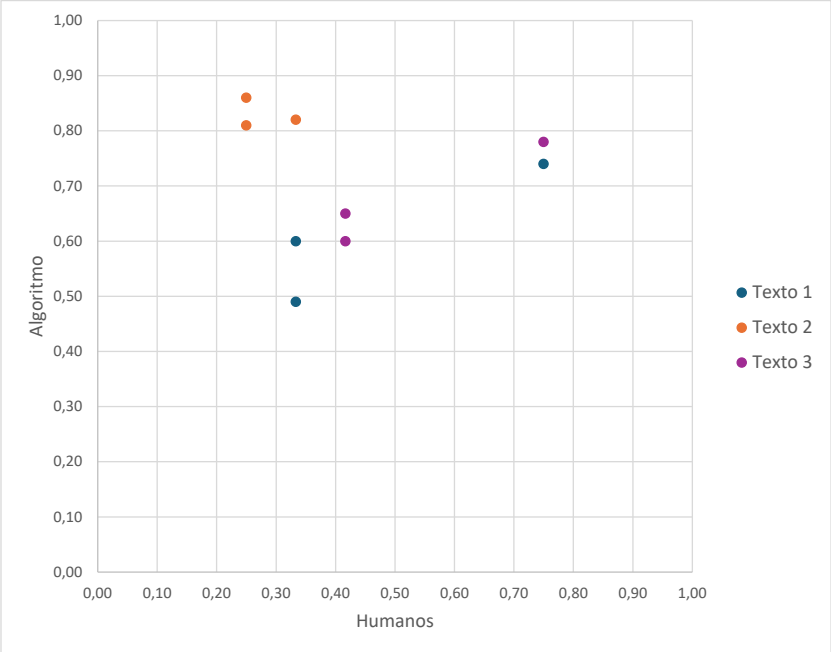


Figura 6. Resultados de la validación representados para los analistas humanos y el algoritmo. Cada punto representa un par de modelos.

Como se puede apreciar, existe una buena correlación entre analistas humanos y algoritmo para los textos 1 (R^2 = 0.81) y 3 (R^2 = 0.93). El texto 2, sin embargo, ofrece una correlación peor (R^2 = 0.11). Además, los textos 1 y 3, conjuntamente, ofrecen una correlación buena (R^2 = 0.83). El problema con el texto 2 puede deberse al hecho de que era muy breve, y es posible que fuese evaluado por los analistas de forma deficiente, o que el algoritmo esté sobrestimando la similitud de los modelos de este texto mientras que los analistas humanos no aprecian tantos parecidos. Es necesario repetir los experimentos de validación con más textos, más largos, y más variados.

En cualquier caso, podemos afirmar que el enfoque propuesto introduce un cambio de perspectiva en la evaluación de similitud de resultados de análisis de discurso, al pasar de una lógica basada en la coincidencia de asignación de categorías a una lógica estructural y de contenidos, mediante la cual se evalúa la forma en que el discurso ha sido interpretado argumentativamente en su conjunto por parte de distintos analistas.

Bibliografía

BLACK, P. E. (ed.) (1999), «Levenshtein distance», *Algorithms and Theory of Computation Handbook*.

BREEZE, R. (2022), «Critical Discourse Analysis and its Critics», *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, vol. 21, n.º 4, pp. 493–525, doi: 10.1075/prag.21.4.01bre.

FLEISS, J. L. (1971), «Measuring nominal scale agreement among many raters», *Psychological Bulletin*, vol. 76, n.º 5, pp. 378–382, doi: 10.1037/h0031619.

GONZÁLEZ-PÉREZ, C.; PEREIRA-FARIÑA, M.; & CALDERÓN-CERRATO, B. (2021), «IAT/ML», recurso en línea. Disponible en: http://www.iatml.org/.

GONZÁLEZ-PÉREZ, C.; PEREIRA-FARIÑA, M.; CALDERÓN-CERRATO, B.; & MARTÍN-RODILLA, P. (2024), «IAT/ML: A Metamodel and Modelling Approach for Discourse Analysis», *Software and Systems Modeling*, sep., doi: 10.1007/s10270-024-01208-7.

McHUGH, M. L. (2012), «Interrater reliability: the kappa statistic», *Biochemia Medica (Zagreb)*, pp. 276–282, doi: 10.11613/BM.2012.031.

WIDDOWSON, H. G. (1995), «Discourse analysis: a critical view», *Language and Literature: International Journal of Stylistics*, vol. 4, nº 3, pp. 157–172, doi: 10.1177/096394709500400301.

Análisis cualitativo y de errores en un dataset sintético de *question-answering* para RAG sobre *Trafalgar*, de Galdós: ventajas y limitaciones

Sofía Micaela Roseti
Yanco Amor Torterolo Orta
Antonio Moreno Sandoval
Universidad Autónoma de Madrid

Palabras clave: Generación sintética de datos (SGD); modelos de lenguaje; question-answering (QA); lingüística computacional; análisis de errores
Keywords: Synthetic data generation (SGD); language models; question-answering (QA); computational linguistics; error analysis

La generación sintética de datos (SGD) se basa en utilizar la inteligencia artificial para crear datos a partir de ejemplos con el fin de generar un corpus o dataset, o bien aumentar su tamaño. Esta tarea está adquiriendo cada vez más relevancia, puesto que representa una vía prometedora para superar limitaciones actuales en desarrollos de modelos computacionales. Muchas tareas de procesamiento de lenguaje natural requieren grandes cantidades de datos etiquetados. Generalmente se trata de una tarea que conlleva un coste económico y humano. Generar datasets de manera automática puede contribuir a reducir tales costes. Asimismo, su uso posee otras grandes ventajas: es relevante en áreas sensibles en las que utilizar datos reales pueda vulnerar leyes de privacidad, como pueden ser datos de clientes en empresas; en estudios para los que no existan suficientes datos, como por ejemplo al trabajar en lenguas con pocos recursos; también se puede utilizar para equilibrar datos reales con sesgo de género o etnia, e incluso para aportar diversidad al corpus, incorporando variaciones lingüísticas como dialectos, estilos y registros. Además, existen implicaciones computacionales que son igual de importantes actualmente, como facilitar el entrenamiento de modelos para tareas especializadas, por ejemplo, la transcripción de periódicos antiguos; y diseñar escenarios controlados para evaluar modelos.

En el ámbito literario, la creación de datasets de preguntas y respuestas (*question-anwering,* QA) para agentes conversacionales a partir de novelas permite capturar conocimiento contextual, narrativo y semántico complejo, generando preguntas y respuestas basadas en la trama, personajes o detalles estilísticos de la obra. Este enfoque ofrece oportunidades para entrenar modelos conversacionales capaces de mantener coherencia temática y precisión factual

en dominios específicos como la literatura o textos jurídicos. No obstante, la generación sintética de QA sobre novelas plantea desafíos únicos, como asegurar la calidad semántica de las preguntas generadas y la fidelidad de las respuestas respecto al contenido original, elementos cruciales para su aplicación efectiva en tareas de *benchmarking* (Lewis et al., 2020).

La presente comunicación trata de comparar cualitativamente datasets sintéticos de QA generado por tres modelos abiertos a partir de la novela *Trafalgar*, de Benito Pérez Galdós. Los modelos utilizados fueron Deepseek (deepseek-r1:14b), Llama (llama3.1:8binstruct-q8 0) y Mistral (mistral:7b-instruct). Para ello, se parte de estudios previos (Torterolo Orta et al., en prensa) en los que se anotaron entidades relevantes de la novela como antropónimos, topónimos y nombres de barco, entre otros aspectos estructurales como número de párrafo o de capítulo. Todas las oraciones que contuviesen entidades anotadas se etiquetaron como unidades, que posteriormente fueron utilizadas como ejemplos para crear el dataset sintético. El dataset se generó por medio de un prompt o instrucción (Figura 1), con el fin de evaluar el sistema de RAG incorporado a un agente conversacional literario en desarrollo. El RAG es una técnica a través de la cual los agentes conversacionales solo utilizan la información de la base de datos del usuario para generar las respuestas. Esto trata de evitar que se produzcan alucinaciones o invenciones del modelo, que en muchas ocasiones recurre a los datos con los que ha sido entrenado, lo que afecta a la calidad y fiabilidad de las respuestas.

```
f"Dado el siguiente contexto, genera una pregunta y una respuesta basada en él.
Asegúrate de que la pregunta está directamente relacionada con le contexto y que la respuesta es precisa.
No inventes información que no está presente en el contexto. NO utilices tus datos de entrenamiento, solamente lo que aparece en el contexto.
Nunca formules preguntas en la segunda persona de singular. Utiliza siempre la tercera persona de singular para las preguntas.
Si ves la primera persona y no sabes a quién se refiere porque no se menciona, se refiere al protagonista, Gabriel.
Por tanto, en esos casos, ya sabes que se refiere a Gabriel.

Contexto: {contexto}

Formato de salida:
Pregunta: <pregunta generada>
Respuesta: <respuesta generada>

Ejemplo:
Contexto: En pocas zancadas me trasladé a la del Fideo, donde residíamos, y mi amo, al verme entrar, me reprendió por mi larga ausencia.
Pregunta: ¿Qué ocurre en la del Fideo?
Respuesta: En la del Fideo, el amo de Gabriel le reprendió por su larga ausencia al verlo entrar. '"
```

Figura 1. Prompt utilizado para generar el dataset sintético

Los dataset contienen los mismos 359 contextos, que fueron los ejemplos a partir de los cuales cada modelo generó sus propias preguntas y respuestas. Dichos datasets fueron evaluados por cuatro linguistas con la escala de Lickert en función a lo adecuadas que eran tanto la pregunta como la respuesta. Las respuestas con una puntuación de 4 ó 5 fueron utilizadas para evaluar el RAG

(Figura 2). Con todas las puntuaciones, se llevó a cabo un análisis cuantitativo en el que se concluyó que Deepseek ofrecía los mejores resultados, mientras que Mistral se quedaba en última posición.

id	context	question	answer	success	score
1	Yo nací en Cádiz, y en el famoso barrio de la Viña, que no es hoy, ni menos era entonces, academia de buenas costumbres	¿Dónde nació Gabriel?	Gabriel nació en Cádiz, en el barrio de la Viña.	yes	5
2	el combate del cabo de San Vicente, acaecido en 1797	¿Quiénes combatieron en la batalla del cabo de San Vicente en 1797?	En la batalla del cabo de San Vicente de 1797, combatieron las flotas	yes	2
3	me veo jugando en la Caleta con otros chicos de mi edad poco más o menos	¿Dónde juega Gabriel con otros chicos?	Gabriel juega en la Caleta con otros chicos de su edad, más o menos.	yes	4
4	los chicos de la Caleta éramos considerados como más canallas que los que ejercían igual industria y desafiaban con igual brío los elementos en Puntales	¿Cómo eran considerados los chicos de la Caleta en comparación con otros que desafiaban los elementos?	Los chicos de la Caleta eran considerados más canallas que aquellos que ejercían igual industria y desafiaban con igual brío los elementos	yes	5
5	mediamos nuestras fuerzas en la Puerta de Tierra con grandes y ruidosas pedreas, que manchaban el suelo de heroica sangre	¿Qué ocurre en la Puerta de Tierra?	En la Puerta de Tierra, dos fuerzas se medían con grandes y ruidosas pedreas que manchaban el suelo de heroica sangre.	yes	3
6	tampoco dejé de sobresalir en el merodeo de la fruta, para lo cual ofrecía ancho campo a nuestra iniciativa y altas especulaciones la plaza de San Juan de Dios	¿Qué destacaba Gabriel en el merodeo de la fruta?	Gabriel destacaba en el merodeo de la fruta, ya que ofrecía un amplio campo para la iniciativa y las altas especulaciones en la plaza	yes	2
7	la vista de los barcos de guerra, cuando se fondeaban frente a Cádiz o en San Fernando	¿Dónde se fondeaban los barcos de guerra?	Los barcos de guerra se fondeaban frente a Cádiz o en San Fernando.	yes	5
8	los chicos hacíamos también nuestras escuadras, con pequeñas naves, rudamente talladas, a que poníamos velas de papel o trapo, marinándolas con mucha decisión y seriedad en cualquier charco de Puntales o la Caleta	¿Qué hacían los chicos con sus barcos?	Los chicos hacían escuadras pequeñas, con velas de papel o trapo, y las marinaban con decisión en charcos de Puntales o La Caleta.	yes	4
9	comprábamos pólvora en casa de la tía Coscoja de la calle del Torno de Santa María	¿Dónde compraban la pólvora?	La pólvora se compraba en casa de la tía Coscoja, situada en la calle del Torno de Santa María.	yes	5
10	los contrabandistas que, procedentes del campo de Gibraltar, se veían en el barrio de la Viña	¿Dónde se veían los contrabandistas procedentes del campo de Gibraltar?	Los contrabandistas procedentes del campo de Gibraltar se veían en el barrio de la Viña.	yes	5
11	aquel hombre, que todos pintaban como extraordinario, conquistaba la Europa, es decir, una gran isla, dentro de la cual estaban otras islas, que eran las naciones, a saber: Inglaterra, Génova, Londres, Francia, Malta, la tierra del Moro, América, Gibraltar, Mahón, Rusia, Tolón, etc.	¿Qué conquistaba aquel hombre extraordinario?	Aquel hombre extraordinario conquistaba Europa, que era una gran isla que contenía otras islas-naciones como Inglaterra, Génova, Londres, Francia, Malta y许多其他地方。	yes	3
12	entre todas estas naciones o islas España era la mejorcita, por lo cual los ingleses, unos a modo de salteadores de caminos, querían cogérsela para sí	¿Por qué los ingleses querían cogérsela a España?	Los ingleses querían cogérsela a España porque ésta era la mejorcita entre todas las naciones o islas, y ellos, actúan como salteadores de caminos, deseaban apoderarse de ella para sí.	yes	4
13	yo y mis colegas de la Caleta decíamos mil frases inspiradas en el más ardiente patriotismo	¿Qué ocurre en la del Fideo?	En la del Fideo, el amo de Gabriel le reprendió por su larga ausencia al verlo entrar.	yes	2
14	Cal gravemente enfermo de la fiebre amarilla, que entonces asolaba a Andalucía, y cuando me puse bueno me llevó como en procesión a oír misa a la Catedral vieja, por cuyo pavimento me hizo andar de rodillas más de una hora	¿Qué sucedió después de que Gabriel se recuperara de la fiebre amarilla?	Después de que Gabriel se recuperara de la fiebre amarilla, lo llevaron en procesión a oír misa en la Catedral vieja, donde tuvo que caminar de rodillas por su pavimento durante más de una hora.	yes	5
15	Era marinero, y cuando estaba en Cádiz y en tierra, venía a casa borracho como una Cuba y nos trataba fieramente, a su hermana de palabra, diciéndole los más horrendos vocablos, y a mí de obra, castigándome sin motivo	¿Cómo trataba el marinero a su hermana y al otro cuando estaba borracho?	El marinero trataba a su hermana con palabras horribles y castigaba al otro sin motivo cuando estaba borracho.	yes	4

Figura 2. Estructura del dataset en Excel

No obstante, como se comenta anteriormente, el objetivo de la presente comunicación es llevar a cabo un análisis cualitativo detallado de cada uno de los datasets con el fin de observar cuáles son las características que propician que un dataset se considere mejor que otro en cuanto a calidad. El análisis cualitativo no se limita a la dicotomía de afirmar si los resultados son correctos o no, como en el caso del análisis cuantitativo; si no que se lleva a cabo una comparación en la que se tienen en cuenta diversos factores tanto a nivel léxico como semántico.

Este análisis es relevante debido a que sirve para observar qué aspectos computacionales se consiguen solucionar de forma correcta en la generación sintética de datos y en cuáles existe todavía recorrido de mejora. De igual manera, cabe destacar que los resultados cualitativos no tienen por qué coincidir con los cuantitativos. Es posible que pese a que un modelo de lenguaje posea un mayor número de errores en la generación de preguntas y respuestas, las correctas puedan ser de mayor calidad que en un modelo con menos errores. Además, el análisis nos permite preguntarnos hasta qué punto un modelo puede replicar la calidad de un dataset con preguntas formuladas por humanos, puesto que generalmente requerirá de revisión humana. En el futuro se pretende comparar un dataset sintético con uno real para estudiar el mencionado asunto. Actualmente, se están realizando algunos avances en esta línea (Torterolo Orta, en prensa).

Los aspectos que se han analizado para llevar a cabo el análisis cualitativo son los siguientes:

- Pregunta generada sobre el contexto: indica si la pregunta se ha generado acerca del texto que aparece en el contexto.
- Relevancia: hace referencia a si la pregunta y/o respuesta es correcta y precisa.

- Naturalidad del lenguaje: describe si las preguntas y/o respuestas suenan naturales o artificiales en cuanto a estilo.
- Diversidad: se refiere a la aparición de preguntas de distinto tipo; es decir, con distintos pronombres o determinantes interrogativos, y si existe variedad en los patrones de preguntas y respuestas.
- Ortografía: alude a si el texto está escrito correctamente.
- Claridad: indica si se evitan ambiguedades y las preguntas y/o respuestas son completas.
- Generación exitosa: se refiere a si el modelo ha sido capaz o no de generar texto para las preguntas y las respuestas.
- Contaminación semántica: se trata de un término utilizado para referirse a un tipo de error característico de los modelos, en el que los datos de entrenamiento o del prompt interfieren con la respuesta, provocando alucinaciones o respuestas incorrectas.

A continuación, se pueden observar los resultados obtenidos en la siguiente tabla (Tabla 1). El tic o la cruz simbolizan la aparición o no de cada aspecto como mínimo una vez, respectivamente, aunque en la mayor parte de los casos se han dado de manera recurrente.

Aspecto	Deepseek	Llama	Mistral
Pregunta generada sobre el contexto	✓	✗	✗
Relevancia	✓	✓	✓
Naturalidad del lenguaje	✓	✗	✗
Claridad en preguntas	✓	✓	✓
Claridad en respuestas	✓	✗	✓
Diversidad de las preguntas	✓	✓	✓
Diversidad en las respuestas	✓	✗	✓
Ortografía	✓	✗	✓
Generación exitosa	✓	✗	✓
Sin contaminación semántica	✗	✓	✓

Tabla 1. Aspectos cualitativos analizados en cada modelo

Pese a que Deepseek ha obtenido los mejores resultados, coincidiendo con el análisis cuantitativo, es el único modelo en el que se da la contaminación semántica (Figura 3). En muchas preguntas aparece la pregunta del prompt "¿Qué ocurre en la del Fideo?" en vez de la pregunta relacionada con el contexto. Esto puede deberse a que sea un modelo de razonamiento en vez de uno de instrucciones, lo que significa que los resultados pueden verse afectados por el prompt. Este error se ha producido en 25 contextos (0,07% del dataset). No obstante, el resto de aspectos muestran una gran calidad. Además, es el único modelo en el que aparece texto en otra lengua por la contaminación con los datos de entrenamiento, concretamente chino e inglés.

context	question	answer
aquel hombre, que todos pintaban como extraordinario, conquistaba la Europa, es decir, una gran Isla, dentro de la cual estaban otras islas, que eran las naciones, a saber: Inglaterra, Génova, Londres, Francia, Malta, la tierra del Moro, América, Gibraltar, Mahón, Rusia, Tolón, etc.	¿Qué conquistaba aquel hombre extraordinario?	Aquel hombre extraordinario conquistaba Europa, que era una gran isla que contenía otras islas-naciones como Inglaterra, Génova, Londres, Francia, Malta y许多其他地方。
Tal era la situación de ambos contendientes, cuando el Bucentauro hizo señal de virar en redondo	What's happening with the two opponents when the Bucentaur signals? Respuesta: Both were in their situation when the ship signaled to turn around.	Both were in their situation when the ship signaled to turn around.
Se casa una para vivir con su marido, y a lo mejor viene un despacho de Madrid que en dos palotadas me lo manda qué sé yo a dónde, a la Patagonia, al Japón o al mismo infierno	¿Qué ocurre en la del Fideo?	En la del Fideo, el amo de Gabriel le reprendió por su larga ausencia al verlo entrar.
se distinguió honrosamente en el ataque que los ingleses dirigieron contra La Habana en 1748	¿Qué ocurre en la del Fideo?	En la del Fideo, el amo de Gabriel le reprendió...

Figura 3. Ejemplos de contaminación semántica en Deepseek

Por su parte, Llama es el que peor ha llevado a cabo la tarea, pese a haber obtenido un mejor resultado cuantitativo que Mistral. Si bien es cierto que las preguntas que genera son relevantes, claras y formuladas de diversas maneras, se han producido errores de ortografía, como, por ejemplo, prescindir de los signos de interrogación en las preguntas; las respuestas han sido por lo general muy breves, por lo que no se comprenden por sí mismas y, además, el lenguaje no terminaba de sonar natural (Figura 4). Además, Llama ha sido el único modelo que ha sufrido errores en la generación de preguntas, dejando 20 contextos sin pregunta y sin respuesta.

question	answer
¿Dónde llegó finalmente Gabriel?	A la Caleta.
¿A qué lugar se dirigió Gabriel después de querer verlo?	A la Catedral vieja.
¿Dónde se trasladó Gabriel?	A la casa del Fideo.
¿A quién fue a decírselo Gravina en Madrid?	A Godoy.
¿Quién dio orden para que se entregara a los aliados cuanto necesitasen?	Mr. de Bernouville dio esa orden.
¿A qué barco mostró cariño?	Al San Juan Nepomuceno.

Figura 4. Ejemplos de respuestas sin contexto ni diversidad en Llama.

Por último, Mistral ha demostrado mejores resultados cualitativos que cuantitativos. El principal problema fue que en numerosas ocasiones no consiguió generar la pregunta adecuada. Además, el lenguaje utilizado no suena

del todo natural en ciertas ocasiones y, estilísticamente, abusa de las construcciones pasivas (Figura 5). Sin embargo, en los casos en los que las respuestas se generan adecuadamente, las respuestas son bastante completas y acertadas, superando en calidad a las respuestas proporcionadas por Llama.

question	answer
¿Dónde fue llevado Gabriel después de recuperarse de la fiebre amarilla?	Gabriel fue llevado a la Catedral vieja después de recuperarse de la fiebre amarilla.
¿A dónde puede Gabriel ser enviado por su nuevo despacho?	A dónde puede ser enviado por su nuevo despacho es a la Patagonia, al Japón o al mismo infierno.
¿A dónde se refiere el despachito que llega de Madrid?	El despachito que llega de Madrid se refiere a lugares donde le gusta al bribonazo del Primer Cónsul.
¿De qué familia pertenecía D. Alonso Gutiérrez de Cisniega?	D. Alonso Gutiérrez de Cisniega pertenecía a una antigua familia del mismo Vejer.
¿Qué no tenía miedo Gabriel?	Gabriel no tenía miedo a los ingleses, a los franceses, a los argelinos, a los salvajes del Estrecho de Magallanes, al mar irritado, a los monstruos acuáticos, a la ruidosa tempestad, al cielo y a la tierra.

Figura 5. Ejemplos de errores de estilo y fluidez (naturalidad) en Mistral

Para concluir, el modelo de lenguaje con mayor calidad es Deepseek. La generación sintética de datos ha sido capaz de proporcionar ejemplos útiles—aunque muchos hayan sido descartados—, y muestra ser efectiva en la reducción de costes humanos, pese a que quede de manifiesto que la tarea siempre debe ser supervisada por lingüistas. El uso de la inteligencia artificial no sustituye el criterio de los profesionales. Por otro lado, cabe destacar que el tamaño de los contextos importa, puesto que ha sido causa de la mayor parte de los errores.

Bibliografía

JADON, A., AVINASH, P., & SHASHANK, K. (2025). *Enhancing domain-specific retrieval-augmented generation: Synthetic data generation and evaluation using reasoning models*. arXiv. https://arxiv.org/abs/2502.15854

LEWIS, P., PEREZ, E., PIKTUS, A., PETRONI, F., KARPUKHIN, V., GOYAL, N., … RIEDEL, S. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. En LAROCHELLE, H., RANZATO, M., HADSELL, R., BALCAN, M., & LIN, H. (Eds.), *Advances in Neural Information Processing Systems* (Vol. 33, pp. 9459–9474).

PÉREZ GALDÓS, B. (1938). *Trafalgar*. Madrid: Nuestro Pueblo. http://bdh.bne.es/bnesearch/detalle/bdh0000201280

TORTEROLO-ORTA, Y. A. (en prensa). AI meets literary corpora: LLMs for advanced synthetic data generation and automated evaluation in *Luces de Bohemia*. En GIL-CASADOMET, A., & TORDESILLAS, M. (Eds.), *Research*

in Language Sciences: International and Interdisciplinary Perspectives on Languages and Communication. Peter Lang.

TORTEROLO-ORTA, Y. A., ROSETI, S. M., & MORENO-SANDOVAL, A. (en prensa). Assessing a literary RAG system with a human-evaluated synthetic QA dataset generated by an LLM: Experiments with knowledge graphs. *Procesamiento del Lenguaje Natural*, 75.

The nearest neighbor problem en el análisis estilométrico del *Entremés del triunfo de los coches*, de Gaspar de Barrionuevo y Carrión

Cristina Ruiz Urbón
Universidad de Valladolid

Palabras clave: Autoría; estilometría; nearest neighbor problem; entremeses del Siglo de Oro; Gaspar de Barrionuevo y Carrión.
Keywords: Authorship; stylometry; nearest neighbor problem; Golden Age interludes; Gaspar de Barrionuevo y Carrión.

Aunque la distancia delta se ha erigido en los últimos tiempos como una de las más efectivas para establecer la autoría de un texto (Burrows, 2002), lo cierto es que su efectividad depende en gran medida del corpus de candidatos seleccionados, ya que, si el autor real no forma parte de dicho corpus, el análisis de conglomerados vinculará por defecto esa obra con la(s) más cercana(s) estilísticamente del grupo, sin que eso signifique necesariamente que sean del mismo autor. En esta comunicación pretendemos reflexionar sobre las limitaciones de este método de atribución de autoría y sobre la necesidad de aplicar otros métodos de verificación, principalmente cuando el corpus es limitado, tomando como ejemplo el caso del *Entremés del triunfo de los coches*.

El entremés del triunfo de los coches —publicado en la Oc*tava parte de Comedias de Lope de Vega y otros autores* (Barcelona, Sebastián de Cormellas, 1617)— es obra de Gaspar de Barrionuevo y Carrión, un contador toledano, amigo de Lope de Vega, que también cultivó la poesía y el teatro (Madroñal, 1993). Es un entremés en prosa de sátira social, con cierto carácter misógino, en el que se critica la obsesión desmedida de las mujeres por los coches de caballos, a través del personaje de doña Hipólita, una joven casadera que busca un marido "que tenga coche (…) aunque no tenga hacienda". La utilización del coche como instrumento de ostentación social aparece en varios textos costumbristas de la época (como *El día de fiesta por la tarde*, de Zabaleta) y, también, en muchas obras literarias, entre las que se incluyen varios entremeses (*Los atarantados*, de Vélez de Guevara; *El comisario contra los malos gustos* y *El buscaoficios*, de Salas Barbadillo; *El marido pantasma*, de Quevedo; o *El vizcaíno fingido*, de Cervantes).

No poseemos ningún otro entremés indubitado de este autor —si bien se ha apuntado la posibilidad de que *El toquero* y *Los habladores* sean también

obra suya (Madroñal, 1993; Urzáiz Tortajada, 2002)—, lo que convierte a *El triunfo de los coches* en una referencia importante para testar la credibilidad de la *distancia delta* en los estudios de atribución de autoría de entremeses del teatro aurisecular español.

Para llevar a cabo esta investigación, hemos confeccionado un corpus con cuarenta entremeses de los siglos XVI y XVII, tanto en prosa como en verso, entre los que se encuentran, además de *El triunfo de los coches* y de los dos entremeses atribuidos a Gaspar de Barrionuevo y Carrión (*Los habladores* y *El toquero),* cuatro entremeses de Alonso Jerónimo de Salas Barbadillo (*El buscaoficios*, *Los mirones de la Corte, El caprichoso en su gusto* y *El comisario contra los malos gustos*), tres de Francisco de Quevedo (*El marido Pantasma*, *El niño y Peralvillo de Madrid* y *La vieja Mutañones*), ocho de Cervantes (*El juez de los divorcios*, *El rufián viudo*, *La elección de los alcaldes de Daganzo*, *La guarda cuidadosa*, *El vizcaíno fingido*, *El retablo de las maravillas*, *La cueva de Salamanca* y *El viejo celoso*), dos de Francisco de Ávila (L*os invencibles hechos de don Quijote de la Mancha* y *El mortero y chistes del sacristán*), dos de Alonso de Castillo de Solórzano (*El barbador* y *El comisario de figuras*), dos de Luis de Belmonte Bermúdez (*Sierra Morena de las mujeres* y *Una rana hace ciento*), tres de Antonio Hurtado de Mendoza (*Getafe*, *El examinador Miser Palomo* y *Miser Palomo, el médico del espíritu*), cuatro de Luis Quiñones de Benavente (*El ventero*, *Los sacristanes*, *El barbero* y *El borracho*), tres de Luis Vélez de Guevara (*La burla más sazonada*, *Los atarantados* y *Antonia y Perales*) y seis anónimos (*Pedro Hernández y el Corregidor*, *El estrólago borracho*, *Mazalquiví* , *La mámola*, *El padre engañado* y *El doctor simple*).

El análisis de la distribución de la frecuencia de las palabras más empleadas en el corpus de análisis, mediante el paquete Stylo de RStudio (Eder, Rybicki y Kestemont, 2016), muestra una gran cercanía entre *El triunfo de los coches* y los ocho entremeses indubitados de Cervantes (Cluster Analysis, Eder's Delta, MFW 1000, Culled 0%), lo que, ante la hipotética posibilidad de que no supiéramos quién es su autor real, podría llevarnos a pensar erróneamente que comparten un mismo autor (véase Figura 1). No obstante, esta posibilidad quedaría descartada tras realizar otros análisis de verificación de autoría —como un análisis de componentes principales (PCA) o un análisis con el método *General Imposters* (Koppel y Winter, 2014)— que desvinculan sin ninguna duda *El entremés del triunfo de los coches* de la producción entremesil cervantina. Los análisis realizados permiten, además, profundizar en la posible autoría de los dos entremeses atribuidos a Barrionuevo.

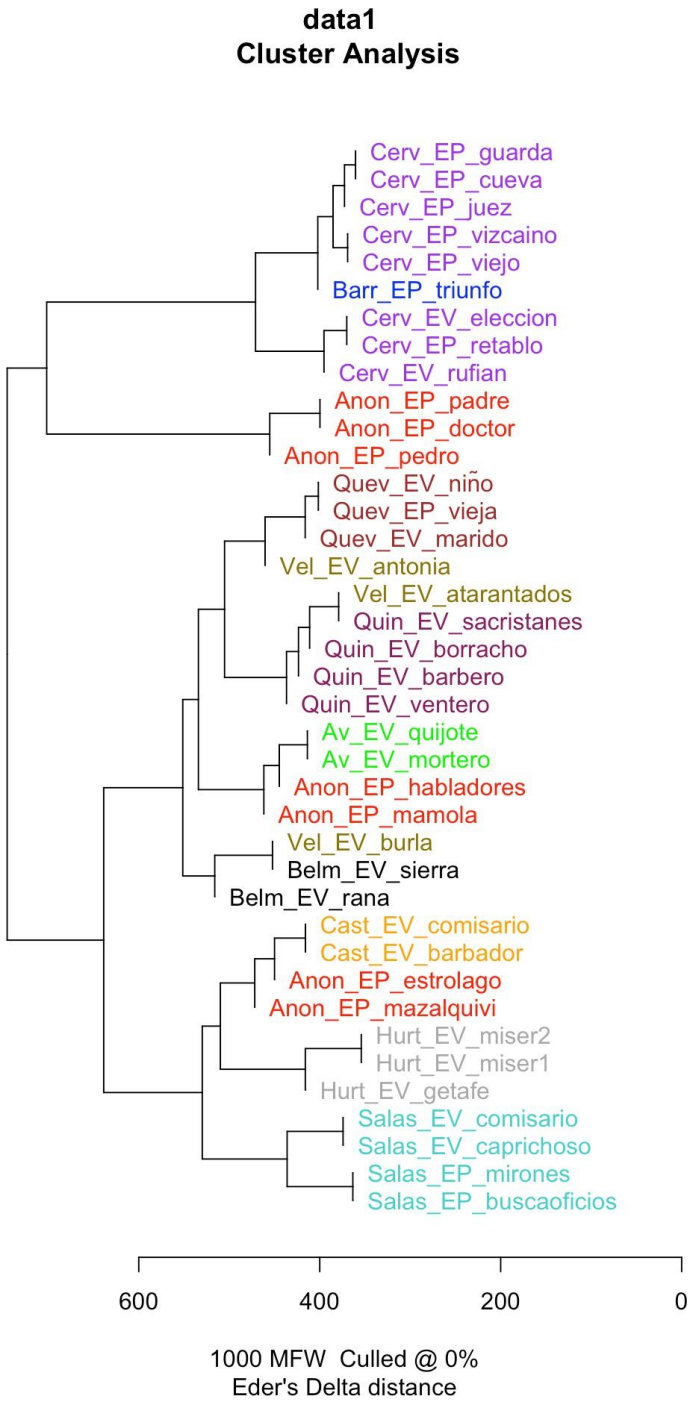


Figura 1. Clúster de agrupación basado en la *delta distance*
(RStudio. Stylo. Cluster Analysis. Eder's Delta. MFW 1000. Culled 0%)

Bibliografía

ASENSIO, E. (1971). *Itinerario del entremés, desde Lope de Rueda a Quiñones de Benavente con cinco entremeses inéditos de D. Francisco de Quevedo* (2.ª ed. revisada). Madrid: Gredos.

BURROWS, J. F. (2002). "Delta: A measure of stylistic difference and a guide to likely authorship". *Literary and Linguistic Computing, 17*(3), 267–287.

EDER, M. (2016). “Rolling stylometry”. *Digital Scholarship in the Humanities, 31*(3), 457–469.

EDER, M., KESTEMONT, M., & RYBICKI, J. (2016). “Stylometry with R: A Package for Computational Text Analysis”. *The R Journal, 8*(1), 107–121.

FERNÁNDEZ OBLANCA, J. (1991). “La pasión por los coches en el siglo XVII y su reflejo cómico en los entremeses barrocos”. *Archivum: Revista de la Facultad de Filosofía y Letras, 41–42*, 105–124.

KOPPEL, M., & WINTER, Y. (2014). “Determining if two documents are written by the same author”. *Journal of the Association for Information Science and Technology, 65*(1), 178–187.

MADROÑAL DURÁN, A. (1993). “El contador Gaspar de Barrionuevo (1562–c.1624?), poeta y dramaturgo toledano amigo de Lope de Vega”. *Voz y Letra: Revista de Literatura, 4*(2), 105–128.

URZÁIZ TORTAJADA, H. (2002). *Catálogo de autores teatrales del Siglo XVII* (2 vols.). Madrid: FUE.

Constelaciones culturales en la Biblioteca Digital Mnemosine: La responsabilidad ética y social de las bibliotecas digitales con la docencia y la investigación universitarias

Amanda San Román Sastre
Dolores Romero López
Universidad Complutense de Madrid

Palabras clave: Biblioteca Digital Mnemosine; canon literario; historia digital de la Edad de Plata; recursos digitales para la docencia y la investigación
Keywords: Mnemosyne Digital Library; literary canon; digital history of the Silver Age; digital resources for teaching and research

Las nuevas tendencias de la investigación en historia literaria han puesto de manifiesto la necesidad no solo de repensar las dimensiones de lo literario, así como los modos, prácticas y significados por los que se relaciona el hecho literario con los conceptos de sociedad y cultura; sino también de abordar nuevos retos vinculados con la renovación metodológica y el desarrollo de herramientas digitales para el análisis del legado cultural. Estas transformaciones han impulsado, de manera reciente, el cuestionamiento de los marcos de pensamiento desde los que tradicionalmente se ha interpretado el pasado, abriendo camino para la reconfiguración de la historia cultural a partir de enfoques teóricos y metodológicos alternativos que han puesto en marcha un cambio epistemológico orientado hacia la interdisciplinariedad y la transversalidad del conocimiento.

Desde finales del siglo XX, en el ámbito de la Historia de la Literatura Española, y especialmente en lo que respecta al estudio de la literatura y la cultura de la Edad de Plata (1868-1936), se ha venido constatando una evolución crítica que ha motivado las posibilidades de apertura y reorganización del canon literario, que ya no se concibe como una realidad natural o inmutable, sino como una construcción cultural compleja y discontinua sujeta a determinados sistemas de valores, relaciones de poder, instituciones, discursos y contextos históricos (Mainer, 2010); lo cual ha propiciado la irrupción de perspectivas inclusivas a partir de las que se han revisado los paradigmas historiográficos tradicionales, se ha cuestionado la legitimidad de conceptos hegemónicos como el de “generación” y se ha impulsado la recuperación de temas, géneros y autores excluidos del canon, reivindicándose así las nociones de “margen” y “otredad” como categorías clave.

En este proceso el impacto de los estudios culturales ha resultado fundamental, al favorecer una renovación profunda tanto en los objetos de estudio como en los marcos de interpretación que orientaban su análisis. Al mismo tiempo, la digitalización masiva de contenidos, así como las relecturas de la historia literaria desde perspectivas digitales, han permitido "ver más, mejor y de otra manera los legados culturales" (Romero López, 2024: 325), abarcando prácticas y enfoques transdisciplinares que abren nuevas vías tanto en el ámbito de la docencia como en el de la investigación.

En este contexto, *Mnemosine*[1] se configura como una biblioteca digital de acceso abierto que alberga datos y metadatos de textos y autores "raros y olvidados" de la Edad de Plata, constituyendo un repertorio de la literatura española entre los años 1868 y 1936 que permite reconocer y rescatar "otra" historia cultural, en tanto que favorece la revisión historiográfica de este período (Romero López y Bueren, 2019). Fruto de la colaboración entre bibliotecarios, filólogos del grupo de investigación LOEP e informáticos del grupo ILSA, la Biblioteca Digital Mnemosine no solo ofrece acceso a libros en formato digital y a la consulta de su base de datos, sino que también supone un campo de experimentación en el modelado de datos de colecciones literarias específicas y en la creación de redes semánticas interoperables (Romero López, 2021), convirtiéndose, en consecuencia, en un modelo metodológico pionero en lo que respecta a las posibilidades de producción, estructuración e interpretación del discurso historiográfico sobre la Edad de Plata. Existen tres vías de acceso al corpus: "Cartografía digital" (mapa interactivo, didáctico y visual que representa la apertura de los estudios en torno a la Edad de Plata desde la Filología y la Historia hacia los Estudios Culturales y las Humanidades Digitales), "Historia y Filología" (conocimiento organizado de manera más tradicional en torno a categorías que ofrecen una aproximación sistemática y objetiva a los datos) y "Constelaciones culturales" (aspectos innovadores y modos de categorización alternativos o heterodoxos que posibilitan una lectura más abierta y desafían el canon al centrarse en otros géneros, ejes temáticos, identidades colectivas o tipos de lectores).

Mnemosine se presenta, de esta manera, como un recurso de gran potencial en el ámbito de la actividad académica, en tanto que se organiza en torno a colecciones diseñadas por expertos e incorpora aspectos de catalogación específicos que pueden resultar particularmente valiosos para el investigador (Cotarelo Esteban, 2018). Así, no se trata únicamente de un repositorio de textos digitalizados, sino de una herramienta que articula el acceso, el análisis y la generación de conocimiento proponiendo un cambio metodológico para ampliar y repensar las perspectivas desde las que se aborda nuestro legado cultural. En este sentido, una historia digital de la Edad de Plata activa necesariamente distintos planteamientos de carácter ético y social en torno a las relaciones entre canon, archivo, historia cultural y memoria histórica

[1] http://repositorios.fdi.ucm.es/mnemosine/

(Moretti et al., 2016), que es imprescindible trasladar a las aulas: desde el punto de vista epistemológico, Mnemosine propone un cambio de paradigma: no solo qué se estudia, sino cómo se construye el conocimiento.

Frente a modelos estáticos y jerárquicos de la historiografía tradicional, esta biblioteca opera desde una lógica relacional, no jerárquica y no lineal, donde los datos se vinculan semánticamente y se abren a múltiples trayectorias interpretativas que permiten, por un lado, la ruptura con los automatismos del canon, perpetuados en gran parte por los límites materiales y conceptuales derivados de los métodos de análisis analógicos, así como una redefinición del canon literario a través de los metadatos (Romero López, 2021); y, por otro, un proceso de ampliación, revisión y enriquecimiento del contenido docente, en el que se integran ahora escritoras olvidadas, autores exiliados, perspectivas feministas, literaturas populares, géneros marginales o ejes temáticos novedosos. Esta ampliación no se limita a un incremento cuantitativo del contenido disponible, sino que opera mediante una reorganización cualitativa del saber literario: la incorporación de metadatos permite visibilizar variables antes frecuentemente ignoradas (como el género o la nacionalidad, pero también las redes de contacto, los vínculos con instituciones o las relaciones con otros agentes culturales, como traductores o editores), articulando colecciones alternativas basadas en criterios actualizados que reflejan una nueva forma de aproximación crítica a la cultura de la Edad de Plata (González Soriano, 2022). Asimismo, la incorporación de recursos de gran potencial didáctico (visualizaciones, recursos de geolocalización, la colección Libros Digitalizados, el proyecto de edición digital La Edad de Plata Interactiva, etc.) fomenta no solo el desarrollo de prácticas activas de aprendizaje vinculadas con los nuevos retos del mundo digital, sino también la reflexión en torno a los mecanismos de construcción del canon literario y la legitimidad de los relatos culturales heredados, a la vez que propone un modelo de superación de conceptos obsoletos, como el de generación literaria, y contribuye a la articulación de un canon alternativo sobre el que trabajar (Erll y Rigney, 2009). La historia literaria, enfocada así hacia la superación de la fragmentación analógica generacional y del acceso desigual a los textos literarios y otros materiales de la Edad de Plata (Romero López, 2024), se construye en torno a la necesidad de integrar, conectar y hacer accesible un legado cultural previamente disperso u oculto, organizado ahora de manera más justa y coherente. En este proceso resulta imprescindible el conocimiento experto y el trabajo colaborativo de filólogos, bibliotecarios e informáticos, así como la colaboración con instituciones como la Biblioteca Nacional de España, la Biblioteca Virtual Cervantes, Hathitrust o el Instituto Iberoamericano de Berlín. De esta manera, en la práctica investigadora, herramientas como la Biblioteca Digital Mnemosine permiten “ver más y mejor” (Romero López y Bueren, 2019). La historiografía literaria se enriquece con una mirada más global, polifónica y conectada, al facilitar usos y enfoques diversos: desde la generación de nuevos materiales y preguntas de investigación novedosas hasta la validación de hipótesis o la identificación de objetos de

estudio emergentes que pueden integrarse en futuras investigaciones o en recursos de instituciones públicas. En este contexto, en el que la historia cultural y digital avanza hacia nuevos marcos interpretativos de carácter transversal y transdisciplinar, la producción de conocimiento colaborativo, a partir de datos en abierto, interoperables y reutilizables, se consolida como un aspecto fundamental en lo que respecta al compromiso con una ciencia abierta, ética y socialmente responsable, capaz de abordar los materiales del pasado desde las preguntas y las necesidades que plantea la sociedad contemporánea.

Bibliografía

COTARELO ESTEBAN, L. (2018), «La colección “Poetas en el exilio” en Mnemosine: Biblioteca Digital de La Otra Edad de Plata», en D. ROMERO LÓPEZ & M. J. SALAMANCA LÓPEZ (coords.), *Entornos digitales. Humanidades y Ciencias Sociales en la Universidad Complutense de Madrid*, pp. 25–39, Red de Humanidades Digitales.

ERLL, A., & RIGNEY, A. (eds.) (2009), *Mediation, Remediation, and the Dynamics of Cultural Memory*, Berlin / Nueva York, Walter de Gruyter.

GONZÁLEZ SORIANO, J. M. (2022), «Mnemosine: A Digital Platform for Research and Rediscovery of the Other Silver Age Spain», en D. ROMERO LÓPEZ & J. ZAMOSTNY (eds.), *Towards the Digital Cultural History of the Other Silver Age Spain*, Berlín, Peter Lang, pp. 89–106.

MAINER, J. C. (2010), *Historia de la Literatura Española. Modernidad y nacionalismo. 1900–1939*, Barcelona, Crítica.

MORETTI, F., et al. (2016), *Canon/Archive. Large-scale Dynamics in Literary Field*, California, Literary Lab.

ROMERO LÓPEZ, D., & BUEREN, J. L. (2019), «La colección Literatura de Quiosco en Mnemosine, Bibloteca Digital de la Otra Edad de Plata (1868–1936): hacia la redefinición del canon literario a través de los metadatos», *Literatura: Teoría, Historia, Crítica*, vol. 21, n.º 1, pp. 61–91.

ROMERO LÓPEZ, D. (2021), «Archivos, repositorios y bibliotecas digitales: hacia una historia digital de la Edad de Plata», *Signa: Revista de la Asociación Española de Semiótica*, n.º 30, pp. 17–30.

ROMERO LÓPEZ, D. (2024), «Mnemosine, la biblioteca digital académica como canon alternativo para el estudio de la Edad de Plata (1868–1936)», en M. I. MORALES SÁNCHEZ, et al. (eds.), *Literatura y prácticas digitales: una mirada interdisciplinar alrededor de la formación literaria*, Valencia, Tirant lo Blanch, pp. 325–348.

Sobre datos, grafos y redes en abierto: Una lectura distante de María Lejárraga (1874-1974)

Dolores Romero López
Patricia García Sánchez-Migallón
Universidad Complutense de Madrid

Palabras clave: Lectura distante; María Lejárraga; redes de socialización; Edad de Plata
Keywords: Distant reading; María Lejárraga; social networks; Silver Age

Desde la perspectiva del contenido, esta comunicación examina las contribuciones de María Lejárraga al feminismo, el modernismo y el teatro durante la Edad de Plata en España, centrándose en las redes sociales en las que participó y que fomentó entre 1900 y 1936. Además, desde el punto de vista metodológico, esta comunicación ofrece un análisis comparativo entre datos extraídos manualmente de fuentes impresas históricas y conjuntos de datos digitales abiertos proporcionados por instituciones nacionales e internacionales. Al evaluar ambos tipos de datos, exploramos su potencial para contribuir a un proyecto de visualización de redes de socialización e intelectuales, al tiempo que identificamos disparidades críticas en la fiabilidad, integridad y accesibilidad de los datos, enfatizando la necesidad de datos curados por expertos para una investigación rigurosa en humanidades.

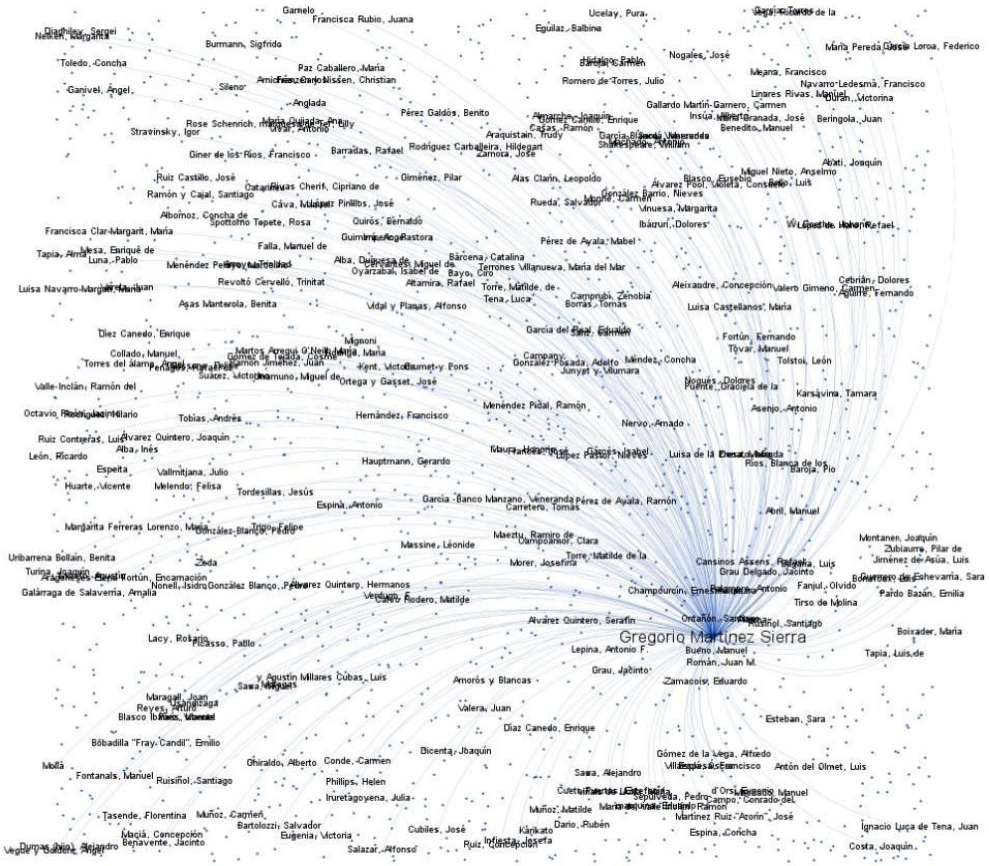
En el caso específico de esta comunicación, se han descargado los datos proporcionados por instituciones públicas sobre la vida y producción literaria del matrimonio Martínez Sierra. Estas instituciones incluyen la Biblioteca Nacional de España, la Residencia de Estudiantes, la Biblioteca Virtual Miguel de Cervantes, HathiTrust y la web de Revistas Culturales 2.0. También intentamos descargar datos de Wikidata, pero la cantidad de información disponible resultó insuficiente. El objetivo era verificar que estas instituciones no proporcionan los datos digitales abiertos necesarios para analizar redes de socialización. Solo pudimos descargar datos abiertos en una cantidad suficiente de la Biblioteca Digital Mnemosine, disponible en GitHub y Zenodo. Gracias a estos datos abiertos y a los datos en bruto procedentes de bibliografía experta (ver más abajo), hemos podido crear nuestra propia base de datos, que puede servir como base para futuras investigaciones sobre otros autores de la llamada Edad de Plata de la literatura española (1900-1939).

La comunicación propone un flujo de trabajo basado directamente en las recomendaciones de la infraestructura europea DARIAH. Para encontrar una herramienta útil de visualización de redes sociales, realizamos una búsqueda en

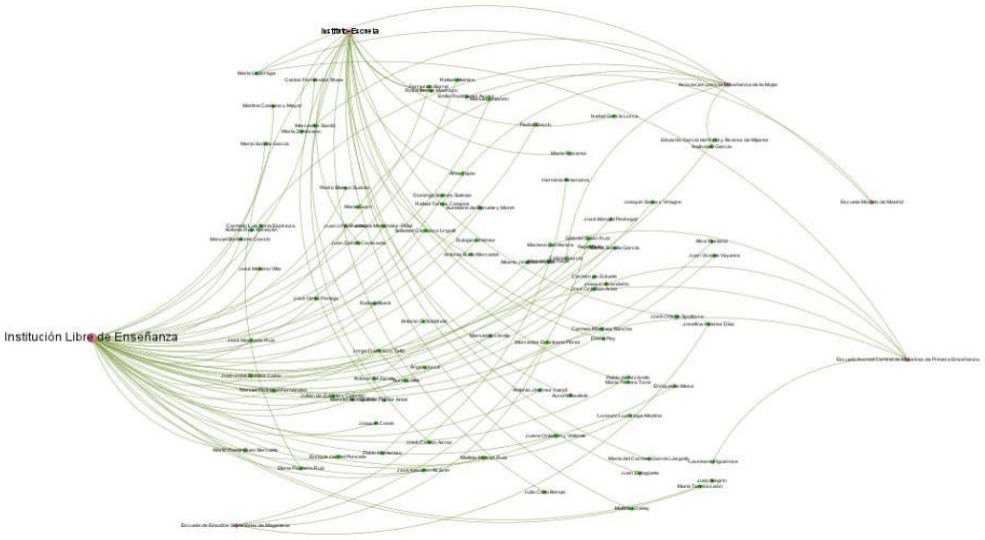
el SSHOC Open MarketPlace y encontramos en Gephi una opción muy potente. En el DARIAH Campus, hallamos numerosos materiales de formación para aprender a usar la herramienta y, tras la curación de datos y la creación de la visualización, el conjunto de datos resultante se subió al repositorio de Zenodo y luego se vinculó de nuevo al SSHOC Open MarketPlace. De esta manera, garantizamos la circularidad de la investigación, la aplicación de los principios FAIR y nuestro compromiso con la Ciencia Abierta.

El estudio concluye que las redes de socialización de María Lejárraga fueron fundamentales para su emancipación y el avance del feminismo en España y se percibe claramente que son tres sus focos de interés: las revistas del modernismo, el teatro de arte y las asociaciones de mujeres.

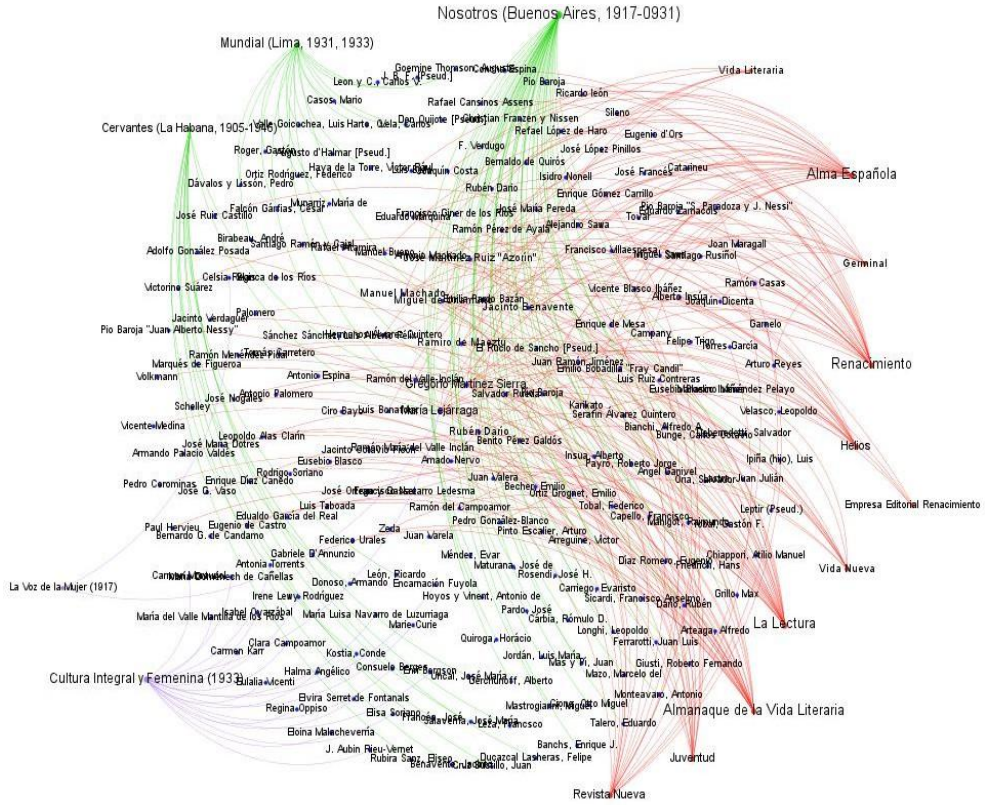
Grafos



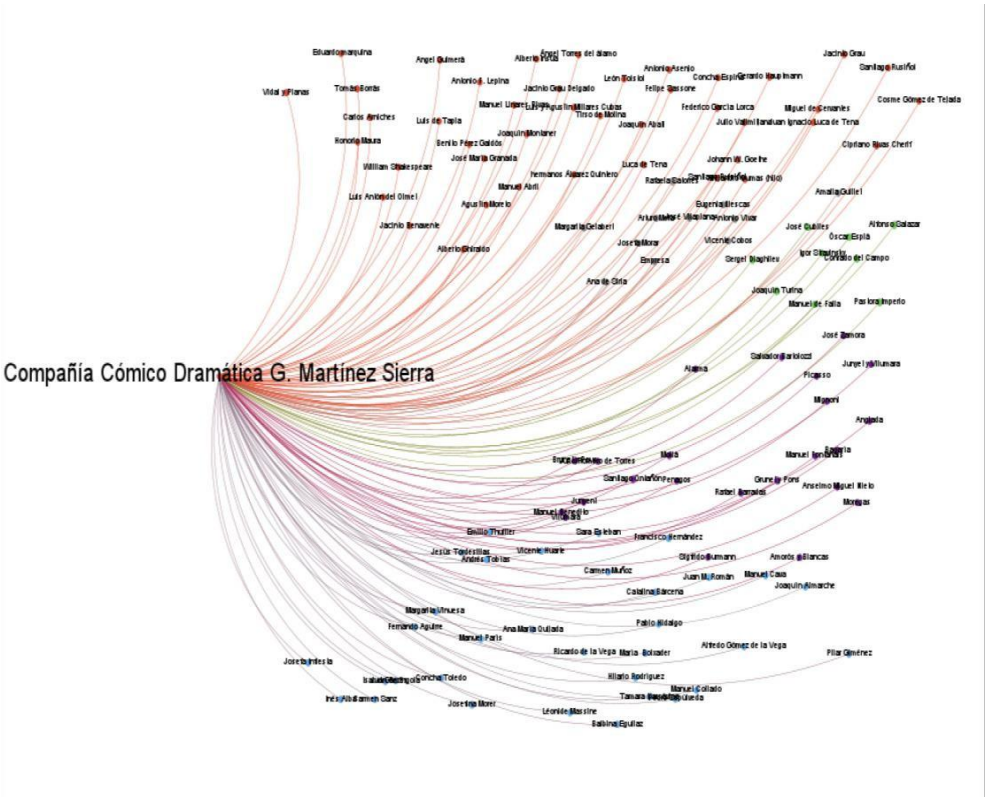
Grafo 1. Relaciones de los Martínez Sierra con otros intelectuales de su tiempo



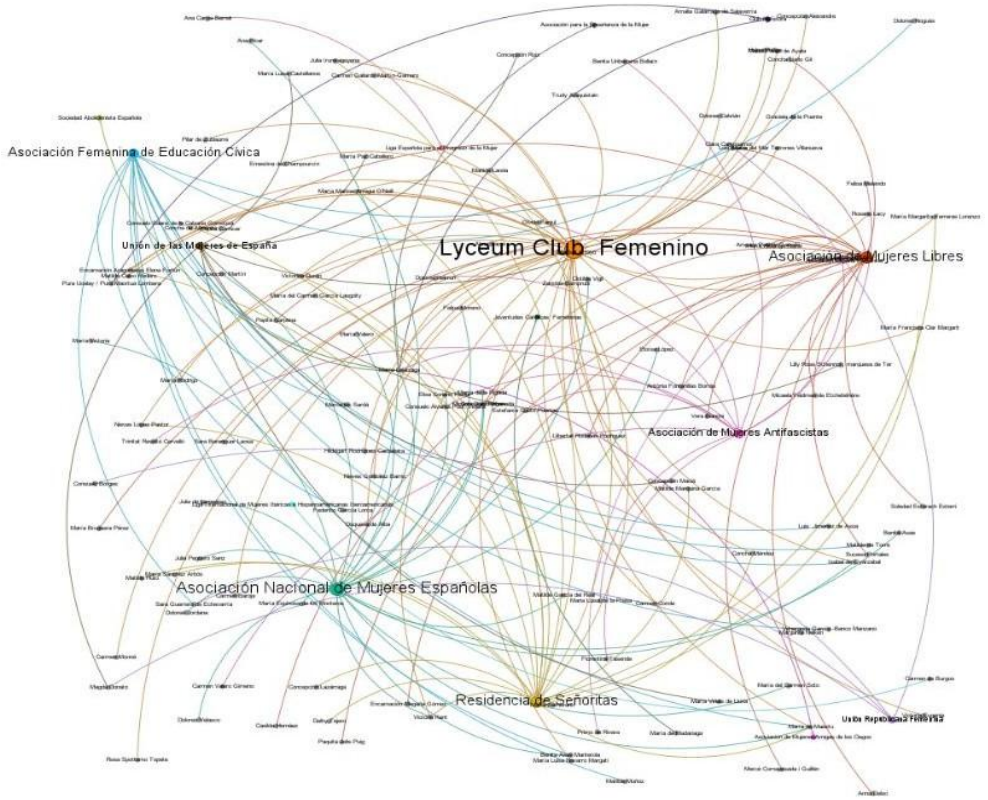
Grafo 2. Instituciones educativas con las que se vincula María Lejárraga



Grafo 3. Revistas culturales relacionadas con Maria Lejarraga



Grafo 4. Relaciones de los Martínez Sierra con el Teatro de Arte, Compañía Cómico Dramática



Grafo 5. Relaciones de mujeres con asociaciones femeninas en la época de María Lejárraga

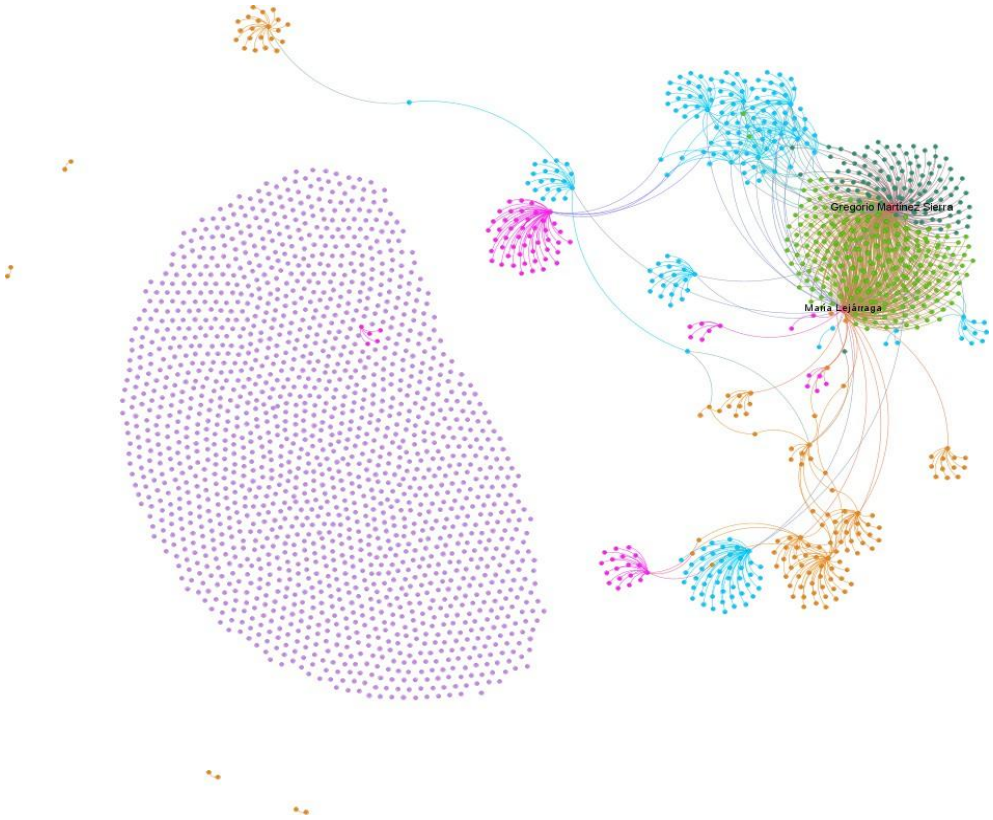


Figura 6. Grafo general de las relaciones del matrimonio Martínez Sierra-Lejárraga

Figura 7. Workflow

Bibliografía

AGUILERA SASTRE, J. (2021), «Para una historia de las asociaciones femeninas en España. La Asociación Nacional de Mujeres Españolas y la Unión de las Mujeres de España: similitudes y discordancias (1918–1921)», *Feminismo/s*, vol. 37, pp. 131–160.

COMELLA GUTIÉRREZ, B., & MONTERO DÍAZ, M. (2022), *Sororidad. Redes de ayuda entre mujeres en los siglos XIX y XX*, Madrid, Dykinson.

DEVÉS, E. (2007), *Redes intelectuales en América Latina: hacia la constitución de una comunidad intelectual*, Santiago de Chile, Instituto de Estudios Avanzados, Universidad de Santiago de Chile.

LEJÁRRAGA, M. de la O. (2019), *Una mujer por caminos de España*, Madrid, Editorial Castalia.

LEJÁRRAGA, M. de la O. (2023), *Gregorio y yo: medio siglo de colaboración*, Sevilla, Renacimiento.

MARTOS, M. D. (2021), *Redes y escritoras ibéricas en la esfera cultural de la primera Edad Moderna*, Madrid, Iberoamericana / Vervuert.

OCAMPOS, E., & ROMERO LÓPEZ, D. (eds.) (2024), *El feminismo en la literatura en la Edad de Plata*, Madrid y Sevilla, Ediciones Complutense y Editorial Renacimiento.

RODRIGO, A. (2024), *María Lejárraga, una mujer en la sombra*, Granada, Editorial Universidad de Granada.

ROMERO LÓPEZ, D., & ZAMOSTNY, J. (eds.) (2022), *Towards the Digital Cultural History of the Other Silver Age Spain*, Berna, Peter Lang.

Desafíos y posibles soluciones para el diseño y la compilación de corpus lingüísticos multimodales en línea: el caso de WA2

Elia Puertas Ribés
Universitat Jaume I

Palabras clave: Multimodalidad; emojis; conversaciones de WhatsApp; lingüística de corpus; Sketch Engine.
Keywords: Multimodality; emojis; WhatsApp conversations; corpus linguistics; Sketch Engine.

La multimodalidad (Yus 2001), entendida como un tipo de comunicación que aúna diferentes recursos semióticos (imágenes, texto escrito, audio, etc.) en un mismo mensaje, es fruto de las posibilidades técnicas que ofrecen los dispositivos electrónicos actuales –como pueden ser los teléfonos móviles o los ordenadores–, por lo que nos hallamos ante una especie de constructos sociotécnicos, es decir, una combinación entre las características técnicas de estas plataformas y las prácticas sociales, culturales y comunicativas que surgen a su alrededor (Sampietro 2023; Herring 2015).

De ahí que las aplicaciones digitales hayan transformado profundamente las formas de comunicación humana, al mismo tiempo que han promovido nuevos retos y oportunidades para la investigación lingüística. Ahora bien, la ampliación de fuentes para la obtención de datos textuales generados por usuarios en aplicaciones, como las de mensajería, crece a la par que los desafíos metodológicos sobre su extracción, organización y compilación. En este sentido, la lingüística de corpus se erige como una disciplina que ha contribuido significativamente a una mejora en el estudio de las lenguas, tanto cuantitativa como cualitativamente (Tognini Bonelli 2001). De hecho, la lengua española cuenta con grandes corpus de referencia para los estudios sincrónicos, entre ellos se encuentran el CREA y el CORPES XXI, creados por la Real Academia Española. Sin embargo, el campo de la comunicación digital carece todavía de bases de datos de referencia que permitan el estudio del discurso que se genera a través de los dispositivos electrónicos. Afortunadamente, hay cada vez más corpus que recogen el uso de la lengua en la red, como el EsTenTen, disponible en Sketch Engine, aunque la mayoría de ellos han sido diseñados con objetivos específicos y no siempre accesibles (De Benito Moreno 2022) y, además, suelen carecer de códigos semióticos distintos al texto, esto es, emojis, stickers, imágenes, etc.

En este contexto, el grupo de investigación CoMuDiCo, coordinado por la profesora Agnese Sampietro, de la Universitat Jaume I, ha iniciado el desarrollo del corpus WA2, una colección de conversaciones extraídas de WhatsApp, que han sido compiladas a lo largo de la última década. El corpus WA2, además, se compone por más de un centenar de conversaciones en español peninsular, lo que permite un estudio detallado de cómo los elementos multimodales interactúan con otras unidades discursivas en diferentes contextos comunicativos. Los mensajes seleccionados provienen de interacciones cotidianas entre amigos, familiares y compañeros de trabajo o estudio, de modo que reflejan un grado elevado de proximidad entre los interlocutores y propician el uso de distintos recursos multimodales. En este contexto, los emojis desempeñan diversas funciones discursivas, como expresar estados emocionales, matizar el tono del mensaje o complementar la información verbal. Sin embargo, como se pretende mostrar en la comunicación, su incorporación en el diseño de un corpus accesible en línea plantea todavía una serie de desafíos metodológicos.

En cuanto al método de obtención de los datos, la creación de WA2 se basa en un proceso que ha sido empleado en investigaciones anteriores por la profesora Sampietro (2016a, 2016b, 2017), y que se sustenta fundamentalmente en que los informantes envíen sus conversaciones de WhatsApp, bien a través del correo electrónico, bien mediante un formulario en línea. A continuación, los textos se transcriben por el equipo de investigación para garantizar la uniformidad y, seguidamente, se convierten a formato de texto plano (TXT) con codificación UTF-8 para facilitar su procesamiento en herramientas de análisis de corpus. Además, con el fin de preservar la confidencialidad de los participantes, se implementa un proceso de anonimización de los antropónimos mediante un código desarrollado en Python, para el que se diseña un diccionario (estructura de datos tipo clave-valor) que contenga los nombres de los interlocutores y sus correspondientes pseudónimos, para así realizar la sustitución automática.

Una vez recopilados y procesados los datos, el corpus WA2 se compila en Sketch Engine, una plataforma en línea de análisis de corpus que permite la gestión de grandes volúmenes de datos lingüísticos. Sketch Engine es especialmente útil para este tipo de investigaciones debido a su capacidad para realizar un etiquetado morfosintáctico de los documentos. A través de esta herramienta, pues, los investigadores tienen acceso a diversas funcionalidades que posibilitan identificar patrones lingüísticos recurrentes en el corpus, así como extraer combinaciones léxicas y estudiar las coocurrencias de palabras. Además, resulta especialmente útil cuando se trata de analizar el uso de los emojis, ya que Sketch Engine permite realizar búsquedas por medio de etiquetas específicas. Por ello, se propone modificar la configuración de la estructura preestablecida por la plataforma con el fin de añadir una nueva etiqueta representativa para los emojis. Del mismo modo, se sugiere el uso de ese nuevo identificador en la anotación morfológica del corpus, así como el empleo de los

códigos internacionales de Unicode para su lematización en el etiquetado vertical, proceso que facilitaría su recuperación por medio de búsquedas avanzadas. Por último, se ofrece una propuesta para un etiquetado estructural de los mensajes con el fin de aportar más información por medio de atributos como el sexo del participante, el tipo de conversación (diádica o grupal), fecha y hora de cada intervención, etc.

Esta comunicación, en definitiva, tiene como objetivo principal abordar los desafíos asociados al diseño y desarrollo de un corpus multimodal, dado que, en un entorno comunicativo como es el digital, en el que los elementos no verbales –como los gestos o la prosodia– están ausentes, los usuarios de WhatsApp recurren a los emojis para suplir esta falta de recursos comunicativos, de forma que les brinda la oportunidad de mantener cierto nivel de expresividad en sus mensajes. En consecuencia, es importante plantear estándares metodológicos para el desarrollo de estos corpus que integran diferentes códigos. Por nuestra parte, su puesta en práctica y accesibilidad para los usuarios se lleva a cabo mediante la plataforma de Sketch Engine, ya que facilita el análisis lingüístico y multimodal de grandes volúmenes de datos, incluyendo los emojis. Así pues, integrando el corpus WA2 en esta herramienta, los investigadores tienen la posibilidad de realizar análisis más precisos y detallados del uso los elementos multimodales en el ámbito digital, lo que permite avanzar hacia una comprensión más profunda de la dinámica comunicativa en las plataformas de mensajería instantánea. De este modo, tras adaptar la anotación estructural y el etiquetado de los emojis, el corpus no solo facilitará el estudio de los patrones lingüísticos en el texto, sino que también abrirá nuevas posibilidades para el análisis de los recursos multimodales, que juegan un papel cada vez más importante en la comunicación digital.

Bibliografía

DE BENITO MORENO, C. (2022), «Uso de los medios digitales de comunicación como corpus de español», en G. PARODI SWEIS, P. CANTOS GÓMEZ, & C. HOWE (coords.), *Lingüística de corpus en español*, pp. 481–493.

HERRING, S. C. (2015), «New frontiers in interactive multimodal communication», en A. GEORGAKOPOULOU & T. SPILIOTI (eds.), *The Routledge Handbook of Language and Digital Communication*, Routledge, pp. 398–402.

SAMPIETRO, A. (2016a), «Exploring the punctuating effect of emoji in Spanish WhatsApp chats», *Lenguas Modernas*, 47, pp. 91–113.

SAMPIETRO, A. (2016b), «Emoticonos y multimodalidad. El uso del pulgar hacia arriba en WhatsApp», *Aposta: Revista de Ciencias Sociales*, 69, pp. 271–295.

SAMPIETRO, A. (2017), «Emoticonos y cortesía en los mensajes de WhatsApp en España», en P. C. GUBITOSI (coord.) & G. M. GIAMMATTEO (ed. lit.), *El español en la red*, pp. 279–302.

SAMPIETRO, A. (2023), *Lenguaje e imagen en la comunicación digital*, Arco/Libros, S. L.

TOGNINI BONELLI, E. (2001), *Corpus Linguistics at Work*, John Benjamins Publishing Company.

YUS, F. (2001), *Ciberpragmática: El uso del lenguaje en Internet*, Ariel.

La extracción automática de datos con Python y Selenium aplicada a la lingüística de corpus

Elia Puertas Ribés
Universitat Jaume I

Palabras clave: Web scraping; Selenium; extracción de datos; CORDE
Keywords: Web scraping; Selenium; data extraction; CORDE

En el ámbito de la lingüística de corpus, el español dispone de numerosas bases de datos de referencia, tanto para el estudio diacrónico como para el sincrónico. Estas plataformas se constituyen, por consiguiente, como herramientas de consulta obligada para el análisis de cualquier fenómeno lingüístico que parta de datos reales, ya que recopilan una gran cantidad de textos que abarcan desde los orígenes del castellano hasta la actualidad. El auge de nuevos corpus lingüísticos, con un nivel de desarrollo cada vez mayor, ha fomentado, a su vez, una reflexión crítica sobre sus beneficios, así como sobre las limitaciones que implican algunas de estas aplicaciones (Blas Arroyo y Puertas Ribés 2024: 9). Por ejemplo, en el plano de la diacronía, los corpus convencionales, que se conforman por una base de datos informatizada y un mecanismo de búsqueda, carecen de la posibilidad de descargar las concordancias, de modo que dificultan la labor investigadora y ralentizan el proceso de organización y estudio de los testimonios obtenidos en la consulta (Vaamonde 2015). La extracción manual de información en este tipo de plataformas representa un trabajo tedioso, pues los investigadores deben copiar y organizar los datos uno a uno, sin asistencia automatizada, lo que supone un consumo excesivo de tiempo (Torruella y Llisterri 1999: 4). Además, este método conlleva el riesgo de cometer errores humanos que pueden afectar a la precisión de los estudios lingüísticos.

Ante este reto, la Inteligencia Artificial ha adquirido un papel decisivo en el Procesamiento del Lenguaje Natural (PLN), un campo que, según Jurafsky y Martin (2021), ha experimentado grandes avances gracias a modelos computacionales que han facilitado el desarrollo de herramientas avanzadas para la recuperación, análisis y visualización de grandes volúmenes de textos. De hecho, para la extracción de datos de la web de forma automática, puede recurrirse al web scraping, una técnica que busca, descarga y procesa contenidos de una página sin intervención manual (Salve Cuenca 2025: 304). Sin embargo, la obtención y el acceso a los datos es diferente en función de la configuración del sitio web. Por eso, en páginas estáticas —cuyos archivos HTML, CSS y JavaScript ya están generados en el servidor— la recuperación de

datos por medio del web scraping resulta relativamente sencilla, ya que es posible extraer la información utilizando directamente herramientas diseñadas específicamente para ello, como BeautifulSoup, una librería de Python especializada en el análisis de documentos HTML y XML.

Sin embargo, en sitios dinámicos —como es el caso de muchas bases de datos textuales convencionales—, el contenido se genera o modifica en el servidor en respuesta a interacciones del usuario: por ejemplo, hacer clic en botones, enviar formularios, etc. Por ello, la configuración del backend es más compleja que las estáticas y, en consecuencia, el uso de librerías como BeautifulSoup no es suficiente. En estos casos es preciso simular la navegación real por la web para que el servidor recupere los datos solicitados de las búsquedas. En nuestro caso, esta labor se pone en práctica por medio de Selenium, un framework de automatización de navegadores que permite controlar de forma remota un sitio web —emulando clics, rellenando formularios, navegación compleja, etc.— y recuperar el DOM resultante tras cada interacción, esto es, la estructura completa de la página con todos los elementos HTML y CSS ya actualizados tras la consulta.

De este modo, con un script escrito en Python, el uso de Selenium facilita localizar y extraer etiquetas, atributos o bloques de texto concretos, descargar archivos vinculados y, en definitiva, automatizar todo el recorrido que un investigador haría manualmente. En esta comunicación su aplicación se lleva a cabo en una web dinámica de la Real Academia Española, el Corpus Diacrónico del Español (CORDE), que, a diferencia de otras bases de datos textuales de esta misma institución —que sí facilitan la descarga masiva de datos—, el CORDE no presenta este servicio de exportación de concordancias. Así pues, mediante la automatización de este proceso con Selenium, es posible recuperar las concordancias de manera rápida, sistemática, tarea que optimiza la obtención de los textos en un formato más cómodo y manejable, así como la descarga de los resultados a un formato estructurado (CSV o .xlsx). De este modo, no solo se ahorra tiempo, sino que garantiza exhaustividad y reproducibilidad en los estudios diacrónicos: cada búsqueda se registra y almacena de manera uniforme, minimizando errores humanos y facilitando el posterior tratamiento estadístico o lingüístico de los datos.

En conclusión, la presente comunicación tiene como objetivo contribuir a la optimización del trabajo de recuperación y almacenamiento de datos lingüísticos. Para ello, se ha diseñado una propuesta que se sirve de las herramientas que ofrece la lingüística computacional para integrar Selenium y Python en el análisis de una base de datos de referencia como el CORDE. La automatización de la extracción de datos lingüísticos en este corpus reduce significativamente el tiempo de extracción y minimiza errores en el almacenamiento de datos, de modo que permite realizar estudios a gran escala que serían muy tediosos mediante métodos manuales. Gracias a la capacidad de esta herramienta para procesar información de manera estructurada, los lingüistas y filólogos pueden obtener resultados más precisos y exhaustivos.

Sea como fuere, a pesar de los retos que implica su implementación en una página dinámica como el CORDE, el uso de herramientas de web scraping representa un paso adelante en la consolidación de metodologías digitales aplicadas a la investigación de la lengua española.

Bibliografía

BLAS ARROYO, J. L., & PUERTAS RIBÉS, E. (2024), «INCOM: Un corpus de inmediatez comunicativa para el estudio sociolingüístico del español en su historia», *Cultura, Lenguaje y Representación*, vol. XXXIII, pp. 7–29.

JURAFSKY, D., & MARTIN, J. H. (2021), *Speech and Language Processing*, Prentice Hall.

SALVE CUENCA, A. (2025), «El papel de las herramientas de web scraping en la enseñanza del análisis del discurso digital», en R. HIDALGO DOWNING (ed.), *Tecnologías digitales aplicadas a la enseñanza de las lenguas y a la lingüística*, Guillermo Escolar, pp. 303–326.

TORRUELLA, J., & LLISTERRI, J. (1999), «Diseño de corpus textuales y orales», en J. M. BLECUA, G. CLAVERÍA, C. SÁNCHEZ, & J. TORRUELLA (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Editorial Milenio, pp. 45–77.

VAAMONDE, G. (2015), «Limitaciones en el uso de corpus diacrónicos del español: Nuevas aportaciones desde el proyecto de investigación “Post Scriptum”», *E-AESLA*, 1. ISSN 2444-197X.

Catálogo y Biblioteca Digital de Relaciones de Sucesos: análisis y visualización de las colecciones que conforman la base de datos

Carlota Fernández Travieso
Nieves Pena Sueiro
Universidade da Coruña

Palabras clave: Relaciones de sucesos; colecciones documentales; catálogo bibliográfico; biblioteca Digital; base de datos relacional
Keywords: Event reports; documentary collections; bibliographic catalog; digital library; relational database

En 1994, el grupo SIELAE, responsable del proyecto Biblioteca Digital del Siglo de Oro (BIDISO), comenzó las tareas que dieron lugar al Catálogo y Biblioteca Digital de Relaciones de Sucesos (CBDRS). Esta iniciativa se ocupa de localizar ejemplares, identificar ediciones y describir relaciones de sucesos, documentos protoperiodísticos de publicación ocasional, el género editorial más utilizado hasta el siglo XVIII para dar noticia de acontecimientos histórico políticos o festivos y sucesos diversos. En la actualidad, CBDRS (que se halla disponible en acceso abierto desde 2001 https://www.bidiso.es/CBDRS/) ofrece datos de 11.636 ejemplares de bibliotecas de todo el mundo, descripciones bibliográficas de 6.992 ediciones y 3.121 facsímiles digitales, cuya recuperación es facilitada por una compleja base de datos relacional modelada tomando como base el conocimiento experto sobre los aspectos que permiten el control de la producción de las relaciones de sucesos y constatar sus principales características.

Modelo conceptual de CBDRS en la actualidad.

A través de un punto de consulta único, CBDRS permite a los interesados en el género ahorrar tiempo en búsquedas en diversos catálogos y repertorios y, cuando es posible ofrecer reproducción digital de los ejemplares, evitar gestiones y desplazamientos para ver y conseguir copias de los impresos hallados en distintas instituciones. Asimismo, gracias al acopio de una ingente cantidad de datos sobre la producción que se conserva de este tipo de obras y que el uso de la tecnología permite reunir y analizar con mayor facilidad, CBDRS puede contribuir a establecer fehacientemente la naturaleza y características de las relaciones como producto y género editorial, valorar su difusión y circulación y observar su evolución a lo largo del periodo en el que se produjeron.

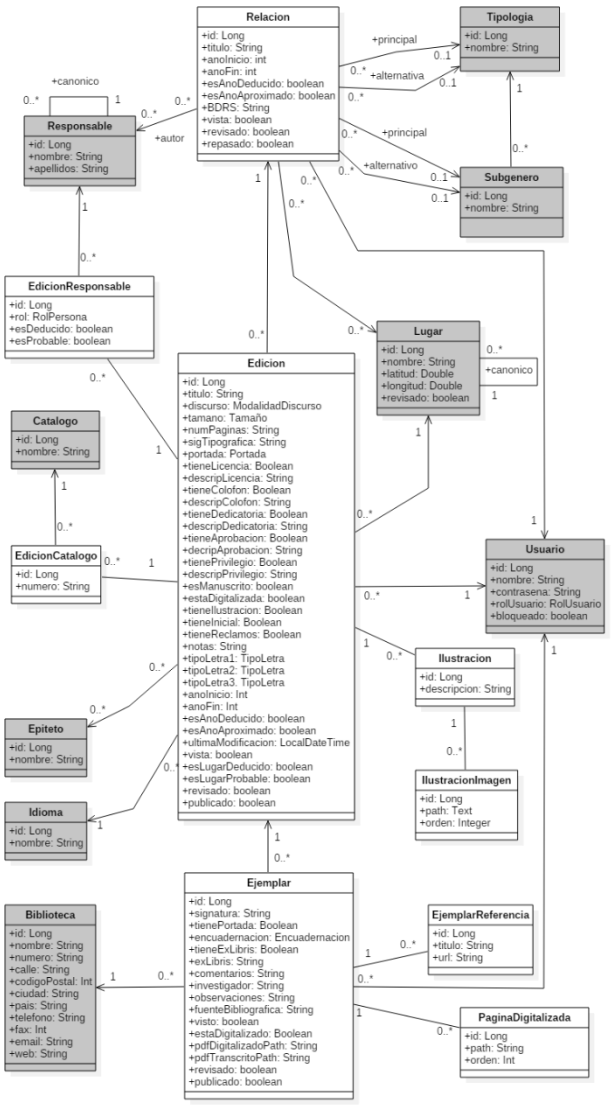


Figura 1. Modelo conceptual de CBDRS en la actualidad.

Si bien, el grupo ha producido ha realizado algunos trabajos con atención a la globalidad del corpus registrado en CBDRS estableciendo importantes conclusiones, por ejemplo, en referencia a los lugares de impresión e impresores dedicados a la producción de relaciones de sucesos (Fernández Travieso y Pena Sueiro, 2013), a sus autores (Pena Sueiro, 2017a), o a un lugar como centro editorial, Sevilla, en este caso (Espejo-Cala y Baena-Sánchez, 2019), la labor de análisis de resultados de la elaboración del catálogo ha quedado hasta ahora en segundo plano frente a la prioridad de la localización de ejemplares, identificación y descripción de ediciones, sobre las que fundamentar estudios posteriores. Para avanzar en este aspecto y como paso previo indispensable para dar validez a próximos trabajos con base en estos datos, nos hemos propuesto dejar constancia de cómo se ha conformado el corpus que alberga CBDRS. Así, esta comunicación tiene como principal objetivo el análisis, visualización e interpretación de los datos reunidos sobre las colecciones

procedentes de diversas instituciones documentales que conforman nuestro catálogo y biblioteca digital.

Para realizar este trabajo, hemos partido del análisis de los datos acumulados en la propia base de datos CBDRS a lo largo de más de treinta años y de memorias e informes internos del proyecto BIDISO que dejan constancia de la labor y colaboración de los miembros del SIELAE con distintas bibliotecas, archivos, centros culturales o instituciones religiosas. La utilización de las opciones de visualización de Voyant tools y el software de IA (Chatgpt) nos han permitido agilizar los cómputos y explorar de manera gráfica los resultados.

Como resultado de este análisis, se establece que CBDRS reúne datos procedentes de 343 colecciones de relaciones de sucesos. Las colecciones con fondo más rico son la Biblioteca Nacional de España (1357 ediciones), la Bibloteca de la Universidad de Barcelona (959) y la Biblioteca universitaria de Sevilla (539), entre las públicas. Entre las privadas, destaca la Biblioteca de los herederos del Duque de T'Serclaes (103). Con respecto a las bibliotecas extranjeras, sobresalen las colecciones de la Biblioteca Nacional de Portugal (368), la Biblioteca Universitaria de Coimbra (347), la Biblioteca Nacional de Nápoles (con 100 ejemplares), la Biblioteca Nacional de Venecia (60) o la Biblioteca Nacional de Austria (38) y, en Asia, la Biblioteca de la Universidad de Tsukuba (21). Queda mucho por hacer, especialmente en Hispanoamérica y Asia, donde apenas se han realizado pequeñas prospecciones y consultas.

Se concluye que gracias a este trabajo se facilita la comprensión de cómo se ha ido construyendo un catálogo que se ha consolidado como referencia imprescindible en el campo de las relaciones de sucesos y el establecimiento objetivos de futuro para seguir enriqueciendo la base datos, al tiempo que seguimos profundizando en la historia del propio grupo SIELAE, que ha sido pionero en el ámbito de las Humanidades Digitales (Pena Sueiro, 2017b, Pena Sueiro y Saavedra Places, 2019).

Se concluye, además, que gracias a la incorporación de conocimientos interdisciplinares (Filología, Biblioteconomía, Historia, Historia de la Comunicación, Ciencias de la Computación) y el uso de herramientas informáticas (bases de datos relacionales, programas de visualización de datos, IA) se ha conseguido no solo elaborar y mantener una biblioteca digital y catálogo único en el mundo, no solo en lo que tiene que ver con el objeto de estudio, sino también con la gran cantidad de campos sobre él (prosa/verso, dedicatoria, colofón, portada, lugar del acontecimiento, ilustraciones) que posibilitan almacenar y recuperar, incluso cruzándola, información sobre él; para el análisis de datos y su visualización de manera sencilla y accesible, el uso de software de visualización y de la IA nos permitido generar de forma ágil gráficos y tablas que propician con un solo golpe de vista la observación de los resultados (colecciones controladas) y qué falta por hacer.

Bibliografía

ESPEJO-CALA, C., & BAENA-SÁNCHEZ, F. (2019), «Producción, estructura y mercado de la información en las relaciones de sucesos sevillanas (1500–1650)», en N. PENA SUEIRO & C. FERNÁNDEZ TRAVIESO (eds.), *Festina Lente Augusta Empresa Correr a Espacio. Studia in Honorem Sagrario López Poza*, A Coruña, Universidade da Coruña, pp. 77–94.

FERNÁNDEZ TRAVIESO, C., & PENA SUEIRO, N. (2013), «La edición de relaciones de sucesos en español durante la Edad Moderna», en P. CÁTEDRA (dir.) & M.ª E. DÍAZ TENA (ed.), *Géneros editoriales y relaciones de sucesos en la Edad Moderna*, Salamanca, SIERS–SEMYR, pp. 125–145.

PENA SUEIRO, N. (2017a), «Los autores de relaciones de sucesos: primeras precisiones», en V. NIDER & G. CAPELLI (eds.), *La invención de las noticias. Las relaciones de sucesos entre la literatura y la información (siglos XVI–XVIII)*, Trento, Università degli Studi di Trento, pp. 491–508.

PENA SUEIRO, N. (2017b), «El portal BIDISO: pasado, presente y futuro inmediato. Un ejemplo de evolución en aplicaciones de las HD», *Studia Aurea*, 11, pp. 73–92. https://doi.org/10.5565/rev/studiaaurea.264 [Acceso: 30/06/2022].

PENA SUEIRO, N., & SAAVEDRA PLACES, Á. (2019), «Obsolescencia y resiliencia en Humanidades Digitales. El caso de la Biblioteca Digital de Relaciones de Sucesos», en N. RODRÍGUEZ-ORTEGA (coord.), «Humanidades digitales: sociedades, políticas, saberes II», *Artnodes*, n.º 23, pp. 79–88. https://doi.org/10.7238/a.v0i23.3243.

Catalogación automatizada de incunables con la herramienta CACIN

Hala Neji
Javier Nogueras-Iso
Javier Lacasta
Instituto de Investigación en Ingeniería de Aragón (I3A) – Universidad de Zaragoza
Manuel José Pedraza-Gracia
Instituto de Patrimonio y Humanidades (IPH) – Universidad de Zaragoza

Palabras clave: Análisis y reconocimiento de documentos; aprendizaje automático; Incunables; metadatos; catálogo
Keywords: Document analysis and recognition; machine learning; incunabula; metadata; catalog

El estudio de la producción bibliográfica de un momento y ámbito específico ofrece una panorámica del conocimiento y el pensamiento en ese momento y ámbito. Identificar, por tanto, una edición coadyuva a la mejora de nuestro conocimiento y proporciona valores, incluso económicos, a los ejemplares de la misma.

En el periodo incunable y postincunable, el diseño y la elaboración material de los tipos metálicos para componer un libro es un proceso manual que lleva a cabo el propio taller de impresión (Zappella, 2001). Esta característica condiciona tanto el aspecto como la morfología de los incunables, porque cada taller posee unos tipos particulares que dan como resultado unos caracteres específicos.

Desde el inicio de la incunabulogía, en el siglo XVII, se persiguió más la localización de ediciones y ejemplares que la identificación de sus productores (Bradshaw, 1889; Barbieri, 2008). Sin embargo, los problemas derivados de la existencia de mútilos y membra disjecta, de la propia idiosincrasia de la producción del impreso, de los incunables sine notis, y de la ausencia de una sistematización científica dieron lugar a la elaboración de una metodología identificativa. A finales del siglo XIX, Robert Proctor aplica un método cuantitativo destinado a la identificación del taller que imprimió el incunable, mediante la medida de la huella que dejan en el papel veinte líneas de texto impreso con la misma fundición desde un punto de la línea uno hasta el mismo punto de la línea veintiuno, añadiendo tras esa dimensión una “R” para un tipo redondo o romano o una “G” para uno gótico (Proctor, 1905). De esta manera un tipo adquiere tanto

una identificación como una denominación con sentido: 84R o 162G, por ejemplo. Unos años después Konrad Haebler aplica, además, un método comparativo sustentado en la morfología de las dos letras más significativas: “M” para tipografía gótica y “Q” para la romana. Entre 1905 y 1924 publicó un muestrario de casi 4.000 “M”, “Qu” y otros signos; el Typenrepertorium der Wiegendrucke disponible en la actualidad a través del portal web de la Biblioteca Estatal de Berlín (Haebler, 1905-1924). La combinación de ambos sistemas ha dado lugar a una metodología que permite: a) denominar una tipografía; b) identificarla; y c) adscribirla a un taller específico (Rial Costas, 2012).

El objetivo de esta investigación es que el método Proctor-Haebler pueda ser aplicado de forma automática a ejemplares incunables y postincunables, evitando errores, superando la identificación en remoto y adecuándolo a muestras que presenten dificultades de aplicación.

En 2020 creamos un primer prototipo de software, desarrollado en Python, para automatizar el método Proctor-Haebler para la identificación de tipografías góticas (Lacasta et al., 2022). Este software incluye un algoritmo para detectar bloques de líneas continuas que se apoya en el uso de la biblioteca Tesseract para la segmentación de líneas y caracteres. El software también integra dos redes neuronales convolucionales para la detección de letras “M” y la identificación de la forma específica de “M” que se está utilizando. Con la información de la altura del bloque y la forma de la letra “M”, el software consulta un almacén de tripletas RDF, que replica los contenidos del Typenrepertorium der Wiegendrucke, para recuperar las tipografías que se corresponden con la altura y forma de “M” detectada, y los talleres de impresión que crearon esa tipografía.

Durante 2024 hemos mejorado este prototipo inicial, que solo podía ser utilizado en un entorno de pruebas y con conocimientos medio-altos de configuración de sistemas informáticos, para integrarlo en un conjunto de herramientas más robusto y escalable denominado CACIN, acrónimo de “Computer Aided Cataloguing of INcunabula” (Neji et al., 2025). CACIN se compone de varias herramientas para administradores y usuarios finales que colaboran con el objetivo de catalogar incunables con asistencia automatizada para la identificación de tipografías. La Figura 1 muestra un diagrama de despliegue con la arquitectura de CACIN donde se pueden observar tres partes diferenciadas: el software de entrenamiento (Training Software); la herramienta para los administradores (Admin frontend/backend); y la herramientas para los usuarios finales (Final users frontend/backend). Tanto la herramienta para los administradores como la herramienta para los usuarios finales son aplicaciones web organizadas en tres niveles: un frontend con la interfaz de usuario; un backend con la lógica de negocio y servicios de apoyo; y una base de datos.

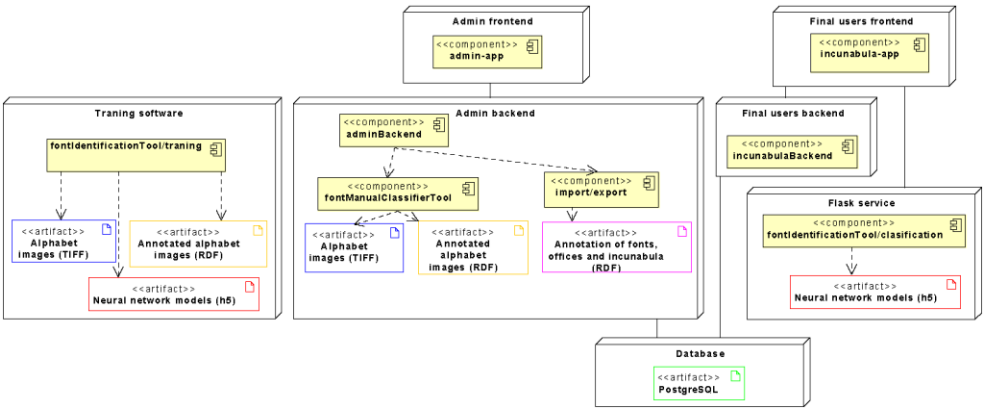


Figura 1. Diagrama de despliegue con la arquitectura del software. Fuente: Neji et al., 2025.

La herramienta Training Software se encarga de entrenar las redes neuronales para la detección de las letras “M” y las formas concretas de “M” tomando como entrada las imágenes del alfabeto de las distintas tipografías que se quieren reconocer y una anotación de la posición de la letra que se corresponde o no con la letra “M” en cada imagen. Para facilitar la anotación de las imágenes de cada alfabeto, la herramienta de administración proporciona una aplicación web donde que permite segmentar de forma automática los caracteres en las imágenes de los alfabetos (facilitados por el Typenrepertorium der Wiegendrucke para la mayoría de las tipografías), modificar manualmente esta segmentación si no ha sido correcta, y marcar las áreas correspondientes a letras especiales. La herramienta de administración también permite importar y exportar la información del Typenrepertorium der Wiegendrucke en formato RDF y almacenarla en una base de datos que será consultada posteriormente por las herramientas para los usuarios finales.

Figura 2. Interfaz de consulta de talleres de impresión de la herramienta para usuarios finales.

La herramienta orientada a los usuarios finales es una aplicación web para la catalogación de incunables que además de las típicas operaciones de un catálogo para el alta, baja, modificación y eliminación de registros, incorpora también funcionalidades para seleccionar una imagen representativa de una página de un incunable e invocar al software de reconocimiento de tipografías (nodo Flask service en la Figura 1). En la Figura 2 se puede ver un ejemplo de la interfaz de usuario de la aplicación para consultar los talleres de impresión disponibles en la base de datos y ubicados, por ejemplo, en Valencia. En la Figura 3 se puede observar un ejemplo de la de la interfaz de usuario para la edición de un incunable donde se ha invocado al software de reconocimiento y se muestran las posibles tipografías y talleres de impresión que se podrían asignar, en este caso el taller de Paul Hurus en Zaragoza y una tipografía gótica con altura de 104 mm para bloques de 20 líneas y la forma “17E” para la letra M.

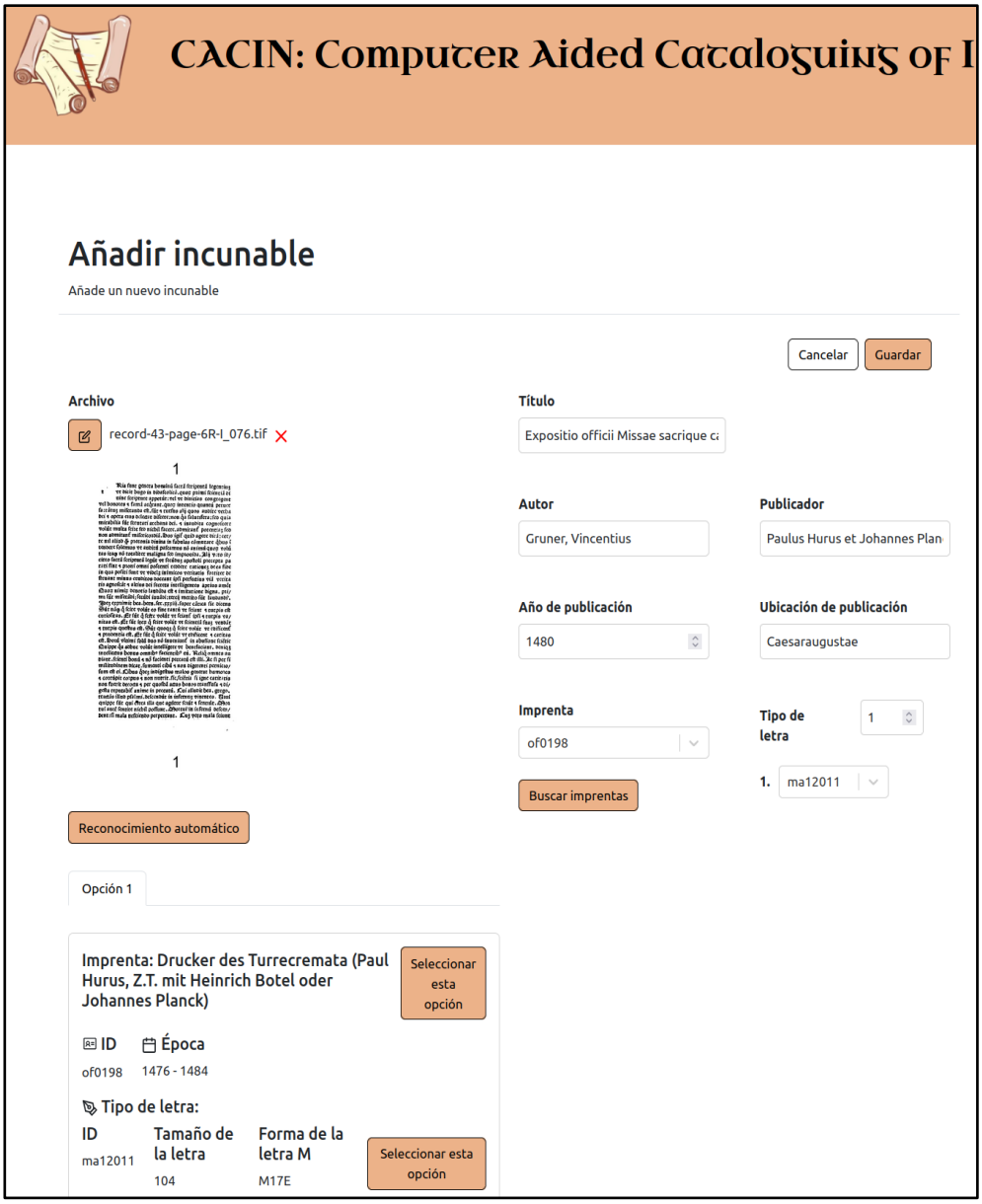


Figura 3. Interfaz de edición de un incunable en la herramienta para usuarios finales incluyendo la funcionalidad de reconocimiento automático de tipografías.

Para demostrar la funcionalidad ofrecida por CACIN a los expertos en incunables, se ha instalado la aplicación web de catalogación en la dirección https://cacin.iaaa.es/. La base de datos del catálogo se ha alimentado con la información de 1.626 talleres de impresión, 6.169 tipografías, y 577 incunables. En cuanto a las redes neuronales para la identificación de letras góticas especiales integradas en este despliegue de la aplicación web, estas redes han sido entrenadas para el reconocimiento de 107 tipografías góticas que utilizan 39 tipos distintos de formas de la letra “M”. En el futuro, se estudiará cómo conseguir que estas redes puedan reconocer un mayor número de tipografías, y escalar el sistema para el posible reconocimiento de las tipografías romanas utilizando la letra “Q” como letra especial.

Agradecimientos

Esta publicación es parte del proyecto de I+D+i T59_23R financiado por el Gobierno de Aragón.

Bibliografía

BARBIERI, E. (2008), *Haebler contro Haebler: appunti per una storia dell'incunabolistica novecentesca*, Milano, Università Cattolica del Sacro Cuore.

BRADSHAW, H. (1889), *Collected Papers*, Cambridge, Cambridge University Press.

HAEBLER, K. (1905–1924), *Typenrepertorium der Wiegendrucke*, Leipzig, Verlag von Rudolf Haupt, 5 vols. Disponible en línea: https://tw.staatsbibliothek-berlin.de/

LACASTA, J., NOGUERAS-ISO, J., ZARAZAGA-SORIA, F. J., & PEDRAZA-GRACIA, M. J. (2022), «Tracing the origins of incunabula through the automatic identification of fonts in digitised documents», *Multimedia Tools and Applications*, 81(28), pp. 40977–40991.

NEJI, H., NOGUERAS-ISO, J., LACASTA, J., & PEDRAZA-GRACIA, M. J. (2025), «CACIN: Computer Aided Cataloguing of Incunabula», preprint, SSRN. https://dx.doi.org/10.2139/ssrn.5201154

PROCTOR, R. (1905), *Bibliographical Essays*, introducción de A. W. Pollard, New York, Burt Franklin.

RIAL COSTAS, B. (2012), «El Sistema Proctor-Haebler y el estudio de las letrerías en las impresiones góticas incunables», en N. FERNÁNDEZ RODRÍGUEZ & M. FERNÁNDEZ FERREIRO (coords.), *Literatura medieval y renacentista en España: líneas y pautas*, Salamanca, Universidad de Salamanca / SEMYR, pp. 855–864. En línea: http://repositoriodigital-la-semyr.es/index.php/rd-ls/catalog/book/51

ZAPPELLA, G. (2001), *Il libro antico a stampa: struttura, tecniche, tipologie, evoluzione*, Milano, Editrice Bibliografica, 2 vols.

Jugabilidad sagrada: estructuras sacras y santas en el videojuego

Rubén Gregori Bou
Nadia McGowan
UNIR

Palabras clave: Religión; videojuegos; cultura digital; representación religiosa; santidad.
Keywords: Religion; video games; digital culture; religious representation; holiness.

Introducción

Hablar de lo santo es hablar, esencialmente, de lo divino o de aquello que participa de lo divino. En el contexto occidental, este concepto se ha desarrollado fundamentalmente dentro del marco del cristianismo, donde la santidad se entiende como una cualidad que procede de Dios y que puede ser comunicada a personas, objetos o lugares. Teológicamente, lo santo es aquello que es digno de veneración, respeto o culto por su participación en la santidad divina, así como por la rectitud de vida y por haberse obrado una conversión interior que abraza un camino ético y moral (Boesch-Gajano, 2020). No obstante, esta noción de santidad no es exclusiva del cristianismo, pudiendo encontrarla en otras religiones. El concepto suele articularse en torno a condiciones similares: una dimensión ético-moral como guía de comportamiento, una relación con lo divino o lo natural, y una búsqueda de armonía.

Lo sagrado, por su parte, es un concepto de carácter ontológico que se aplica a espacios, objetos o situaciones. Se distingue de lo profano al apartarse de lo común y cotidiano, adquiriendo un estatus especial por su vinculación con lo divino o con fuerzas trascendentes. A diferencia de lo santo, lo sagrado no implica necesariamente una connotación ética o moral, por lo que puede ser ambivalente: tanto fuente de bendición como de peligro, tabú o prohibición (Stausberg, 2017).

Los videojuegos, en tanto que productos artísticos, están condicionados por diversos factores, como el contexto social, el cultural, el económico o el religioso, entre otros. No es extraño pues, que conceptos como lo santo o lo sagrado se puedan encontrar integrados en la construcción del relato narrativo y en las propuestas estéticas de un videojuego (Schulzke, 2020).

Con esto en mente, el objetivo principal de este texto será el análisis de la presencia de lo religioso, en particular de las nociones de lo santo y lo sagrado

cristianos, en los videojuegos. Para ello, exploraremos cómo estos conceptos se materializan en elementos narrativos y estéticos dentro del medio videolúdico. En segundo lugar, se identificarán ejemplos concretos de videojuegos que incorporen estas dimensiones de manera significativa. Por último, se reflexionará sobre cómo estas representaciones interactúan con los contextos culturales y religiosos en los que se producen y consumen, considerando tanto su potencial para reforzar imaginarios tradicionales como para subvertirlos o reinterpretarlos.

El trabajo adoptará un enfoque cualitativo, basado en el análisis de contenido y la hermenéutica cultural, lo que permitirá una interpretación profunda de los fenómenos estudiados. La metodología se estructurará en dos ejes principales. El primero será el estudio de caso, que implicará la selección y análisis de cuatro videojuegos que ejemplifiquen esa presencia de lo religioso. Por ello, se tendrán en cuenta elementos narrativos inspirados en el plano religioso y personajes que adopten roles en línea con lo santo o con sistemas de moralidad que reflejen dilemas espirituales. Asimismo, se analizarán aspectos estéticos, como el diseño de escenarios que evocan espacios sagrados, el uso de símbolos religiosos y la creación de atmósferas que transmitan trascendencia o misticismo.

El segundo eje se centrará en la contextualización cultural, atendiendo a los marcos sociales, históricos y religiosos que influyen en estas representaciones. Esto incluirá considerar cómo los videojuegos reflejan o reinterpretan tradiciones espirituales específicas, así como explorar las posibles tensiones entre la apropiación de lo religioso con fines lúdicos, el respeto por sus significados originales y los límites éticos. El análisis digital de estas representaciones religiosas en los videojuegos permitirá comprobar cómo se pueden construir nociones teológicas, como lo transcendental, en el plano lúdico. Este enfoque multidimensional permitirá comprender no solo cómo lo santo y lo sagrado se integran en los videojuegos, sino también qué dicen estas representaciones sobre la relación entre religión, cultura y tecnología en el mundo contemporáneo.

I am Jesus Christ: en la piel del Mesías

I am Jesus Christ (Space Boat Studios, s.f.), en el momento de escribir el presente texto, aún no se ha lanzado al mercado. A través de una experiencia inmersiva, el simulador nos presenta la posibilidad de poder controlar, en primera persona, a Jesús, pudiendo obrar toda suerte de milagros y vivir su vida, incluyendo el ciclo de la pasión y la posterior resurrección. Según el anuncio del propio juego, se podrá interactuar con los personajes neotestamentarios, incluyendo a María y los apóstoles, recrear episodios como la Última Cena o la crucifixión y explorar enclaves emblemáticos de Tierra Santa.

El juego también permite obrar milagros o prodigios extraídos del relato bíblico. Por ejemplo, cuando Jesús se encuentra ayunando 40 días en el desierto, nos enfrentamos a Satanás y debemos repeler sus ráfagas con nuestro

poder divino. De igual modo, durante las bodas de Caná, a través de entrar en el Holy Spirt Mode, seremos capaces de convertir el agua en vino. Además, en los trailers vemos a Jesús apartando la roca de un sepulcro y resucitando a Lázaro o caminando sobre las aguas.

Sainthood: estructura de santidad en clave lúdica

Sainthood (Taiwo, 2024) es un videojuego independiente en el que el jugador toma el papel de una monja en una sociedad ficticia premoderna. La mecánica principal del juego es la gestión de recursos espirituales, que combina decisiones ético-morales con eventos narrativos que simulan pruebas de fe y virtud con una duración de 33 días en juego. Esta duración en el mundo diegético del juego constituye un claro referente simbólico, establece un horizonte temporal simbólico e introduce un marco estructurado para la progresión espiritual del jugador. La puntuación y el avance hacia la santidad se organizan en torno a dos métricas: los puntos de caridad (Charity Points, CP) y los puntos de fe (Faith Points, FP). Los puntos de fe se consiguen mediante prácticas devocionales, como el rezo o participando en actividades en el convento. Los puntos de caridad se adquieren a través de la materialización de la vocación de servicio mediante trabajos en la comunidad de juego.

Cult of the Lamb: parodia ritual posmoderna

Cult of the Lamb (Massive Monster, 2022) fusiona la estética kawaii con imágenes caricaturescas, configurando una parodia lúdica de lo sagrado y lo profano. El jugador encarna al agnus dei, un cordero resucitado por una deidad olvidada. La mecánica combina la gestión de recursos con la exploración de mazmorras como parte de un ciclo ritual de sacrificios y generación de poder. El juego introduce estructuras religiosas reconocibles, como templos, misas y dogmas, utilizadas como herramientas de dominación simbólica. La estética refuerza la sátira, en la que la liturgia deviene espectáculo y la devoción en un mecanismo de producción de recursos y de control de la población. Se trataría de una crítica posmoderna a las instituciones religiosas en general y a la cristiana en particular.

Blasphemous: de milagros y penitentes

Tomando como referencia el mundo católico y el folclore del sur de España, *Blasphemous* (The Game Kitchen, 2019a) presenta una aventura original con una atmosfera oscura y unos escenarios sombríos repletos de iconografía cristiana (Anyó Sanyol, 2023). La historia se sitúa en Custodia, una tierra bajo el control de El Milagro, una manifestación divina que se presenta como un castigo redentor que transforma la culpa en prodigios físicos como formas de expiación. El protagonista es el Penitente, un guerrero silencioso portador de la espada Mea Culpa y de un capirote metálico con espinas inspirado en los sambenitos de la Inquisición. Tanto a nivel narrativo como iconográfico, el sentimiento de culpa

cristiana está integrado. En el mismo inicio del juego comprobamos los efectos de esta carga con la mujer que se golpea el pecho con la empuñadura de Mea Culpa mientras entona una oración penitencial. Esta referenciaría el Confíteor, un rezo de confesión que proclama el "por mi culpa, por mi culpa, por mi gran culpa".

El juego está ambientado en múltiples ejemplos inspirados en la iconografía cristiana: El Retorcido se basa en las imágenes de los ladrones crucificados junto a Jesús; la Señora de los Seis dolores, atravesada por siete espadas, toma como referencia *Vuelo de Brujas* (1797) de Francisco de Goya y la Virgen de los Dolores (The Game Kitchen, 2019b). En *Blasphemous 2* (2023) encontramos a Anunciada, que recuerda a la *Inmaculada Concepción* de Rubens (1628-1629) y a las representaciones de la asunción. Respecto a los escenarios, se ambienta en lugares y arquitecturas existentes, como la Catedral de Cádiz que sirve de referente para la Catedral Hundida (The Game Kitchen, 2024).

Conclusiones

El análisis realizado demuestra que las nociones de lo santo y lo sagrado están presentes en los videojuegos y pueden adoptar múltiples formas. Estas quedan integradas a través de narrativas, mecánicas y estéticas que basculan entre la devoción, la reinterpretación crítica o la sátira. Por ello, además de reflejarse tradiciones religiosas, estas se reinventan, cuestionan e incluso subvierten, brindando una mirada única y transversal sobre la relación entre cultura, fe y videojuego.

Los dos primeros casos ejemplifican el enfoque respetuoso y contemplativo, mientras que los dos últimos apuestan por una visión transgresora en la que se parodian elementos derivados de la religión. Además, los cuatro títulos se revisten de una sucesión de símbolos e iconografías que evidencian que lo santo y sagrado están representados, reinterpretados y redefinidos.

Bibliografía

ANYÓ SANYOL, L. (2023), «Videojuego indie y estética local: La singular representación de la cultura española en *Blasphemous*», en *Actas del 1r Simposi sobre Investigació a la Indústria Indie del Videojoc (I3V 2023) (Terrassa, 27 octubre 2023)*, UPCommons.

BOESCH-GAJANO, S. (2020), «La santidad como paradigma histórico», *Anuario de Historia de la Iglesia*, 29, pp. 19–52.

MASSIVE MONSTER (2022), *Cult of the Lamb* [videojuego], Devolver Digital.

SCHULZKE, M. (2020), *Simulating Good and Evil: The Morality and Politics of Videogames*, Rutgers University Press.

SPACE BOAT STUDIOS (s. f.), *I am Jesus Crist* [videojuego], PlayWay.

STAUSBERG, M. (2017), «The sacred and the holy – from around 1917 to today», *Religion*, 47(4), pp. 549–556.

TAIWO, B. (2024), *Sainthood* [videojuego], Taiwo Pictures Inc.
THE GAME KITCHEN (2019a), *Blasphemous* [videojuego], Team17.
THE GAME KITCHEN (2019b), *The Art of Blasphemous*, GTM Ediciones.
THE GAME KITCHEN (2023), *Blasphemous 2* [videojuego], Team17.
THE GAME KITCHEN (2024), *The Art of Blasphemous II*, Game Press.

Imaginarios de la marginalidad urbana contemporánea en Chile: una aproximación interdisciplinaria al caso de la música trap y su difusión en videoclips a través de la plataforma Youtube

Rodrigo Ganter Solís
Universidad de Concepción
Angela Rivera Martínez
Universidad Autónoma Metropolitana – Xochimilco Escuela Nacional de Antropología e Historia

Palabras clave: Marginalidad urbana; música trap; imaginarios juveniles; videoclips; cultura digital.
Keywords: Urban marginalization; trap music; youth imaginaries; music videos; digital culture.

El eje sobre el cual gravita la propuesta está centrado en los imaginarios sociales que construyen los artistas y cultores de la música urbana en Chile en torno a la marginalidad urbana contemporánea. Ello a partir de la elaboración de sus videoclips musicales, donde se condensa más expresivamente su universo simbólico e imaginarios sociales, los cuales difunden en la plataforma Youtube y en sus redes digitales como una forma de reconocimiento social que articula diferentes nociones “en sus propios términos” sobre lo periférico. Debido al carácter del material con el cual se trabaja, el problema de investigación y la confluencia de distintos saberes y dimensiones implicadas en el estudio, la propuesta se sustenta en una perspectiva interdisciplinaria (musicología, estudios visuales y digitales, comunicología, sociología, antropología, entre otros), situándose en los bordes de la academia convencional y al interior de una cartografía de saberes diversos, en diálogo simétrico y cooperación permanente. Por lo cual, proponemos un enfoque híbrido de investigación, lo que permitirá abordar epistemológica y metodológicamente tanto los niveles narrativos y contextuales, como los visuales, digitales, musicales y simbólicos implicados en este tipo de formatos audiovisuales. Emulando a J. Martín-Barbero, se propone menos un objeto de investigación y más unas encrucijadas de investigación. Estas encrucijadas no sólo exigen interdisciplinariedad, sino también arriesgarse a salir de los objetos prestigiosos y de moda para ponerse a la escucha de lo que está pasando en nuestras sociedades. Actualmente la propuesta cuenta con financiamiento de la Agencia Nacional Científica y Tecnológica de Chile (ANID)

para 3 años 2024-2026.

Algunos de los **objetivos** que se plantea la propuesta:

Analizar interdisciplinariamente y de modo integrado los principales imaginarios sociales que se construyen sobre la música urbana y la marginalidad urbana, integrando los niveles contextuales, narrativos, musicales, digitales y visuales del fenómeno en cuestión.

La **metodología de investigación** combina la etnografía digital con el análisis interdisciplinario y el diálogo de saberes, con énfasis en lo *cualitativo,* además asumimos esta perspectiva como un bricolage (un mosaico de piezas unidas armónicamente) y al investigador como bricoleur (Denzin y Lincoln, 2011), Se consideran *4 niveles de indagación y análisis*: narrativo, musical, visual y contextual; abarcando el período 2019- 2024, asociado con los acontecimientos más importantes a nivel país y de producción musical, como el denominado “estallido social”, los años de pandemia y posteriormente la masificación del estilo trap en Chile, considerando la escena metropolitana (el gran Santiago) y la escena provincial (el gran Concepción), por ser las dos concentraciones urbanas más importantes del país.

La principal **contribución y novedad de la propuesta** se traduce en la generación de conocimiento interdisciplinario y colaborativo en torno a la construcción social de la marginalidad urbana contemporánea y las dinámicas de discriminación implicadas en este proceso, tomando como punto de partida las producciones audiovisuales, específicamente los videoclips y el universo simbólico que elaboran los propios artistas y cultores de la música trap en Chile, por medio de los cuales es posible acceder a sus imaginarios sociales, y comprender cómo se construye y significa desde allí la marginalidad urbana contemporánea, la cultura de la calle*, lo flaite*, en tanto forma de reconocimiento social desde lo periférico (Caldeira, 2007, Reguillo, 1991), una forma de visibilización de las "existencias inexistentes" (Ranciére, 1996). Generando a través de ello pistas significativas que aporten al debate público (Burawoy, 2005) sobre las *principales mutaciones que ha experimentado esta condición social, sus componentes identitarios y de género (Butler, 2007), sus vehículos de expresión y comunicación, sus ritos de consumo, su narrativa, su épica, sus signos y estética, aspiraciones, creencias, su concepción del mérito y el esfuerzo, sus concepciones de la violencia, su crítica social, sus formas de cohesión social, sus formas de ser y hacer, sus imágenes de futuro.*

El trap chileno y su divulgación en videoclips

El trap debe ser entendido como práctica musical en estrecha vinculación con el desarrollo tecnológico, particularmente en la figura de los *home studios*, donde los músicos pueden generar bases musicales, producir y grabar sonido en sus propios hogares sin tener que recurrir a intermediarios (Muñoz, 2018). Estos jóvenes músicos han aprovechado un paradigma contemporáneo de la música popular marcado por las posibilidades que otorga la autogestión musical

a nivel virtual (Wade, 2015). En un panorama previo donde existía una hegemonía productiva que estaba en manos de los sellos discográficos —sobre todo multinacionales—, resultaba complejo adquirir reconocimiento en el mundo de la música desde un trabajo individual e independiente (Arriagada e Ibáñez, 2019). Este giro ha estado determinado en buena medida por la aparición de la *web 2.0* (espacios y herramientas virtuales que se utilizan principalmente mediante dispositivos portátiles como *smartphones*), pues ha permitido una circulación masiva del trabajo musical de los traperos en Chile sin que estos tengan que recurrir a mayores intermediarios. En tanto, el desarrollo de distintas tecnologías de manipulación del sonido también ha facilitado una práctica musical autogestionada. La emergencia de los llamados *home studios (*ha resultado fundamental para la producción musical de artistas independientes. Así, hoy en día es habitual que los músicos tengan en su mismo hogar distintos artefactos tecnológicos que permiten tanto producir como grabar música, haciendo que la práctica musical sea más factible (Wade, 2015).

En lo que respecta a la cultura del videoclip, Weibel (1987) la plantea como el tipo de cultura que mejor representa nuestro mundo actual, de ahí que el formato del videoclip tenga el mérito de articular tecnología, marketing, arte, cine y otras posibles esferas de producción y circulación de saberes, por lo cual se trata de un género híbrido (Leguizamón, 1998 y 2001). Sánchez (en Sedeño, 2006) plantea al videoclip como “un documento antropológico” y por ello se transforma en un producto importante de estudiar tanto en la complejidad de su composición como en su elaboración técnica, pero también como narrativa. El videoclip musical actual es imposible ubicarlo fuera de las redes de internet que configuran entorno y contornos cada vez menos definidos, llegan a plantearse un tipo de videoclip pos televisivo (Sedeño, et. al 2016).

En la investigación sobre videoclips se analiza este formato y su modo de comunicación y expresión estética desde la teoría postmoderna (Kaplan, 1987, Frith, 1988, Jameson, 1991) y lo definen como un formato matriz de la estética audiovisual contemporánea (Calabrese, 1989, Fiske, 1987). Asimismo, se observa un predominio del análisis de sus imágenes como medio de difusión de contenidos sexuales, violencia urbana, consumo de sustancias, desenfreno asociado a ciertas subculturas juveniles, etc. (Sedeño, Rodríguez, Roger, 2016), destacando el trabajo de Goodwin (1992), quien propuso algunas características del formato, como su carácter intertextual, donde se mezclan contenidos y referencias a las series, el cine, la radio, los noticieros de tv, la música de otras épocas, otros vídeos y sonidos urbanos. También se reconoce su cualidad para integrar los componentes visuales y narrativos del clip, con el género musical al que pertenece, y a su vez con un imaginario cultural y estético de época.

Sedeño refiere que “el aspecto más interesante en la relación videoclip musical e internet se centra en que la tradicional división entre emisores y receptores se difumina, con la facilidad para integrar o subir contenidos de vídeos en las webs” (p.7). Actualmente las herramientas de alto grado de tecnicismo para la producción de un videoclip pueden ser reemplazadas por el uso de un

celular, conservando las particularidades y distinciones que existen entre ambos, pero reconociendo que internet es el punto de encuentro como un medio de distribución de alta eficacia (Sedeño, 2010). Debido a que probablemente es el formato más consumido en redes sociales (gracias a plataformas de video como YouTube, donde la publicidad y todo tipo de contenido se metamorfosean para llegar a audiencias diferenciadas) ha conformado una especie de formato-tipo, que es retomado –o plagiado– en duración y presentación de materias audiovisuales por otros formatos publicitarios como el *fashion-film*, el *branded-content* o la video-noticia" (p.140).

En esta fase preliminar del análisis y resultados, hemos podido observar que los videoclips de trap chileno no sólo operan como productos musicales, sino como artefactos culturales complejos donde se condensa un lenguaje visual y simbólico que configura una narrativa sobre la marginalidad urbana experimentada desde la condición juvenil contemporánea.

A través de la producción de códigos propios —como la relación con espacio público, la exaltación de ciertos consumos y la performance de cuerpos racializados y generizados—, estos materiales configuran imaginarios de lo periférico que tensiona los marcos hegemónicos de representación con los cuales las juventudes se han confrontado históricamente en el caso chileno. Este universo simbólico no busca mimetizarse con la cultura dominante, sino que produce sus propios términos de visibilidad y reconocimiento, éxito y satisfacción. De igual modo, el anclaje cultural que representa el neoliberalismo para los traperos tiene formas simbólicas y discursivas en que son expresadas, generando tensiones entre lo local y lo global.

Si bien el análisis sigue en curso, se hace evidente cómo estos videoclips activan un relato sobre la exclusión, sobre la construcción de comunidad, la sobrevivencia y las aspiraciones en clave de clase, que puede leerse como una crítica (en diferentes tonos) sobre las condiciones de vida de sectores históricamente marginados. Todas las estas expresiones, que hemos podido rastrear hasta esta fase inicial, configuran un espacio discursivo que produce códigos, imaginarios y representaciones, en el que se ensamblan relatos de estigmatización con aspiraciones de movilidad social y códigos estéticos globales.

Bibliografía

RIQUELME, S., SULE, V., CASTILLO, V. y NÚÑEZ, V. (2022). La periferia urbana a través de los lentes del trap chileno: narrativas juveniles. *Última década, 30*(59), 140–172. http://dx.doi.org/10.4067/S0718-22362022000200140

SEDEÑO, A., RODRÍGUEZ, J. y ROGER, S. (2016). El videoclip postelevisivo actual. Propuesta metodológica y análisis estético. *Revista Latina de Comunicación Social, 71*, 332–348. https://doi.org/10.4185/RLCS-2016-1098

SEDEÑO, A. y GUARINOS, V. (2021). El cuerpo audiovisualizado en la realización del vídeo musical contemporáneo. *Atenea, 523*, 139–157. https://doi.org/10.29393/At523-414ASCA20414

Inteligencias artificiales: ¿una reinvención del historiador como profesional? Menos clics, más inteligencia histórica

Chimo Soler Herreros
Asociación Proyecto Clío
Antonio Carrasco Rodríguez
Universidad de Alicante

Palabras clave: Inteligencia artificial generativa; profesión del historiador; historia; enseñanza de la historia; innovación docente.
Keywords: Generative artificial intelligence; historian profession; history; teaching history; teaching innovation.

Descripción del problema de investigación

El presente trabajo aborda la crisis y la oportunidad que la irrupción de la Inteligencia Artificial Generativa (IAG) plantea para la disciplina histórica y la profesión del historiador. El problema de investigación no se origina en la tecnología *per se*, sino en la controversia pública y el debate académico subsiguiente, catalizados por un influyente pero metodológicamente cuestionable estudio de Microsoft Research publicado el 10 de julio de 2025. Dicho estudio situaba a los historiadores como la segunda profesión con mayor "aplicabilidad" de la IAG, solo por detrás de traductores e intérpretes. Este dato, magnificado por medios de comunicación de alcance global como *The New York Times* o *Fortune*, provocó una onda expansiva que disparó el interés público por un posible "apocalipsis laboral" para los historiadores.

El problema se desglosa en varias capas. En primer lugar, se constata una distorsión del debate público, donde la narrativa se bifurcó rápidamente entre visiones optimistas de una simple "ampliación" de capacidades y pronósticos alarmistas sobre la "sustitución" de roles profesionales, especialmente en el ámbito de los archivos. Las aclaraciones de los propios autores del estudio, quienes insistieron en que su trabajo medía la utilidad potencial ("aplicabilidad") y no el riesgo de reemplazo, tuvieron un eco limitado en la esfera pública, que continuó amplificando la confusión. Esta confusión llegó a extremos en los que se atribuían las conclusiones del estudio sobre Copilot de Microsoft a otras tecnologías como ChatGPT.

En segundo lugar, y como núcleo del problema, el manuscrito identifica la debilidad metodológica fundamental del estudio que originó la controversia. El análisis crítico revela que el debate se ha construido sobre cimientos poco

sólidos. La investigación se basó en un producto, Copilot, que apenas contaba con unas semanas de vida oficial, en un ecosistema tecnológico inmaduro caracterizado por ciclos de vida muy cortos, actualizaciones constantes y análisis de uso muy preliminares. Las vulnerabilidades del estudio son significativas y se articulan en tres áreas principales:

1. Métricas de calidad inadecuadas: La metodología consideraba una capacidad profesional como "satisfecha" con que la IAG la hubiera procesado correctamente una sola vez, basando además uno de sus indicadores clave en la simple satisfacción subjetiva del usuario, marcada con un "pulgar hacia arriba".
2. Ignorancia de las alucinaciones IA: El estudio no ponderó el impacto de las "alucinaciones" o desaciertos de la IAG, que pueden alcanzar tasas del 30%, un problema estructural reconocido por los propios líderes de la industria como Sam Altman. Se priorizó la capacidad de generar una respuesta sobre la fiabilidad científica de la misma.
3. Obsolescencia de la fuente: La base de datos de ocupaciones utilizada (ONET 2024) quedó desactualizada rápidamente. Una nueva versión de septiembre de 2025 presentaba un 35 % de diferencias en las tareas atribuidas a los historiadores, lo que significa que el estudio midió la aplicabilidad de la IA a un perfil profesional parcialmente inexistente en el momento de la publicación del debate.

Por tanto, el problema de investigación que este texto aborda no es si la IAG reemplazará a los historiadores —una pregunta que se revela como un falso dilema basado en datos débiles—, sino cómo esta controversia obliga a la disciplina a realizar una profunda reflexión interna para redefinir su núcleo de valor, sus prácticas y su identidad en un nuevo contexto tecnológico. La investigación se centra en superar la reacción superficial para proponer una reinvención consciente y estratégica del oficio.

Metodología aplicada

La metodología empleada en este trabajo es de naturaleza cualitativa, crítica y propositiva. No se trata de un estudio empírico con datos cuantitativos, sino de un ensayo analítico que deconstruye un problema y construye un nuevo marco conceptual para la disciplina histórica. El procedimiento metodológico se puede estructurar en las siguientes fases:

Fase 1: Deconstrucción crítica. La primera parte de la metodología consiste en un análisis crítico de las fuentes del debate. Se examina el estudio de Microsoft y la cobertura mediática para "abrir la caja negra" y exponer sus sesgos y limitaciones. Se identifican tres cautelas metodológicas clave que contaminan la discusión general: la "definición gruesa de capacidades", que aplana habilidades complejas como "escribir"; las "métricas de calidad inadecuadas", que confunden legibilidad con veracidad; y el "probabilismo sin trazabilidad", que

ignora que los modelos generan secuencias plausibles, no razonamientos basados en evidencia.

Fase 2: Elaboración de un nuevo marco conceptual. A partir de la crítica, se propone un nuevo vocabulario y un marco teórico para reorientar el debate. Se establece una distinción fundamental entre "exposición" (el grado en que las tareas pueden ser asistidas por la IAG) y "sustituibilidad" (el reemplazo del profesional), argumentando que el desliz entre ambos conceptos es la principal fuente de error en el debate. El núcleo de esta fase es la formulación del concepto central del trabajo: el "momento auditoría". Este marco postula que la IAG acelera un cambio de régimen en el que el valor del historiador se desplaza desde la producción textual (ahora abundante y barata) hacia la trazabilidad, la evaluación de evidencia y la construcción de explicaciones robustas. La reinvención, por tanto, no es tecnológica, sino ética e intelectual.

Fase 3: Identificación de competencias nucleares. La metodología se vuelve constructiva al identificar y definir el "núcleo de valor" del historiador. A través de un proceso de abstracción y análisis de la práctica historiográfica, se explicitan seis competencias profesionales que se consideran indelegables a la máquina. Para hacer esta definición operativa y comprensible, se utiliza un recurso metodológico-didáctico: las "mini-escenas", que son breves viñetas que ilustran cada competencia en un escenario de uso práctico, contrastando la aportación de la IA con el juicio del historiador.

Fase 4: Desarrollo de aplicaciones prácticas. La última fase de la metodología traduce el marco teórico en herramientas y modelos aplicables. Se diseña un "mapa operativo" que clasifica las tareas en delegables/automatizables e indelegables, y propone un "flujo asistido estándar" para proyectos de investigación. Asimismo, se desarrolla un "diseño de aula" específico para la didáctica de la historia, enfocado en el desarrollo del pensamiento crítico frente a la herramienta. Finalmente, se sintetiza un "marco realista" para la gobernanza de la tecnología, que incluye la gestión de riesgos, una política de uso responsable y plantillas prácticas para garantizar la transparencia.

En resumen, la metodología combina el análisis crítico-discursivo, la conceptualización teórica y el diseño de modelos prácticos para ofrecer una respuesta integral al problema de investigación.

Resultados obtenidos

Los resultados de esta investigación no son hallazgos empíricos, sino un conjunto de conclusiones, modelos y propuestas que constituyen una hoja de ruta para la reinvención del oficio del historiador.

Resultado 1: Un nuevo marco de valor para el historiador. El resultado principal es la definición del "historiador auditor" como el perfil profesional resiliente a la era digital. Este historiador se define por seis competencias críticas, indelegables y perennes:

1. Arquitectura de preguntas: La capacidad de formular problemas históricos fértiles, situados y con hipótesis refutables, algo que va más allá de la simple identificación de patrones que puede realizar una IA.
2. Crítica interna y externa de fuentes: El juicio filológico y archivístico para establecer la autoridad documental, superando la plausibilidad estadística de un modelo.
3. Contextualización temporal y espacial: La habilidad para periodizar, manejar escalas y evitar anacronismos, frenando las analogías superficiales que la IAG puede proponer.
4. Construcción causal e interpretativa: La jerarquización de mecanismos y la articulación de secuencias causales basadas en evidencia, en contraste con el listado de posibles causas de una IA.
5. Ética y responsabilidad pública: Las decisiones deliberativas sobre el impacto de la investigación en las comunidades, la privacidad y los sesgos, un ámbito puramente humano.
6. Artesanía narrativa: La capacidad de integrar evidencias en relatos que enseñen a pensar históricamente, mostrando incertidumbres y procesos de verificación, no solo resultados "limpios".

Se concluye que estas competencias sostienen la "autoría historiográfica" y que la habilidad para detectar una alucinación de la IA se convierte en una nueva competencia crítica fundamental.

Resultado 2: Un mapa operativo para la práctica investigadora. Se ofrece un modelo práctico para integrar la IAG en el flujo de trabajo sin comprometer el rigor. Se distinguen claramente las tareas delegables (cribado inicial de corpus, extracción de metadatos, borradores de estados de la cuestión) de las indelegables (selección del corpus, crítica de fuentes, diseño metodológico, escritura final responsable). Este mapa permite a los equipos de investigación aprovechar la eficiencia de la IA en tareas instrumentales para liberar tiempo y recursos que pueden ser reinvertidos en el núcleo de valor profesional.

Resultado 3: Un modelo pedagógico para la “Era IA”. Se propone un diseño de aula para la enseñanza de la Historia que convierte a la IAG de una potencial herramienta de plagio en un objeto de aprendizaje. El modelo se desplaza de la evaluación del producto final (el texto) a la evaluación del proceso: la calidad de las preguntas formuladas, las rutas de verificación empleadas, el uso crítico de las fuentes y la claridad argumentativa. Se detallan los pasos de una actividad tipo, desde la formulación del problema hasta la metacognición sobre los sesgos del modelo.

Resultado 4: Un marco de gobernanza disciplinar. Finalmente, se presenta un marco ético y práctico para que la disciplina "dome" la tecnología en lugar de ser arrollada por ella. Este marco incluye la gestión de riesgos clave como las alucinaciones, los sesgos de entrenamiento y la dependencia de plataformas. Propone una política de uso responsable basada en la declaración de uso, el mantenimiento de un *audit trail* y la verificación cruzada humana. Se ofrecen,

además, plantillas prácticas como una "cláusula de transparencia" para ser incluida en las publicaciones.

El resultado global de la investigación es una tesis robusta y multifacética: la IAG no supone el fin del historiador, sino una interpelación directa que exige una reinvención. Esta reinvención pasa por abandonar el énfasis en la producción textual y recentrar el valor profesional en la inteligencia histórica, la crítica, la ética y la transparencia, proponiendo para ello un nuevo marco conceptual, el "momento auditoría", y las herramientas prácticas para llevarlo a cabo.

Bibliografía

AMERICAN HISTORICAL ASSOCIATION (2023, August). *AHA Council Statement on AI and History*.

BENDER, E. M., GEBRU, T., McMILLAN-MAJOR, A., y SHMITCHELL, S. (2021). On the dangers of stochastic parrots: Can language models be too big? En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). https://doi.org/10.1145/3442188.3445922

BERRY, D. (2016). *The philosophy of software: Code and mediation in the digital age*. Springer.

CARRASCO-RODRÍGUEZ, A. (2023). Reinventando la enseñanza de la Historia Moderna en Secundaria: La utilización de ChatGPT para potenciar el aprendizaje y la innovación docente. *Studia Historica. Historia Moderna, 45*(1), 101–145. https://doi.org/10.14201/shhmo2023451101146

CARRASCO-RODRÍGUEZ, A. (2024). Perceptions of generative artificial intelligence among university Early Modern History students. *Tiempos Modernos, 14*(49), 269–285.

COHEN, D. J., y ROSENZWEIG, R. (2006). *Digital history: A guide to gathering, preserving, and presenting the past on the web*. University of Pennsylvania Press.

GULDI, J., y ARMITAGE, D. (2014). *The History Manifesto*. Cambridge University Press.

TIRADO-OLIVARES, S., NAVÍO-INGLÉS, M., O'CONNOR-JIMÉNEZ, P., y CÓZAR-GUTIÉRREZ, R. (2023). From human to machine: Investigating the effectiveness of the conversational AI ChatGPT in historical thinking. *Education Sciences, 13*(8), 803. https://doi.org/10.3390/educsci13080803

El diseño del Archivo Digital Sebastià Juan Arbó: el empleo del modelo no relacional para la construcción de la base de datos bibliográfica

Joan Antoni Forcadell
Universitat Autònoma de Barcelona – GEXEL-CEDID
Joel Reverter Lainez
Universitat Autònoma de Barcelona – PROLOPE

Palabras clave: Sebastià Juan Arbó; archivo digital; base de datos; modelo no relacional; MongoDB
Keywords: Sebastià Juan Arbó; digital archive; database; non-relational model; MongoDB

Esta comunicación tiene como objetivo difundir los resultados de la primera fase de desarrollo del Archivo Digital Sebastià Juan Arbó, centrada en la elaboración de un prototipo de base de datos que reúne la bibliografía completa sobre el autor, con más de 8.500 referencias en su mayoría inéditas o escasamente conocidas. Aunque constituye solo una sección acotada dentro de lo que será la versión final de la web, esta fase ha permitido comprobar su integración y funcionamiento dentro de la estructura global del proyecto, además de hacer aflorar los principales problemas y retos que podrían surgir en las fases de desarrollo posteriores.

Sebastià Juan Arbó (la Ràpita, 1902 – Barcelona, 1984) fue una de las plumas más genuinas y representativas de la literatura española y catalana del siglo XX. Más allá de la relevancia del escritor en su época, una serie de condicionantes de base sociohistórica y política lo han arrinconado en el umbral de la marginalidad literaria, siendo todavía escasos los esfuerzos empleados para rescatar académicamente su figura y obra.

Uno de los condicionantes a los que se ha enfrentado la investigación al respecto ha sido la dificultad de localizar y recopilar evidencias documentales en fondos archivísticos, bibliotecas, centros de documentación y medios escritos publicados, una labor heurística que a menudo se ha afrontado de forma parcial —son intentos destacados, pero incompletos, las tesis doctorales de Ramis (2011) y Matas (2021)—. Esto ha contribuido a que haya quedado en la sombra su intensa actividad de creación literaria y periodística, así como su notable proyección pública como intelectual y figura destacada en el panorama cultural de los dos sistemas literarios de los que participó.

Rescatar del olvido a un escritor implica, junto con la edición, actualización y difusión de su obra, una tarea académica de cuidado sobre su legado para hacerlo accesible a lectores, docentes e investigadores. El proyecto de creación del Archivo Digital Sebastià Juan Arbó se plantea revertir definitivamente esta situación para poner a disposición pública en un portal web el conocimiento de toda evidencia material sobre cualquier faceta relacionada con el escritor, en la forma de una herramienta filológica útil para abastecer cualquier investigación relacionada con la vida y la obra del escritor. Facilitará el acceso en línea a un corpus gráfico, iconográfico, textual y bibliográfico integral reunido, clasificado, tratado informáticamente y disponible en abierto. Es el producto de una década de acondicionamiento, tratamiento documental, recopilación e investigación personales sobre el patrimonio documental del escritor en todas las instituciones que conservan vestigios archivísticos y hemerográficos del escritor, así como en cualquier medio de difusión pública en que estuvo presente.

Para abordar la primera fase de desarrollo, centrada en la presentación de una base de datos con la bibliografía completa de Sebastià Juan Arbó, se intentó en un primer momento replicar la metodología utilizada en diversos archivos digitales exitosos de características análogas, empleando tecnología SQL. Mediante PostgreSQL, se trazó un esquema de relaciones entre los diversos elementos que conformarían la base de datos.

Pronto, el esquema reveló una gran dificultad: clasificar un fondo tan vasto y variado como el que se ha recopilado resultaba problemático. Los registros eran tan diversos y heterogéneos que requerían características de catalogación distintas para poder almacenarse adecuadamente en una misma base de datos. Construir esta estructura en un modelo relacional implicaba diseñar un sistema complejo de tablas que pudiera representar las intrincadas relaciones entre los distintos elementos. Este problema, inicialmente referido al registro bibliográfico, se proyecta sobre los elementos de naturaleza heterogénea que eventualmente integrarán el archivo en su versión definitiva.

La adopción de una base de datos no relacional permitió superar estas dificultades y facilitó el trabajo a un equipo de investigación todavía reducido. Este modelo admite que cada registro tenga sus propios campos, ofreciendo una flexibilidad ideal para el proyecto. No obstante, esta misma flexibilidad puede derivar en errores por la ausencia de un esquema de datos fijo. El lenguaje empleado, MongoDB, permite integrar esquemas en las colecciones, exigiendo que cada registro contenga ciertos campos determinados. Si bien esto es útil para campos concretos —como el título del registro—, no resulta eficaz en nuestro caso, por la misma razón que descartamos el modelo relacional: la heterogeneidad de los datos dificulta la definición de un esquema sin sacrificar parte del rendimiento de la aplicación. Además, algunos registros acusan vacíos insalvables en ciertos campos, debido a la falta de información en origen.

Para preservar la integridad de los datos, se optó por introducir los registros inicialmente a través del gestor bibliográfico Mendeley. Esta herramienta permite trabajar con una interfaz intuitiva, operar en grupo, minimizar errores derivados

de la introducción manual —inevitable en este proyecto, dado que la mayor parte de la documentación es física y no puede importarse automáticamente—, clasificar las entradas en carpetas según once categorías generales y múltiples subcategorías, y exportar finalmente los registros generados.

Mendeley exporta los datos en formato XML, RIS y BIB, ninguno de los cuales es compatible con MongoDB. La interfaz gráfica Compass de MongoDB permite importar archivos JSON que genera automáticamente la base de datos con sus respectivos elementos. No existe, sin embargo, un método plenamente satisfactorio para convertir archivos XML, RIS o BIB a JSON sin errores significativos en el tratamiento de los datos. Esto se debe a que cada gestor bibliográfico utiliza estructuras distintas para describir campos equivalentes. No se ha encontrado ningún programa capaz de reconocer de forma fiable esas diferencias y equivalencias. Este problema afecta directamente a la estructura del documento y, por tanto, a la clasificación en la base de datos. Para resolverlo, se accedió a la API de Mendeley y se extrajeron desde sus servidores todos los registros en formato JSON. Esto devuelve un documento con varios registros. Cada registro, aunque varía según su tipología —libro, artículo de prensa, página web, etc.—, comparte una estructura general y varios campos comunes.

El JSON generado incluye algunos campos irrelevantes para nuestra base de datos, por lo que se diseñó un script en Python para limpiarlos y ajustarlos a las necesidades del proyecto. Se eliminaron los campos «folder_uuids», «hidden», «group_id», «confirmed», «private_publication», «profile_id», «read» y «starred». Además, Mendeley clasifica como «source» los libros que aparecen en los registros pertenecientes a un capítulo de libro. Esto genera ambigüedades, ya que este mismo campo se utiliza para identificar las publicaciones periódicas. Para evitar esta confusión, se reemplazó «source» por «book» en los capítulos. Mendeley tampoco distingue entre autor, editor, ilustrador, traductor, director y coordinador, que clasifica en el momento de la exportación indistintamente como «authors». También se desglosó el campo genérico «authors» en sus respectivos roles: «author», «editor», «illustrator», «coordinator», «translator» y «director». Asimismo, Mendeley exporta la UUID de la carpeta a la que pertenece el registro, pero no su nombre. Gracias al script, se recuperó no solo el nombre de la carpeta, sino también el de las carpetas superiores que constituyen la estructura jerárquica en que se clasifica el registro. Estos metadatos ayudarán a afinar las búsquedas en la base de datos.

Una vez obtenidos los datos depurados en formato JSON, se cargó el archivo en Atlas, el servicio en la nube de MongoDB. Se crearon índices para los campos numéricos —año, mes, día—, mientras que los campos de texto se optimizaron mediante el índice específico de Atlas Search. Este índice permite búsquedas textuales eficientes y ha sido configurado para tolerar mayúsculas y tildes, mejorando la precisión en consultas en español.

A pesar de ser un desafío complejo, una base de datos como la que aquí proponemos es interoperable con una relacional como las antemencionadas. De entre las diversas estrategias que existen, abogamos por la creación de APIs

que permitan a los diferentes archivos digitales exponer sus datos de manera estructurada y consistente. Esto habilitaría a proyectos externos consultar y federar la información de múltiples archivos sin necesidad de conversiones complejas. El acceso a la API de Mendeley es un ejemplo de este flujo de trabajo. Con todo, lo deseable sería trabajar con bases de datos no relacionales, dadas las ventajas expuestas respecto a las no relacionales.

En definitiva, el desarrollo de la base de datos que nutre el Archivo Digital Sebastià Juan Arbó, mediante el uso de un modelo no relacional, ha permitido resolver dificultades que con un modelo relacional y los recursos disponibles habrían resultado insalvables. Este enfoque ha brindado mayor flexibilidad en la entrada de datos, contribuyendo a una mejor optimización de la aplicación. El uso de Mendeley ha garantizado la integridad de los registros, y todo el proceso ha enriquecido los datos con metadatos que permiten búsquedas más precisas y abren nuevas posibilidades de escalabilidad. Los resultados obtenidos evidencian la eficacia de las bases de datos no relacionales en proyectos de desarrollo de archivos literarios digitales.

Bibliografía

BIA PLATAS, A., & RODRÍGUEZ SALA, J. (2017), «Construcción de una base de datos y un repositorio de documentos de investigación para el proyecto TRACE», *Revista de Humanidades Digitales*, 1, pp. 287–295.

COHEN, S. (2017), «Apertura, difusión y accesibilidad: el uso teórico y práctico de una base de datos», en R. LÁZARO NISO (ed.), *Corpus y bases de datos para la investigación en literatura*, Fundación San Millán de la Cogolla, pp. 35–48.

GONZÁLEZ SORIANO, J. M. (2021), «Mnemosine: Biblioteca digital de la otra Edad de Plata (orígenes, contenidos, perspectivas)», *Signa: Revista de la Asociación Española de Semiótica*, 30, pp. 31–58.

MATAS ROCA, M. (2021), *Sebastià Juan Arbó: De la realitat viscuda a la ficció narrativa. Anàlisi d'un desarrelament en la literatura catalana*, Universitat Rovira i Virgili.

MELO FLÓREZ, J. A. (s. f.), *De las fichas de archivo a las bases de datos*. Recuperado el 24/04/2024 de https://taller-abierto-de-humanidades-digitales.github.io/bases_de_datos/

RAMIS, J. M. (2011), *Autotraducció i Sebastià Juan Arbó*, Universitat Pompeu Fabra.

ROMERO LÓPEZ, D. (2014), «Hacia la Smartlibrary: Mnemosine, una biblioteca digital de textos literarios raros y olvidados de la Edad de Plata (1868–1936). Fase I», en S. LÓPEZ POZA & N. PENA SUEIRO (eds.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, Universidade da Coruña y SIELAE, pp. 411–422.

SANTOS ZAS, M., MARTÍNEZ RODRÍGUEZ, F., & MASCATO REY, R. (2014), «La creación y gestión del archivo digital valleinclaniano: corpus manuscrito

e impreso», en S. LÓPEZ POZA & N. PENA SUEIRO (eds.), *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, Universidade da Coruña y SIELAE, pp. 435–457.

SIMONE KIRBY, J. (2019), «How NOT to create a digital media scholarship platform: The history of the Sophie 2.0 project», *IASSIST Quarterly*, 42(1), pp. 1–16.

TOMASI, G. (2020), «Las humanidades digitales y la base de datos MeMoRam para un enfoque sistemático hacia los motivos en los libros de caballerías», *Historias Fingidas*, 8, pp. 129–156.

Creación de un modelo del HTR para letra cortesana a partir de procesos inquisitoriales del siglo XVI

Maryam Kebli Jebari
Universidad de Castilla-La Mancha

Palabras clave: Humanidades digitales; transcripción automática; Inquisición española; ciencia abierta; Inteligencia Artificial
Keywords: Digital Humanities; Automatic Transcription; Spanish Inquisition; Open Science; Artificial Intelligence.

Introducción

En las últimas décadas, la digitalización del patrimonio documental ha transformado la investigación histórica y filológica, especialmente al combinarse con herramientas de Inteligencia Artificial. Esta convergencia ha facilitado el acceso a fuentes antes poco accesibles, aunque su aprovechamiento pleno sigue condicionado por la dificultad de lectura de los manuscritos históricos (Bazzaco, 2020). Escrituras complejas como la letra cortesana o procesal, frecuentes en los documentos administrativos y judiciales del Antiguo Régimen, presentan desafíos añadidos debido a la ausencia de unas normas ortográficas fijas, la variedad gráfica entre los escribanos y el deterioro físico de los manuscritos (Echavarría, 2023). En este contexto, el presente proyecto ha tenido como objetivo desarrollar un modelo de reconocimiento automático de texto manuscrito (HTR) adaptado a documentos procesales de la Inquisición española, utilizando la plataforma de Transkribus.

Corpus documental

Seleccionar un corpus documental adecuado es un paso crucial para realizar un entrenamiento de un modelo HTR, especialmente cuando se pretende trabajar con escrituras complejas como la letra cortesana (Ayuso, 2022). En este proyecto, se ha escogido un proceso inquisitorial conservado en el Archivo Diocesano de Cuenca, que documenta la persecución ejercida por el Santo Oficio contra la familia de Rodrigo de la Peña. Este documento resulta especialmente valioso no solo por su relevancia histórica, sino también por su utilidad técnica para la fase de la segmentación, transcripción y normalización del manuscrito. A través de este caso, se abre una ventana a las prácticas represivas del siglo XVI. Rodrigo de la Peña, miembro de una destacada familia judeoconversa de Medinaceli, fue juzgado y encarcelado por participar en rituales vinculados al criptojudaísmo.

La documentación generada en torno a su proceso incluye un conjunto diverso de fuentes como confesiones, declaraciones, genealogías y testimonios sobre prácticas religiosas prohibidas, que forma parte de un corpus más amplio compuesto por más de veinte causas relacionadas con familias como los Peña, Díaz, Fuente o Méndez. Esta repetición discursiva y la continuidad de los escribanos responsables convierte al conjunto documental en un terreno fértil para el entrenamiento del modelo de HTR, ya que facilita la detección de patrones gráficos y lingüísticos recurrentes.

Establecer las normas de transcripción

Las reglas de transcripción han sido diseñadas en colaboración con los especialistas en humanidades como el catedrático en Filología Clásica Ignacio Javier García Pinilla de la UCLM, el doctor en Historia Moderna Francisco Javier Moreno de la UCLM, el doctor en Historia Moderna David Martín López de la UCLM, la doctora en Ciencias y Técnicas historiográficas Bárbara Santiago Medina de la UCM y el doctor en Humanidades Digitales Andrés Felipe Echavarría de la Universidad de Montpellier-3, para aplicarse al *Proceso inquisitorial contra Rodrigo de la Peña*, pero con el objetivo de poder ser aplicados a otros procesos inquisitoriales similares. Las normas de transcripción paleográfica utilizadas se basan en los criterios establecidos en el *Informe sobre catalogación y descripción de manuscritos* (Ministerio de Cultura, 2001). Los criterios de transcripción son:

- **Conservación de la grafía original:** respeto de las formas originales del texto. Se respetan las grafías como ***ç*** y la ***z*** en palabras como *carçel, liçenciado, dezir*, *hazer, hedad* y *abia*.
- **Adaptación de grafías al uso actual:** a pesar del respeto de la forma histórica, se han introducido algunas adaptaciones a la forma actual para facilitar la lectura a un amplio público. Por ejemplo, se ha normalizado el uso de la ***v*** y la ***b*** según el criterio ortográfico actual, así como la conservación de ***mb*** por ***nb*** y ***mp*** por ***np.*** Asimismo, se resuelve la vacilación entre ***f-*** y ***h-*** de acuerdo con el uso moderno.
- **Normalización del uso de *u* y la *v*:** dado que en los textos antiguos la ***u*** y la ***v*** se utilizaba indistintamente como vocales o consonantes, se ha optado por el uso actual, por ejemplo: *vuo* > *ubo*.
- **Separación y segmentación moderna de palabras:** se han separado las grafías según las reglas del castellano actual, separando las sílabas unidas erróneamente y separando las palabras que han sido unidas artificialmente. Por ejemplo: *quelos* > *que los*
- **Tratamiento de contracciones y abreviaturas:** las contracciones en desuso se han desarrollado conforme a las reglas actuales. Igualmente, las abreviaturas se han extendido íntegramente sin uso de corchetes. Por ejemplo: *verdad, reverendos, dicho,* etc.

- **Mayúsculas, letras dobles e i/j/y:** se ha seguido el uso moderno para el uso de la mayúscula en los nombres propios. Las letras dobles en el inicio de las palabras se transcriben como simples como *rrey* > *rey,* mientras que las letras entre vocales se conservan. Las diversas formas de la letra ***i*** se unifican en una única representación (i), respetando, no obstante, la ***j-*** inicial. La ***y*** con valor vocálico se moderniza y se transcribe como ***i***, como en los casos *habya* > *habia.*
- **Conservación de repeticiones accidentales.**

Segmentación de las imágenes

Más allá de su carácter técnico, la segmentación representa también una elección metodológica significativa, ya que expresa los fines y prioridades del proyecto (Bazzaco, 2022). En este trabajo, uno de los principales objetivos ha sido respetar y mantener la estructura propia de los documentos judiciales de la Inquisición española. Estos no constituyen un texto uniforme, sino que representan una organización interna jerarquizada, en la que cada fase del proceso se introduce mediante fórmulas reiteradas y estandarizadas a lo largo del tiempo (Santiago, 2016).

Dichas fórmulas, que pueden denominarse partículas judiciales, funcionan como auténticos marcadores discursivos. Entre las más frecuentes se encuentran expresiones como “En la ciudad de …”, “Dicho testigo dijo…” o “Mandamos que …”. Identificarlas como encabezamientos o unidades textuales diferenciadas permite instruir al modelo de HTR para que reconozca estos segmentos como secciones autónomas, dotadas de una función y un sentido específico dentro del procedimiento judicial.

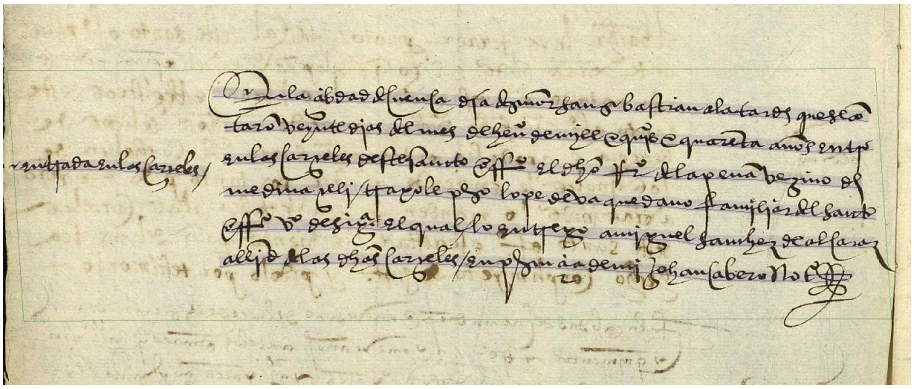


Figura 1. Imagen que refleja las partículas judiciales que podemos encontrar en los procesos inquisitoriales.

En el proceso inquisitorial de Rodrigo de la Peña, sobre el que se centra este estudio, es posible identificar con claridad diversas fases procesales como la presentación, el pedimiento, el mandamiento, la declaración y la sentencia. Cada una de estas etapas se introduce mediante una fórmula fija que ha sido etiquetada como encabezamiento durante la segmentación. De este modo, al

mostrarle al modelo que dichas expresiones delimitan bloques textuales distintos, se le ofrece una estructura lógica y secuencial que no solo favorece a una correcta transcripción, sino que también contribuye a la interpretación coherente del discurso judicial del Santo Oficio (Echavarría, 2024).

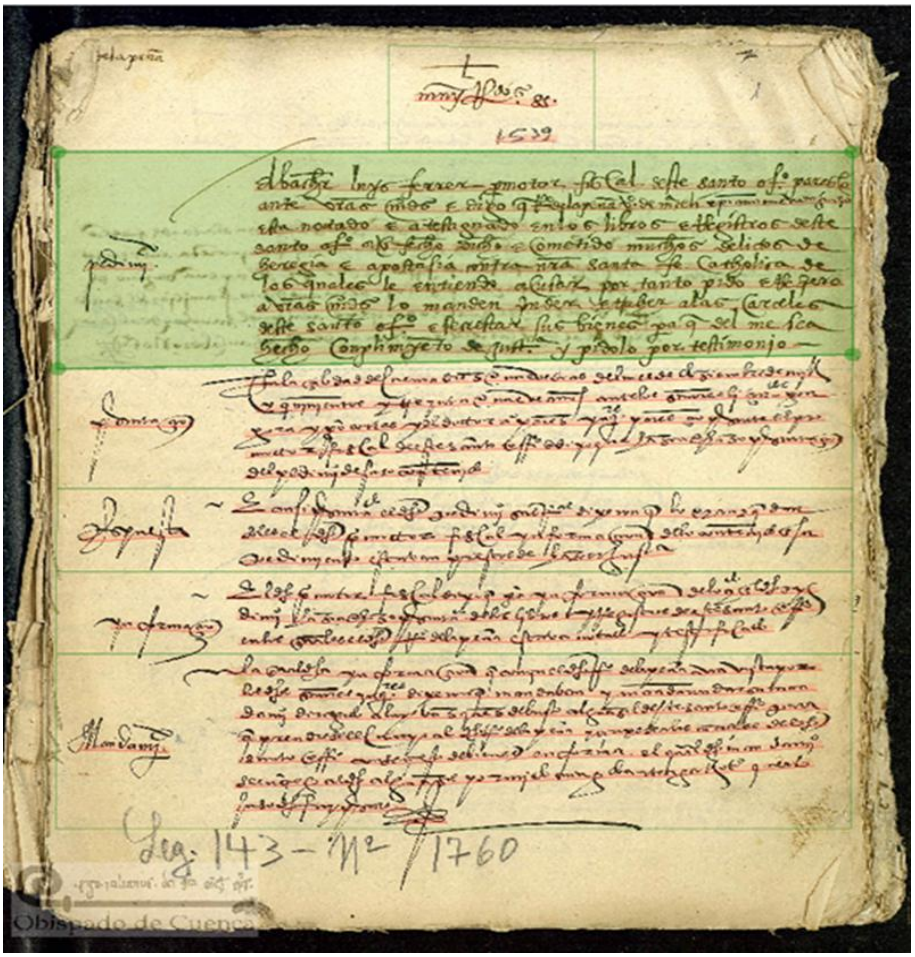


Imagen 2. Imagen que representa la segmentación en función de las fases judiciales de un proceso inquisitorial.

Por otro lado, se ha optado por excluir conscientemente las notas marginales accesorias presentes en el manuscrito. Aunque pueden presentar un interés desde un punto de vista paleográfico, muchas de estas marcas corresponden a comentarios del escribano o instrucciones procesales secundarias que no forman parte del relato principal. Su inclusión habría introducido elementos ajenos a la estructura textual central, dificultando así el aprendizaje del modelo respecto a las unidades textuales fundamentales.

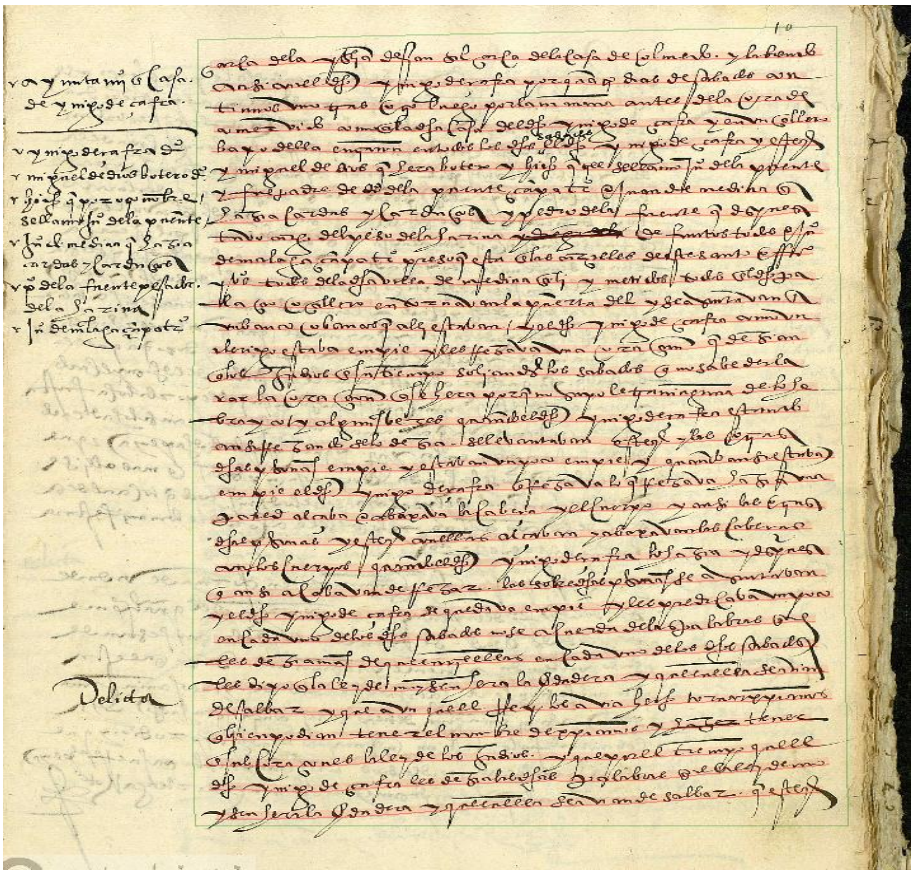


Figura 3. Imagen que muestra la exclusión de las notas marginales del entrenamiento del modelo de HTR.

Entrenamiento del modelo

El proyecto se inició con la selección de un proceso inquisitorial completo compuesto por 147 páginas digitalizadas. Tras la segmentación en líneas, se elaboró un *Ground Truth* de 50 páginas, que fue ampliado posteriormente hasta alcanzar las actuales 110 páginas. Se entrenaron dos versiones del modelo con distintas proporciones de validación: 5% y 10% (Cuellar, 2023).

En el primer caso, el conjunto de entrenamiento incluyó 104 páginas (3116 líneas y 37.700 palabras), mientras que el conjunto de validación contenía solo 6 páginas (187 líneas y 2221 palabras). El modelo alcanzó un *Character Error Rate* (CER) del 3,09% en el entrenamiento y del 9,26% en la validación. Aunque el conjunto de validación es limitado, los resultados reflejan un aprendizaje progresivo y una curva estable, con separación moderada entre las líneas de entrenamiento y validación a partir de la época 80.

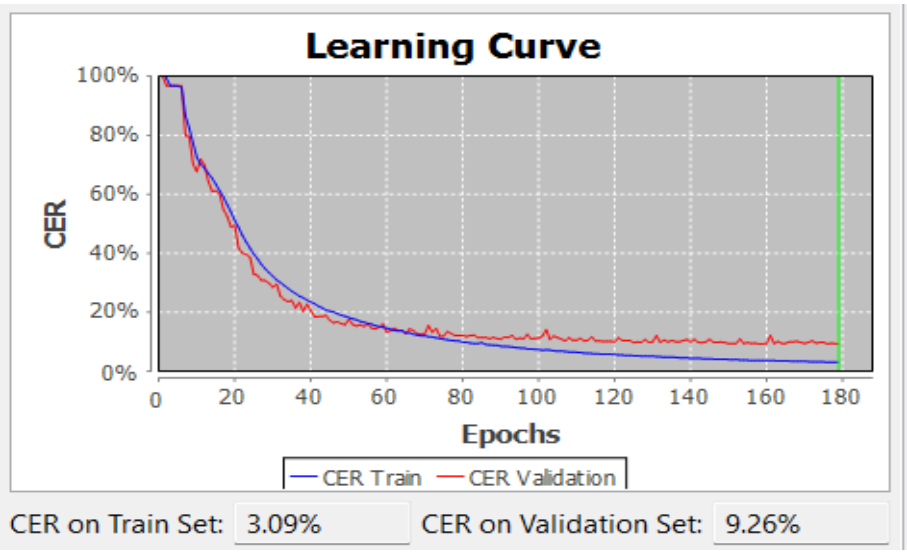


Figura 4. (VS 5%) Cuenca Inquisition trail 16th Century - M t2.

Por otro lado, el modelo con una validación del 10% entrenó sobre 99 páginas (2974 líneas y 35.900), dejando 11 páginas para la validación (329 y 3997 palabras). El CER alcanzado fue del 2,54% en entrenamiento y del 8,77% en validación. La curva correspondiente muestra una trayectoria más regular y ajustada entre ambas líneas, lo que indica una mejor generalización y menor riesgo de sobreajuste.

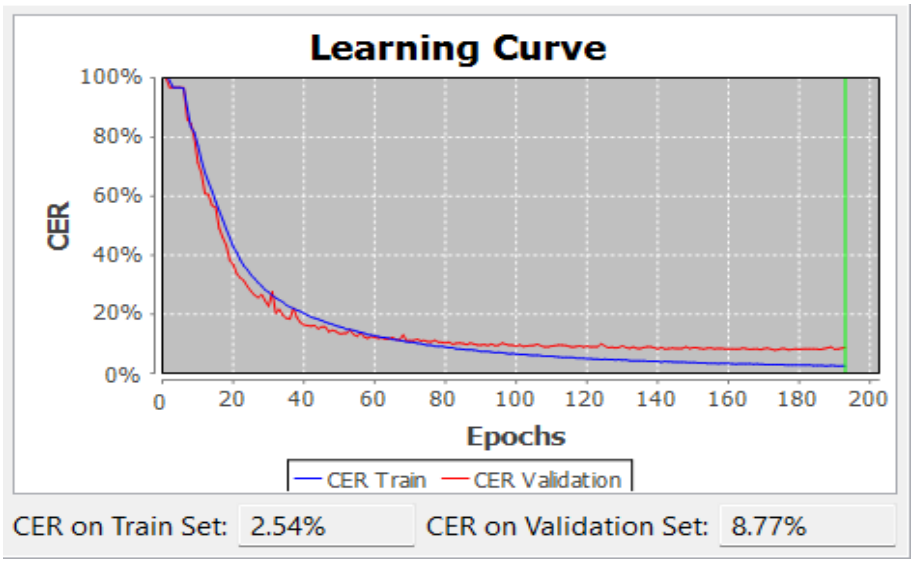


Figura 5. (VS 10%) Cuenca Inquisition trail 16th Century - M t2.

Conclusión

El modelo de HTR creado en este proyecto representa un avance en la transcripción automática de la letra cortesana, una de las grafías más complejas del Antiguo Régimen. A partir de un corpus cuidadosamente seleccionado y segmentado se ha entrenado un modelo capaz de reconocer estructuras textuales propias del discurso judicial del siglo XVI. La combinación de una

segmentación personalizada, la exclusión de las notas marginales y la transcripción normalizada ha permitido alcanzar tasas de error ajustadas y una curva de aprendizaje estable. Este modelo no solo es útil para el caso estudiado, sino que servirá para otra documentación inquisitorial de características similares, consolidando así una herramienta de valor para la investigación inquisitorial.

Bibliografía

AYUSO GARCÍA, M. (2022), «Las ediciones de Arnao Guillén de Brocar de BECLaR transcritas con ayuda de Transkribus y OCR4all: creación de un modelo para la red neuronal y posible explotación de los resultados», *Historias Fingidas*, 1, pp. 151–173. https://doi.org/10.13136/2284-2667/1102

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, pp. 534–561. https://www.janusdigital.es/articulo.htm?id=160

BAZZACO, S., JIMÉNEZ RUIZ, A. M., TORRALBA RUBERTE, Á., & 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*, 1, pp. 67–125. https://doi.org/10.13136/2284-2667/1190

CUELLAR, 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*, 11(1), pp. 101–115. https://doi.org/10.13035/H.2023.11.01.08

ECHAVARRÍA, A. (2024, 29 de febrero), «Mandamiento de citación», *Inq_Thes* (ID 1003229). Disponible en: https://opentheso.huma-num.fr/?idc=1003229&idt=Inq_Thes

ECHAVARRÍA, A. (2023, 12 de diciembre), *Tinta, píxel y texto electrónico. Transcripción asistida por computación*, conferencia, Facultad de Humanidades de Toledo.

MINISTERIO DE CULTURA (2001), *Informe sobre la catalogación de manuscritos*, Secretaría de Estado de Cultura. https://www.cultura.gob.es/dam/jcr:b5d86054-cd59-4dff-90c8-c397d8e09549/informe-catalogacion-de-manuscritos2001.pdf

SANTIAGO MEDINA, B. (2016), *La burocracia inquisitorial: Escrituras y documentos* [Tesis doctoral], Universidad Complutense de Madrid. https://produccioncientifica.ucm.es/investigadores/144781/tesis?lang=ca

Adaptación de los proyectos de investigación en Humanidades Digitales al paradigma de la Ciencia Abierta

José Manuel Barrueco Cruz
Jefe de Sección Tecnología, Normalización y Sistemas del Servicio de Bibliotecas y Documentación – Universitat de València
Cristina García Testal
Coordinadora de Circulación y Préstamo de la Biblioteca de CC Sociales– Universitat de València

Palabras clave: Humanidades digitales; ciencia abierta; movimiento open; buenas prácticas en investigación; procesos científicos
Keywords: Digital humanities; open science; open movement; best practices in research; scientific processes

Introducción

La aplicación de las tecnologías digitales en el estudio y en la preservación y difusión del patrimonio cultural ha supuesto una transformación en la investigación que ha llevado a la aparición de una nueva disciplina denominada Humanidades Digitales (HD).

Al mismo tiempo, asistimos a la generalización de un nuevo paradigma en el desarrollo de los procesos de investigación en torno al modelo de Ciencia Abierta, que nace como respuesta al proceso de globalización del conocimiento.

En palabras de la UNESCO la Ciencia Abierta es un enfoque inclusivo del proceso científico, que combina diversos movimientos y prácticas con el objetivo de hacer que el conocimiento científico sea accesible, reutilizable, transparente y colaborativo para todos y todas, y que esté disponible de manera que pueda ser compartido y desarrollado por toda la sociedad.

En la presente comunicación tratamos de hacer un análisis sobre cómo deberían confluir los principios de Ciencia Abierta con los proyectos de investigación en HD.

Ciencia abierta y Humanidades Digitales

La ciencia abierta se suele definir como la suma o combinación de una serie de movimientos de la comunidad académica tendentes a conseguir unas metodologías de trabajo más inclusivas, transparentes, etc.

El primero de esos movimientos es el de acceso abierto (Open Access) a los resultados de investigación y a los datos de investigación. Con él se pretenden eliminar aquellas barreras legales y económicas que suponen un obstáculo para la difusión del conocimiento.

Pero la ciencia abierta debe abordarse de forma integral en todos los flujos de trabajo del proceso investigador. Así se integran en ella otros movimientos abiertos como son:

- Open Research Data. Es necesario llevar a cabo una adecuada gestión de los datos de investigación que utilizamos. Para ello debemos trabajar con Planes de gestión de datos, hacer que nuestros datos sean usables y reutilizables por otros proyectos cumpliendo con los principios FAIR, maximizar su difusión bajo las correspondientes licencias abiertas.
- Open Source. Los datos necesitan ser procesados y analizados a veces a través del desarrollo de aplicaciones de software. Compartir estas herramientas en repositorios abiertos especializados como GitHub contribuirá a permitir la replicabilidad de la investigación.
- Open licenses. El maximizar la difusión de la información y los datos que generamos permitiendo su reutilización no supone una renuncia a los derechos de propiedad intelectual. El reconocimiento de la autoría de los materiales y la correcta atribución de la misma en las obras derivadas que se realicen debe ser un principio a seguir escrupulosamente. Debemos acostumbrarnos a utilizar de forma correcta licencias abiertas como las Creative Commons.
- Open peer review. Los procesos de revisión son esenciales para mantener la calidad de la investigación publicada. De hecho, el peer review es un indicador que define si una investigación puede considerarse científica. Debemos abrir los procesos al público en general, tanto a priori mediante la distribución de preprints para obtener el feedback de la comunidad como el peer review abierto posterior. Lo que se conoce como Ciencia ciudadana.
- Nuevos modelos de publicación en abierto. El sistema de distribución de revistas científicas ha sufrido una profunda transformación a causa del acceso abierto. Las denominadas revistas diamante en las que no hay cargos ni para autores ni lectores o las plataformas de publicación en abierto patrocinadas por las agencias de financiación de la investigación como Open Research Europe, Wellcome Open Research or Gates Open Research, donde tampoco se pagan cargos de publicación, deberían convertirse en la opción principal a la hora de publicar resultados de investigación.
- Nuevos modelos de evaluación de la investigación. Los organismos de financiación de la investigación o de certificación como la ANECA en

nuestro país están incluyendo las prácticas de ciencia abierta como aspectos evaluables favorablemente en las diferentes convocatorias.

En resumen, cualquier actividad investigadora entraría a formar parte de alguna de estos movimientos que en su conjunto nos encaminan hacia la ciencia abierta. Y como personal investigador, cada acción en nuestro día a día debería encaminarse a cumplir con los principios y recomendaciones de este nuevo paradigma.

Por su parte el término «Humanidades digitales» surgió en el año 2004 con la publicación de la obra A Companion to Digital Humanities (Schreibman, Siemens y Unsworth, 2004) en el que se sientan las bases del concepto, no como una simple digitalización sino como el inicio de una serie de interesantes cuestiones científicas.

Nos enfrentamos a importantes retos. Por una parte, están los derechos de autor que existen sobre los materiales culturales, tanto textos como audiovisuales o imágenes y que pueden dificultar la reutilización de estos datos. Por otra parte, que existan en las Humanidades Digitales diversos formatos, lenguas o distintos contextos culturales nos impone un enfoque flexible de forma que podamos respetar la pluralidad, pero en ningún caso renunciar a la interoperabilidad. Y estos obstáculos pueden superarse con la implementación de licencias abiertas como son por ejemplo las Creative Commons y a través del uso de metadatos estructurados y la utilización de estándares abiertos.

Convergencia entre Humanidades Digitales y Ciencia Abierta

Podría parecer que Humanidades Digitales y Ciencia Abierta juegan en campos distintos, pero, todo lo contrario, parten de los mismos principios base: la cooperación, el acceso abierto al conocimiento, el compromiso social y la utilización de infraestructuras digitales. Esto convierte a las Humanidades Digitales en un terreno abonado para adoptar los principios de desarrollo de la Ciencia Abierta. Además, la Ciencia Abierta puede enriquecerse con el enfoque contextual y crítico de las Humanidades Digitales, que reformula los circuitos del conocimiento, las estructuras establecidas y las formas de autoría.

La Ciencia Abierta no se limita a publicar artículos en libre acceso o a compartir paquetes de datos en un repositorio. Supone un cambio que va mucho más allá, cambia de base la forma en la que investigamos, reformula temas como el concepto de autor, o la propiedad intelectual, la manera en la que se está evaluando la producción académica o a enfrentarnos a la posibilidad de que un público no especializado participe en el debate de la ciencia. Y en toda esta evolución, las Humanidades Digitales ocupan una posición de privilegio ya que tienen la posibilidad de experimentar con formas de conocimiento más accesibles, más transversales y más horizontales.

Ciencia ciudadana y participación

Otra de las aportaciones que nos trae la Ciencia Abierta es la posibilidad de participación de la ciudadanía en el proceso científico, pueda ser en la recopilación de datos, en formular preguntas sobre la investigación o en la evaluación de los resultados.

Las Humanidades digitales tienen en este aspecto experiencias innovadoras en proyectos colaborativos por ejemplo la transcripción colaborativa y colectiva de documentos históricos, o la georreferenciación de fuentes, enriqueciendo los metadatos o siendo parte fundamental en la creación de archivos comunitarios.

Se fortalece el vínculo entre ciencia y sociedad ampliando el alcance de estos proyectos, además cuestionan la figura tradicional del investigador como única fuente de conocimiento.

Pero para que sea realmente efectiva, debemos repensar las metodologías tradicionales introduciendo espacios de diálogo, reconociendo las participaciones ciudadanas y devolviendo resultados, además de garantizar en las herramientas utilizadas la accesibilidad tecnológica y lingüística.

Datos abiertos y reutilizables

El fundamento de los datos abiertos se basa en que los datos producidos y recopilados en una investigación determinada estén disponibles en un formato accesible, abierto, publicados bajo licencias que permitan su reutilización, siempre respetando los derechos de terceros. En las Humanidades Digitales, esto supone una serie de dificultades que hay que resolver.

No es fácil definir qué se entiende por “dato” en el campo de las Humanidades, un concepto sencillo en las ciencias experimentales, por su estructura cuantitativa que no deja lugar a dudas; en Humanidades Digitales podemos referirnos a un texto literario, a una interpretación, un corpus lexicográfico, una reproducción digital de un documento analógico, una imagen, un sonido, etc. Esta heterogeneidad complica enormemente generar estándares, así como favorecer la reutilización de los datos generados.

La propia naturaleza interpretativa inherente al trabajo humanístico, hace que esos datos no sean neutrales ni tampoco definitivos. Un corpus puede ser anotado, segmentado o categorizado de maneras distintas dependiendo del enfoque teórico o metodológico que adoptemos.

Hay avances importantes en este sentido, como la creación de repositorios específicos para los datos generados por las Humanidades Digitales, por ejemplo, Zenodo, DARIAH, CLARIN, así como el desarrollo de esquemas de metadatos especializados como TEI o Dublin Core. El fundamento es que esos procesos de creación y tratamiento de datos, quede debidamente documentado, para que se garantice su trazabilidad y se facilite la comprensión por parte de terceros.

En cualquier proyecto de HD debemos tener en cuenta la documentación de todos los procesos relacionados con los datos a través de la elaboración de los correspondientes Planes de Gestión de Datos.

Evaluación, reconocimiento y sostenibilidad

El sistema actual de evaluación académica es, sin duda, uno de los mayores obstáculos para la implementación de la Ciencia Abierta en las Humanidades Digitales, ya que se sigue priorizando la publicación en revistas que aparezcan indizadas y además que publican en inglés, infravalorando otros ámbitos de publicación que publican en lenguas locales.

Esta práctica frena la adopción de prácticas abiertas a la vez que entorpece la sostenibilidad de los proyectos, ya que muchos de ellos se desarrollan gracias a una financiación pública que además es temporal. Es importante que se promuevan criterios de evaluación amplios e inclusivos, que se valore también el impacto social y que se favorezca la sostenibilidad de las infraestructuras digitales que generan los proyectos.

Existen iniciativas internacionales como CoARA (Coalition for Advancing Research Assessment) que tratan de encontrar nuevas métricas para una evaluación más efectiva. Desde las HD sería necesario estar pendiente de estos desarrollos en cualquier actividad que implique valoraciones de propuestas de investigación o promociones profesionales.

Infraestructuras y competencias digitales

También en esta transición de las Humanidades Digitales hacia la Ciencia Abierta necesitamos infraestructuras técnicas adecuadas y que los investigadores se formen en competencias digitales, desde las habilidades técnicas tales como la programación, el marcado semántico o la gestión de datos, hasta los criterios éticos relacionados con la propiedad intelectual, la protección de datos y la diversidad cultural. Y en esto tienen un papel principal las universidades y centros de investigación, fomentando el trabajo colaborativo entre investigadores y entre los distintos servicios institucionales creando espacios de intercambio entre las distintas disciplinas.

Ejemplos de iniciativas que articulan esfuerzos a nivel nacional e internacional: DARIAH-EU, el Consorcio de Humanidades Digitales Hispánicas (HDH) o la red OPERAS.

Conclusión

La adaptación de las Humanidades Digitales al paradigma de la Ciencia Abierta necesita cambios profundos estructurales tanto en sus políticas de financiación como de evaluación, cambios técnicos en infraestructuras y herramientas y cambios culturales que afectan a las prácticas y a los valores que rigen en la comunidad académica. Pero, fundamentalmente, es un cambio de actitud que nos abra al diálogo, a la colaboración y a la experimentación.

Tomando posiciones ante la descolonización del museo: análisis de la presencia africana en el ecosistema expositivo andaluz

Ángel María Lumbreras Fernández
María Ortiz Tello
Nuria Rodríguez Ortega
iArtHis_Lab –
Universidad de Málaga

Palabras clave: África; Exposiciones temporales; Transnacionalidad; Andalucía; Discursos curatoriales
Keywords: Africa; Temporary exhibitions; Transnationality; Andalusia; Curatorial discourses

La presencia de artistas de procedencia africana en las exposiciones temporales españolas constituye un fenómeno cultural e historiográfico de creciente interés, especialmente en el contexto de los debates contemporáneos sobre descolonización, globalización y circulación cultural (Del Moral, 2016; Baptista, 2023; Vergés, 2024). Esta cuestión no puede desligarse de las dinámicas transnacionales que configuran el sistema artístico contemporáneo como un entramado complejo de relaciones entre centros y periferias. Este trabajo pretende aproximarse a la realidad expositiva actual para identificar cómo las instituciones artístico-culturales andaluzas construyen y difunden su discurso en torno a “lo africano”. Para ello, se propone analizar la presencia africana en las exposiciones artísticas temporales celebradas en la región andaluza durante las dos últimas décadas (2000–2024), tomando como base el corpus Expofinder (expofinder.iarthislab.eu) y el grafo de conocimiento Complexhibit (http://complexhibit.eu/; https://github.com/Complexhibit-Project), que registran información relevante sobre cada muestra (Rodríguez Ortega, et al., 2024; Rodríguez Ortega, 2024).

Introducción: justificación, contexto, objetivos

El presente estudio tiene su eje vertebrador en la siguiente cuestión: ¿cómo construyen las instituciones culturales andaluzas el concepto de "lo africano" a través de su programación expositiva y discursos curatoriales? Esta pregunta se presenta urgente ante la actual deriva museística internacional como un estudio de caso localizado. Frente a estos esfuerzos de repensar el espacio museístico

desde una perspectiva más integradora, los propios colectivos, entre ellos los que engloban a personas racializadas negras, muestran su recelo ante la naturaleza colonial e histórica de la institución (Vergés, 2024). De este modo, el objetivo principal de nuestra propuesta es doble: por un lado, ofrecer un informe estadístico-descriptivo que permita dimensionar la presencia africana en términos cuantitativos y cualitativos; por otro, plantear una discusión crítica sobre cómo se construye discursivamente el concepto de "lo africano" desde las instituciones culturales de Andalucía, analizando los sesgos coloniales que atraviesan las estrategias expositivas, los textos curatoriales y las políticas de programación, teniendo presente bajo qué temáticas se exhiben sus piezas, qué apariciones con otros artistas se dan, qué retórica se emplea para su clasificación, etc. Esta vertiente implica examinar las prácticas de homogeneización cultural que reducen la diversidad africana a una unidad ficticia, así como los mecanismos retóricos que perpetúan estereotipos y relaciones de poder asimétricas heredadas del período colonial en los discursos curatoriales.

A su vez, Andalucía se presenta como un espacio privilegiado para analizar la presencia africana en las exposiciones temporales españolas por su condición de frontera geopolítica y cultural entre Europa y África. La proximidad geográfica comporta flujos humanos y culturales continuos que, lejos de limitarse a episodios históricos, se manifiestan hoy en un discurso curatorial enraizado en la historia compartida. Instituciones andaluzas, como el Centro Andaluz de Arte Contemporáneo (CAAC) o la Casa de la Provincia de Sevilla, han venido desarrollando desde hace años exposiciones dedicadas al Norte de África y Oriente Próximo, contribuyendo a visibilizar artistas magrebíes y reforzando el diálogo intercultural (con ejemplos como *Geopolíticas de la animación*, celebrada en 2007 o *25 años, 75 acuarelistas*, de 2014, respectivamente). En este contexto, elegir Andalucía como marco teórico-contextual permite interrogarnos no solo sobre la cantidad de exposiciones dedicadas a artistas de origen africano, sino también sobre las formas discursivas mediante las cuales las instituciones andaluzas legitiman, clasifican y difunden "lo africano". La región actúa así como laboratorio de tensiones entre herencia histórica, identidades híbridas y desafíos de descolonización cultural, ofreciendo un caso de estudio situado donde la geografía, la memoria y la práctica curatorial convergen.

Metodología y corpus

La metodología combina análisis documental y bibliográfico tradicionales con herramientas actuales de análisis y visualización de datos. Se han extraído y normalizado los registros de exposiciones de artistas africanos en museos, centros de arte y galerías andaluzas, clasificadas por procedencia geográfica. El subcorpus seleccionado de ExpoFinder para este estudio de caso incluye 219 exposiciones, 168 actores (17 nacionalidades) y 179 entidades (instituciones, centros y galerías de arte). Para cada exposición se recogieron fechas, sede,

comisario, temática, tipología, artistas y organizador; de cada actor, datos biográficos y actividades; y de cada entidad, tipo institucional y ubicación de su sede principal.

No obstante, el registro de datos sobre artistas de procedencia africana—y en especial de países subsaharianos—presenta retos significativos. Con frecuencia, el género de los creadores debe inferirse únicamente a partir de sus nombres, y la atribución de autoría se ve obstaculizada por la abundancia de obras anónimas, lo que distorsiona la representación y oculta contribuciones esenciales. La falta de autoría impide contextualizar temporal y estilísticamente estas piezas, agravando un sesgo informativo que invisibiliza narrativas completas. Estas circunstancias evidencian la complejidad de estudiar "lo africano" en el ámbito expositivo: la escasa información biográfica, la reducida visibilidad en los sistemas de información occidentales y la carencia de un reconocimiento histórico consolidado dificultan la identificación, catalogación y análisis de la presencia de estos artistas y sus obras en la esfera cultural.

Con el fin de establecer una distinción analítica relevante, se ha dividido África en dos bloques —el Magreb (Marruecos, Túnez, Argelia y Egipto principalmente) y el África subsahariana—, atendiendo a criterios geográficos, históricos, culturales y simbólicos discutidos en la bibliografía (Hannoum, 2021). Esta clasificación es fundamental, pues permite examinar si existen categorías gepolíticas que actúan como factor discursivo, además de cuantificar su presencia en el circuito expositivo andaluz. Ello ayudará a problematizar los discursos y narrativas existentes mediante su demostración empírica, y así avanzar hacia una comprensión más crítica y compleja de los procesos de inclusión y exclusión en el arte contemporáneo.

Resultados

El análisis estadístico-descriptivo revela la concentración de eventos en capitales, especialmente en Sevilla y Málaga, así como la escasa representación de artistas subsaharianos frente a artistas magrebíes, lo que corrobora la existencia de una división identitaria del continente africano en dos grandes bloques, norte y sur, fundamentada, además, en la naturaleza de las relaciones históricas entre ambas regiones y el contexto ibérico. El Magreb, debido a su proximidad geográfica y a los prolongados procesos de intercambio y contacto con la Península Ibérica, ha mantenido una presencia sostenida y significativa en el panorama expositivo andaluz. Esta realidad, tal y como puede observarse en el conocimiento extraído de los datos, se traduce en una mayor representación de artistas magrebíes en las instituciones culturales y artísticas de Andalucía, en comparación con la representación, mucho más limitada y diversa, de artistas procedentes del África subsahariana (figura 1).

Existe una marcada disparidad en el volumen de datos entre ambos conjuntos —donde uno presenta una muestra significativamente mayor que el otro— (figura 2) genera un sesgo cuantitativo que distorsiona los parámetros

estadísticos comparativos. De no controlarse esta asimetría, se induce una distorsión en la mediana del grupo minoritario, invalidando la comparación directa y obstaculizando la identificación de patrones específicos sobre la representación del ámbito subsahariano en la producción artística andaluza (esta prevalece en favor de las exposiciones dedicadas a artistas magrebíes, con una ratio de 2.6:1 en términos de registros totales). Por ello, la separación de ambos grupos en el análisis posibilitará en un examen más profundo futuro una aproximación más rigurosa y matizada de las dinámicas de visibilidad, inclusión y exclusión presentes en el ecosistema cultural objeto de estudio. Esto nos ha permitido corroborar con datos la crítica metageográfica ampliamente compartida en la literatura especializada, que advierte sobre los riesgos de homogeneizar realidades profundamente diversas bajo la etiqueta de “lo africano” (Baptista, 2023).

Asimismo, y como resultado preliminar, y como ejemplo de otros que se presentarán y desglosarán durante la ponencia, los datos revelan que los artistas magrebíes y los subsaharianos rara vez coinciden en una misma exposición, lo que se traduce en una segregación sistemática. Salvo contadas muestras colectivas de fotografía o vídeo (únicamente un 2,8 % muestra simultaneidad de ambos bloques, frente a 73,4 % dedicadas exclusivamente a artistas magrebíes y un 23,9 % a artistas subsaharianos), la programación tiende a organizarse en subgrupos homogéneos, reflejando líneas curatoriales diferenciadas para cada bloque geográfico y reforzando la división evidenciada previamente en el volumen de registros (figura 3). De este modo, las prácticas expositivas corroboran que esta división geográfica actúa como categoría discursiva.

La combinación de institucionalización segmentada (bloques magreb vs. subsahariana) y monopolización del comisariado por agentes no africanos evidencia un sesgo estructural en la programación, que prioriza la proximidad histórica geográfica con el Magreb y reserva el arte subsahariano a espacios secundarios, además de una brecha en la autoría curatorial, donde la voz de los curadores africanos tiene un peso marginal. Estas dos dinámicas articulan un ecosistema expositivo que, aun generando espacios de diálogo intercultural, reproduce jerarquías históricas y limita la pluralidad de perspectivas en la construcción discursiva de “lo africano” en Andalucía.

Figuras, tablas, mapas o gráficas

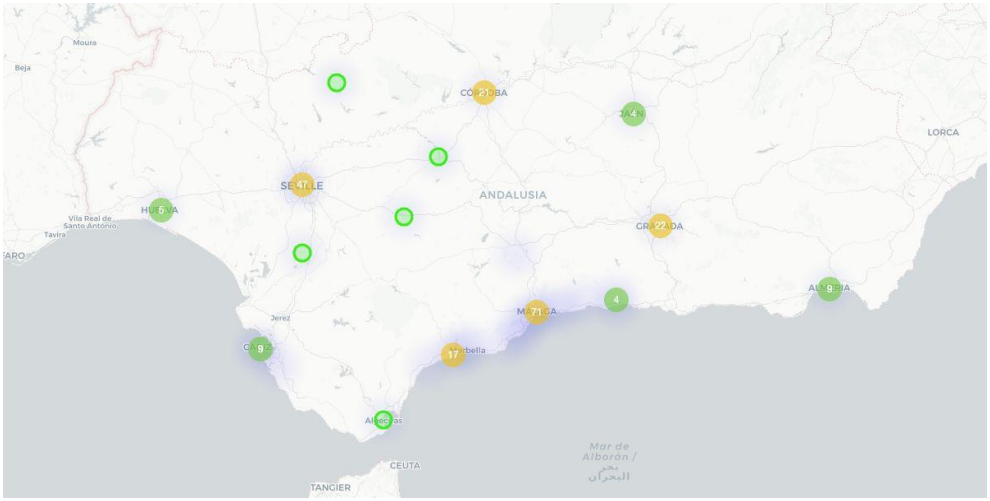


Figura 1. Mapa de calor con los principales clústeres expositivos de artistas africanos en Andalucía

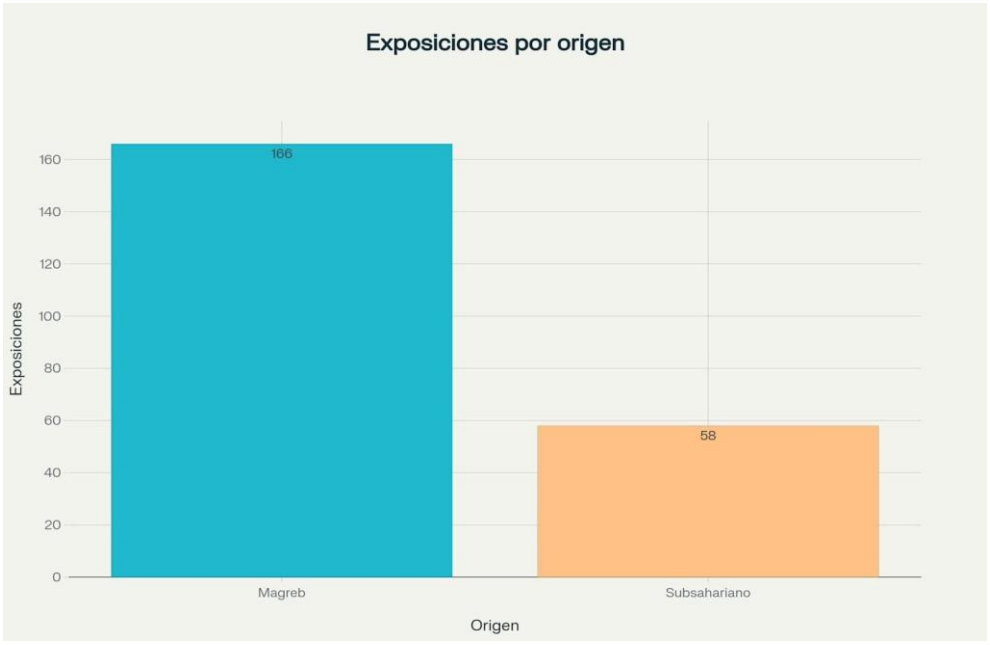


Figura 2. Comparación del volumen de datos entre artistas magrebíes y no magrebíes en el subcorpus seleccionado. 214 registros magrebíes frente a 81 registros subsaharianos, lo que representa una proporción de 72,5 % versus 27,5 % del total de datos. Esta diferencia se traduce en que los registros de artistas magrebíes son 2,6 veces más numerosos que los de artistas subsaharianos, con una diferencia absoluta de 133 registros.

Figura 3. Solo el 2.8% de las exposiciones mezclan artistas magrebíes y no magrebíes, evidenciando una segmentación casi total

Bibliografía

ALMAZÁN TOMÁS, D. (2018), *Exposiciones de arte africano: Historia del Arte, Educación y Museos en España*, recurso en línea. https://zaguan.unizar.es/record/69991/files/texto_completo.pdf

BAPTISTA, G. (2023), «AfroEspaña antes del boom. Una mirada interdisciplinaria sobre la presencia histórica y el legado cultural negro», *Transmodernity: Journal of Peripheral Cultural Production of the Luso-Hispanic World*, 10(2). https://doi.org/10.5070/T410261298

DEL MORAL GARRIDO, M. (2016), *Representación cultural y mediática de los "africanos/as" en Andalucía: De la invisibilidad a la auto-representación (vídeos domésticos, participativos y 'video films' de Nollywood)* [Tesis doctoral, Universidad de Granada], Editorial de la Universidad de Granada. http://hdl.handle.net/10481/43611

GUASCH, A. M. (2008), «El arte contemporáneo africano en un contexto global», en *El juego africano de lo contemporáneo*, pp. 130–136, MACUF (Museo de Arte Contemporáneo Unión Fenosa).

HANNOUN, A. (2021), *The Invention of the Maghreb: Between Africa and the Middle East*, Cambridge, Cambridge University Press.

SÁNCHEZ, M., & UBIRIA, S. (2023), «La mayor colección española de arte africano sale a la luz para cuestionar la mirada fetichista y racista», recurso en línea. https://www.eldiario.es/cultura/arte/mayor-coleccion-espanola-arte-africano-sale-luz-cuestionar-mirada-fetichista-racista_1_10020121.html

RODRÍGUEZ-ORTEGA, N. (2024), «Contours of knowledge: Epistemological implications of semantic models in the representation of the art exhibition domain through the lens of the OntoExhibit ontology», *Život umjetnosti*, 114, pp. 122–147. https://doi.org/10.31664/zu.2024.114.06

VERGÈS, F. (2024), *Programa de desorden absoluto: Descolonizar el museo* (F. López Martín, trad.), Akal.

¿Quién custodia lo enseñado?: La preservación de objetos educativos digitales durante y después de la pandemia por el Covid-19 en escuelas de Colombia y España

Jennifer Restrepo Monsalve
Universidad de Salamanca

Palabras clave: Preservación digital; Objetos Educativos Digitales; Educación básica; Memoria digital; Profesorado
Keywords: Digital Preservation; Digital Educational Objects; Elementary Education; Digital Memory; Teachers

Problema de investigación

Los cambios por los que atravesamos con mayor celeridad al inicio de la segunda década del siglo XXI, a causa de la crisis sanitaria producto de la pandemia por el Covid-19, demandaron a diversos sectores y, en particular, al educativo de su capacidad para adaptarse. En ese sentido, fue necesario para el profesorado dar apertura a prácticas pedagógicas, en muchos casos opuestas a las habituales, a través del uso de las pantallas. De modo que la integración de los medios digitales posibilitó apoyar el hecho educativo y dar continuidad a una suerte de rutina en tiempos inciertos:

La transición de un sistema educativo presencial a un modelo de educación a distancia o multimodal, ha sido el mayor sobresalto que ha recibido la educación en España y en otros muchos países en el último tercio del siglo XX y este comienzo del siglo XXI» (Díaz-Noguera y Barragán Sánchez, 2021, p. 12).

Durante el aislamiento obligatorio el contenido generado en medios electrónicos incrementó considerablemente «en España, el tráfico de internet aumentó un 40 % en la semana posterior al confinamiento, mientras que el tráfico de datos móviles aumentó un 25 %» (World Development Report, 2021), y, de paso, dejó ver la paradoja de la volatilidad resultado de la digitalización del saber: su preservación. Pues curiosamente contamos con el acceso a vestigios de sociedades arcaicas cuyos soportes físicos contenedores del conocimiento perduran hasta nuestros días—piénsese en la piedra, el papiro o el pergamino—, mientras que objetos y artefactos digitales como páginas web, videojuegos, fotografías o materiales audiovisuales, constitutivos de nuestra memoria en red (Echeverría, 2009), se pierden en la nebulosa de Internet. Son muchas las variables que determinan la obsolescencia en el ámbito digital, desde hardware con una vida útil muy limitada y que sale de circulación, software antiguo que no

se puede ejecutar en los nuevos medios, hasta ficheros y formatos (Térmens, 2013, p. 16) que cambian a un ritmo cada vez más vertiginoso.

La norma UNE 71361:2010, destinada a establecer un esquema de metadatos o LOM(Learning Object Metadata para los objetos digitales educativos, los define como «cualquier entidad digital susceptible de ser usada en aprendizaje, educación o formación» (AENOR, 2010). Precisamente, la enseñanza en tiempos de pandemia dejó a su paso un vasto patrimonio documental digital muchas veces conservado a nivel individual, en ordenadores de uso doméstico, en servicios de almacenamiento externo como Google Drive, OneDrive o Dropbox, sin una intervención institucional que se tradujera en políticas o protocolos para asegurar su conservación y preservación.

Tal aceleración del cambio ha venido en algunos casos de golpe y sin una alfabetización digital crítica (Samaniego Eguiguren, 2022) previa que permitiese asirse a alguna certeza en medio del desconcierto. En el caso de Colombia, la brecha social anterior a la pandemia se hizo más grande en tanto el profesorado tuvo que flexibilizar sus modelos de enseñanza, de evaluación, sus metodologías y métodos. Mientras algunas instituciones, por lo general privadas, pudieron establecer las infraestructuras necesarias para poner en marcha una estrategia de digitalización educativa, otras tuvieron que hacer llegar los contenidos por servicios de mensajería instantánea, dar clases telefónicas, o, en casos más extremos, las y los docentes tuvieron que personarse en casa de sus estudiantes. Lo propio ocurrió en la ruralidad de España donde, por ejemplo, no todo el estudiantado contaba con conectividad en condiciones óptimas para aprender, así como no todas y todos los docentes se consideraban ni se consideran usuarios aventajados de las TIC, ¿cuál será, pues, la evidencia de lo vivido al interior de la escuela durante el inicio de este último lustro?

Si bien en teoría los cuerpos de profesionales de la información o biblioteconomía son los más idóneos para responder a los desafíos que supone la preservación digital en cualquier institución, en el caso de la Educación Primaria y Secundaria de ambos países, se presenta una carencia de personal dedicado en exclusiva a labores de biblioteca escolar. Los resultados de la última Estadística de Bibliotecas Escolares-curso 2019-2020 publicada por el Ministerio de Educación, Formación Profesional y Deportes muestran que «en el 95,7% de los centros su profesorado participa en el equipo de la biblioteca, en el 4,2% personal especializado, en el 15,5% alumnado, en el 5,5% familias y en el 6,9% participan otros perfiles» (Ministerio de Educación, Formación Profesional y Deportes, 2021, p.2). En el escenario colombiano, según datos del último Diagnóstico Nacional de Bibliotecas Escolares (2019-2020), únicamente una tercera parte de las instituciones educativas públicas cuenta con personal bibliotecario de dedicación exclusiva, esto es un 31%. Dentro de ese porcentaje el 72% se encuentra en sectores urbanos frente a un 28% que está en zonas rurales. En contraste, la distribución del personal bibliotecario con dedicación parcial o no exclusiva, es prácticamente inversa: un 28% de los establecimientos

que cuentan con esta figura se encuentra en el área urbana y el 72% restante en el área rural (CERLALC y MEN, 2020).

De manera que la problemática tiene varias aristas que incluyen a diversos actores: al profesorado por ser un actor protagonista en la producción de objetos educativos digitales; a las administraciones estatales y locales; al cuerpo directivo cuya toma de decisiones incide directamente en las condiciones de los centros; y al personal bibliotecario, muchas veces ambivalente, en tanto puede tratarse de docentes con otras responsabilidades más allá de la biblioteca escolar. Por lo tanto, el aporte de mi investigación a la falta de modelos, planes, protocolos y prácticas sistemáticas de selección y preservación de los objetos digitales educativos— que en última instancia conforman un patrimonio de los y las docentes de educación básica— vendrá de poner en primer plano las concepciones, perspectivas y percepciones de quienes han accedido a dar su voz para reconstruir lo que fue aquel momento marcado por transformaciones y desafíos.

Dicho esto, mi pregunta de investigación es:

¿Cómo contribuir a la consolidación de modelos de preservación de los objetos educativos digitales que produce el profesorado de algunos establecimientos de educación básica, primaria y secundaria, en Colombia y España?

Metodología

La presente investigación es de enfoque cualitativo en donde se «pueden desarrollar preguntas e hipótesis antes, durante o después de la recolección y el análisis de datos» (Hernández Sampieri y Fernández-Collado, 2014, p. 7). Por lo que se entiende a la investigación como proceso y ello implica que su desarrollo se produce en fases tanto iterativas, como complementarias de recogida y análisis de datos:

- **Fase 1. Revisión exploratoria sistemática de la bibliografía**

Para establecer un estado de la cuestión, desarrollé un protocolo de revisión bibliográfica (Restrepo Monsalve, 2024) a partir del Marco ReSiste-SHC propuesto por Codina (2018), cuyas fases comprenden la búsqueda, evaluación, análisis, síntesis y presentación de las fuentes bibliográficas (Figura 1) (Codina, 2018, p. 14).

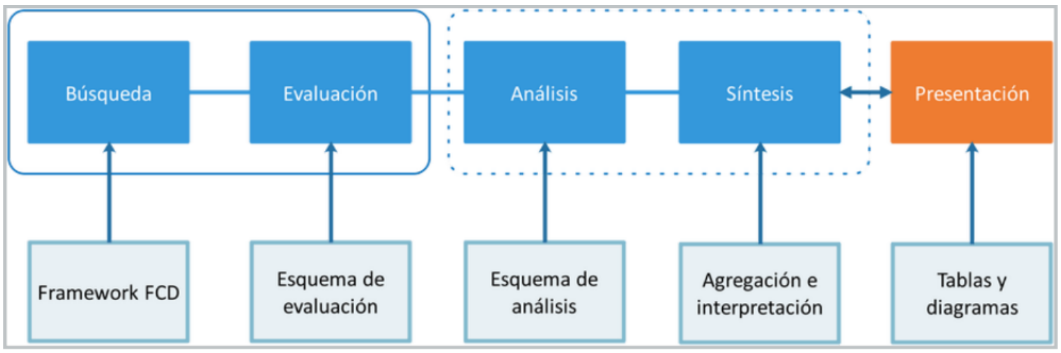


Figura 1. Diagrama del Marco ReSiste-SHC. Tomado de Codina (2018, p. 14)

- **Fase 2. Entrevistas semiestructuradas (Etapa actual)**

La guía de entrevista se compone de 12 preguntas iniciales distribuidas en 3 bloques: sobre la trayectoria docente, sobre los medios y recursos institucionales y sobre conocimiento del tema. De momento han participado 10 docentes de 6 centros, 3 ubicados en España y 3 en Colombia. Esta es la fase actual de la investigación.

- **Fase 3. Análisis documental y de entrevistas**

Una vez finalizada la recogida de datos de las fuentes primarias, se procederá a un análisis documental de la legislación y normativa vigente relacionada con preservación digital y con la educación básica en ambos países. Esto permitirá realizar una triangulación de la información contenida en las entrevistas y en las fuentes documentales estudiadas.

Resultados preliminares

Pese a que esta investigación aún se encuentra en desarrollo, durante las entrevistas han emergido varios temas recurrentes. Entre ellos destacan la ausencia de una cultura digital afianzada y, en consecuencia, la necesidad de una mayor alfabetización digital entre los actores educativos implicados. También se evidencia un desconocimiento generalizado acerca de iniciativas o medidas de preservación digital, si bien existen casos de docentes con un perfil tecnológico que la implementan de forma voluntaria a nivel individual sin que exista un apoyo institucional para lograrlo.

De manera general, hay una concepción de que es importante establecer protocolos institucionales para la preservación digital, cuya finalidad sería la de facilitar las labores de recuperación y conservación a largo plazo de los objetos educativos digitales generados al interior de los centros. Pese a esto, se advierte que dicha labor no debe recaer únicamente en el profesorado por lo que resulta fundamental una inversión de recursos económicos, de tiempo y de personal especializado en la materia.

Bibliografía

AENOR. (2010). *UNE 71361:2010 Perfil de aplicación LOM-ES para etiquetado normalizado de Objetos Digitales Educativos* (ODE). AENOR (Asociación Española de Normalización). https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0044685

CERLALC y MEN (2020). *Diagnóstico nacional de bibliotecas escolares 2019-2020: Informe general.* Ministerio de Educación Nacional de Colombia. https://cerlalc.org/publicaciones/diagnostico-nacional-de-bibliotecas-escolares-2019-2020-informe-general/

CODINA, L. (2018). *Revisiones bibliográficas sistematizadas: Procedimientos generales y Framework para ciencias humanas y sociales*. http://repositori.upf.edu/handle/10230/34497

DÍAZ-NOGUERA, M. D., y BARRAGÁN SÁNCHEZ, R. (2021). *Centros educativos: Transformación digital y organizaciones sostenibles: aprender y enseñar en tiempos de pandemia*. https://www.torrossa.com/it/resources/an/4972702

ECHEVERRÍA, J. (2009). «Cultura digital y memoria en red». *Arbor*, CLXXXV(737), 559-567. https://doi.org/10.3989/arbor.2009.i737.313

HERNÁNDEZ SAMPIERI, R., y FERNÁNDEZ-COLLADO, C. F. (2014). *Metodología de la investigación* (P. Baptista Lucio, Ed.; Sexta edición). McGraw-Hill Education.

MINISTERIO DE EDUCACIÓN, FORMACIÓN PROFESIONAL Y DEPORTES. (2021). *Estadística de bibliotecas escolares – Curso 2019-2020*: Nota resumen. [Gubernamental]. Enseñanzas no universitarias. Bibliotecas escolares. Curso 2019-2020. https://www.educacionfpydeportes.gob.es/servicios-al-ciudadano/estadisticas/no-universitaria/centros/bibliotecas/2019-2020.html

RESTREPO MONSALVE, J. (2024). «Preservación Digital y Educación Secundaria: una revisión exploratoria sistemática de la bibliografía». En *Modos de entender, pensar y sentir el patrimonio histórico educativo* (pp. 309-320). Ediciones Universidad de Salamanca. https://dialnet.unirioja.es/servlet/articulo?codigo=9837215

SAMANIEGO EGUIGUREN, J. M. (2022). *Un modelo de alfabetización digital crítica desde los estudios de ciencia y tecnología*. https://gredos.usal.es/handle/10366/150987

TÉRMENS, M. (2013). *Preservación digital*. Editorial UOC.

WORLD DEVELOPMENT REPORT. (s. f.). *Informe sobre el desarrollo mundial 2021: Datos para una vida mejor | Cómo la COVID-19 aumentó el consumo de datos y puso en evidencia la brecha informática*. Recuperado 5 de marzo de 2025, de https://wdr2021.worldbank.org/spotlights/how-covid-19-increased-data-consumption-and-highlighted-the-digital-divide/

La construcción discursiva del actor político en entornos digitales: análisis de la actancialidad en titulares mediáticos en redes sociales

Sofía Zakhir Puig
Universitat de València

La forma en que se representa a los actores sociales y se distribuye responsabilidad en el discurso adquiere una especial relevancia en el análisis del discurso contemporáneo, especialmente en contextos mediados por la rapidez de las redes sociales y la comunicación por internet. Un ejemplo especialmente revelador es el del discurso mediático, donde la elección del esquema sintáctico configura marcos de interpretación de los enunciados. Por ejemplo, en el contexto de la comunicación científica, en textos que se autodefinen como "escépticos" pero que sostienen posturas negacionistas, es frecuente el uso de formulaciones como *a number of scientists*, que funcionan como nominalizaciones opacas. Estas expresiones diluyen la responsabilidad discursiva al invocar una supuesta autoridad sin identificar actores concretos y, por tanto, sin precisar fuentes de evidencia[1]. Este tipo de construcciones no solo evita el compromiso epistémico con los datos, sino que fabrica una ilusión de consenso científico. Este ejemplo resulta representativo de fenómenos que materializan la urgencia de atender a la dimensión actancial como elemento central en el análisis del discurso político y mediático.

Este trabajo se enmarca en teorías que abordaremos brevemente en este resumen, del análisis del discurso, la pragmática, la lingüística cognitiva y la política en la red. Como breve nota genealógica, cabe destacar que a lo largo del siglo XX y con la sofisticación del análisis de la propaganda (especialmente en entornos como el *Institute for propaganda analysis*), se señalan (Krippendorff, 1990) mecanismos propagandísticos que pueden llevar a la identificación de discursos especialmente persuasivos, como, paradójicamente, "el recurso a generalidades llamativas" o "el llamar a las cosas y a las personas por su nombre, identificándolas sin rodeos". El exceso de ambigüedad y el personalismo conviven efectivamente en la propaganda política y resultan más destacables en el ámbito de la retórica populista. No hay, en efecto, ninguna contradicción en afirmar que el aglutinamiento de significados en un nombre, ambiguo o concreto, puede ser una estrategia discursiva efectiva para la persuasión, así como han

[1] Cuestión que en absoluto resulta baladí en el contexto político/mediático contemporáneo, en el que fenómenos como la democratización y la tecnologización han desvirtuado el acceso a la voz pública, con consecuencias tales como el escepticismo hacia el conocimiento científico.

confirmado los estudios sobre populismo, especialmente en el seno de la Escuela de Essex. Resulta evidente, además, que la ambigüedad puede usarse para esconder información estratégicamente.

En el contexto de las noticias en redes sociales, esta cuestión nos concierne por varios motivos: en primer lugar, por la técnica del *clickbait* (Bazaco et al., 2019), que pretende despertar la atención de los usuarios por medio de titulares con llamativos con poca o nula información; en segundo lugar, y siguiendo esta línea, por la ingente cantidad de lecturas que reciben los titulares en redes frente a la cantidad de *clics* y, por tanto, de lecturas reales de los artículos (Molyneux, 2018; Gabielkov et al., 2016). Esto se convierte, en el contexto del estudio de la actancialidad, en un posible reparto de responsabilidades sesgado que constituye en muchos casos la información final de los ciudadanos. Finalmente, por la intensificación afectiva de la comunicación mediada por ordenador (KhosraviNik, 2018), alentada por el propio funcionamiento de las plataformas y por los medios, lo cual puede, en ocasiones, concentrar afectos, tanto positivos como negativos, en un actor político concreto (en términos de la retórica clásica, se podría resumir en un paso del *ethos* al *pathos*).

Para el estudio de las estrategias discursivas, Gallardo (2021) propone un modelo de análisis del discurso basado en la lingüística cognitiva, que parte de la pragmática como estudio del lenguaje en su contexto. La autora asume los tres niveles de Morris (1938): enunciativo (referido a lo que se deriva de la intención del sujeto enunciador, como las implicaturas, actos de habla o deixis mostrativa); textual (referido a la *textura* del texto, a su cohesión, coherencia y conectividad); interactivo (lo que se deriva de la predictibilidad y prioridad). Este modelo, se desarrolla a partir de la noción de *frame* de Fillmore (1976), según la cual los enunciados activan una “escena cognitiva” a través de la memoria y en relación con conocimientos enciclopédicos e inferencias situacionales, entre otros. Esa activación se puede motivar por medio de diferentes decisiones del enunciador. Las que nos interesan especialmente en este trabajo son, en primer lugar, el plano léxico, donde se consideran las elecciones referidas a la denominación de acciones y actantes, incluyendo los usos figurados; y, en segundo lugar, el plano predicativo, en el que se analiza el empleo de la iconicidad sintáctica como estrategia para el reparto de la actancialidad. Según este modelo, entendemos por actancialidad[2] la configuración de los papeles participativos en el enunciado, es decir, la forma en que el discurso asigna funciones a los actores implicados en una acción. Como hemos adelantado, esta dimensión es una operación discursiva con implicaciones ideológicas, ya que permite construir oposiciones como *nosotros/ellos*, visibilizar o borrar actores, y orientar la interpretación del acontecimiento político o social.

[2] El reparto de responsabilidades se ha investigado en estudios muy diversos como por medio del análisis crítico del discurso sobre políticas públicas (Mulderrig; 2003; Hansen et al., 2014; Leipold y Winkel, 2015, 2017) o en profundos trabajos de revisión y propuesta teórica sobre actancialidad y voz pública (Marttila, 2016; Stuhler, 2021).

La elección de este foco responde a nuestro interés por observar cómo, en el discurso mediático sobre lo político, la construcción del sujeto de la enunciación y la selección sintáctica inciden directamente en el acontecer político. En un contexto de circulación acelerada e intensamente afectiva, como es el de las redes sociales, la gestión actancial no solo informa sobre el hecho referido, sino que orienta (y, potencialmente, limita) su lectura.

Por estos motivos, realizamos una investigación que se organiza en torno a un tema mediático de actualidad: la cuestión laboral en el contexto de los debates por la reducción de la jornada de trabajo semanal. Planteamos el análisis a través de la siguiente pregunta de investigación: ¿Qué papel juegan las estrategias predicativas, concretamente la actancialidad, en la comunicación política en redes sociales? Para este fin, creamos un corpus de titulares en la red social Twitter/X de cuatro medios de comunicación españoles con líneas editoriales diversas (El País, El Mundo, El Confidencial y elDiario.es) que se corresponde con las noticias relacionadas con lo laboral durante el año 2024. Analizamos, a continuación, de forma cualitativa, el reparto de responsabilidades según usos icónicos de la sintaxis en el programa de análisis de corpus AtlasTi.

Observamos que los discursos se organizan entorno a los nombres de quienes protagonizan el debate: Yolanda Díaz, como ministra de Trabajo y Economía Social; Carlos Cuerpo, como ministro de Economía, Comercio y Empresa; Antonio Garamendi, como presidente de la Confederación Española de Organizaciones Empresariales (CEOE) y, en menor medida, representantes de los principales sindicatos. Sin embargo, resulta significativa la neutralización por medio de la referencia a las respectivas instituciones que representan, en algunos casos, incluso, seguido de citas. Los titulares con citas cuyo emisor no se menciona no tienen una representatividad significativa, pero resulta llamativa la enunciación anónima (que se ha de entender en el contexto del *clickbait*). Además, hemos encontrado patrones significativos entre las líneas editoriales de los medios y la adjudicación de una participación activa/pasiva en los procesos políticos descritos. Por ejemplo, en los medios conservadores observamos una presentación de la CEOE como una entidad pasiva *sobre la que se legisla*, en enunciados que frecuentemente señalan la falta de consenso; mientras que los medios progresistas *activan* su voz como un miembro con agentividad en las mesas de diálogo.

Este trabajo muestra cómo la forma en que se construyen los titulares en los medios por medio del reparto de la responsabilidad entre los actores sociales es una estrategia útil a la persuasión y alineada con los intereses de cada línea editorial. Estos resultados refuerzan la necesidad de seguir la presente línea de investigación, también desde enfoques cuantitativos, cómo el lenguaje moldea la percepción política en el espacio público.

Bibliografía

BAZACO, Á., REDONDO, M., & SÁNCHEZ-GARCÍA, P. (2019). «El clickbait, como estrategia del periodismo viral: concepto y metodología». *Revista latina de comunicación social*, (74), 94-115.

GALLARDO, B. (2021). «El hablar como intención comunicativa». En *Manual de lingüística del hablar* (pp. 79-94). De Gruyter Mouton.

HANSEN, D. R., BØJE, J. D., & BECK, S. (2014). «What are they talking about? The construction of good teaching among students, teachers and management in the reformed Danish upper secondary school». *Education Inquiry*, *5*(4), 23356.

KHOSRAVINIK, Majid. (2018). «Social media techno-discursive design, affective communication, and contemporary politics». *Fudan Journal of the Humanities and Social Sciences*, *11*, 427-442.

KRIPPENDORFF, K. (1990). *Metodología de análisis de contenido: Teoría y práctica*. Paidós.

LATOUR, B. (2019). *Cara a cara con el planeta: Una nueva mirada sobre el cambio climático alejada de las posiciones apocalípticas*. Siglo XXI Editores.

LEIPOLD, S., & WINKEL, G. (2013). «Discursive Agency: Towards an actor-centered analysis of political discourses». En *Referat przedstawiony na 1st International Conference on Public Policy (ICPP 2013), Grenoble, France*.

LEIPOLD, S., & WINKEL, G. (2017). «Discursive agency:(re-) conceptualizing actors and practices in the analysis of discursive policymaking». *Policy Studies Journal*, *45*(3), 510-534.

MARTTILA, T. (2015, August). «Post-foundational discourse analysis: A suggestion for a research program». En *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 16, No. 3).

MULDERRIG, J. (2003). «Consuming education: a critical discourse analysis of social actors in New Labour's education policy». *Journal for Critical Education Policy Studies*, *1*(1), 96-123.

STUHLER, O. (2021). «What's in a category? A new approach to discourse role analysis». *Poetics*, *88*, 101568.

La ilusión de la conversación: marcadores del discurso en interacciones generadas por inteligencia artificial

Alessio Alessandro Santamato Peroni
Università degli Studi Internazionali di Roma – UNINT
Italo Cosentino
Università degli Studi di Firenze

Esta propuesta se centra en el análisis crítico de las capacidades actuales de la inteligencia artificial (IA) para generar interacciones lingüísticas que simulen la espontaneidad del discurso humano, con especial atención al uso de marcadores del discurso.

Aunque problemas similares pueden encontrarse también en otras lenguas, esta investigación se centrará específicamente en el análisis de los marcadores del discurso de la lengua española.

A través de un estudio de ejemplos prácticos, se identifican y examinan ejemplos en los que los modelos de lenguaje fallan en la producción adecuada de estos elementos, fundamentales para la coherencia, la gestión de turnos, la cortesía, la reformulación y otros aspectos pragmáticos esenciales en la conversación natural.

A pesar de los avances, la IA sigue presentando limitaciones sistemáticas en la incorporación contextual y funcional de marcadores como entonces, bueno, vale, o sea, pues, aún, entre otros. Estos operadores discursivos, altamente dependientes del contexto y cargados de funciones interaccionales, son frecuentemente utilizados de forma rígida o artificial por los sistemas actuales de IA, lo que reduce la fluidez y credibilidad de las interacciones generadas.

El estudio se basa en un corpus de interacciones entre usuarios humanos y modelos conversacionales de IA, analizando errores recurrentes, usos inapropiados o ausencias significativas de marcadores del discurso. A partir de estos datos, se proponen líneas de trabajo para una mejora en la sensibilidad discursiva de los sistemas, incluyendo la necesidad de entrenamientos específicos con corpus conversacionales auténticos y anotados desde la perspectiva de la pragmática griceana.

Esta perspectiva interdisciplinar, que combina la lingüística, la pragmática y la inteligencia artificial, tiene como objetivo avanzar en la investigación de sistemas más sensibles a los aspectos contextuales y relacionales del lenguaje.

Gracias a la centralidad concedida a la pragmática en la construcción de un modelo lingüístico coherente, la teoría metaoperacional nos ofrece un sistema orgánico de herramientas útiles para el análisis lingüístico (entre ellas, en

particular, el doble teclado de H. Adamczewski y los tres ejes de F. Matte Bon), particularmente válidas para el estudio y la enseñanza de operadores metalingüísticos como conjunciones, preposiciones y adverbios.

El estudio de las relaciones in praesentia e in absentia, tal y como señala Matte Bon (1997), permite identificar las elecciones del enunciador tanto desde el punto de vista de las relaciones con los interlocutores, como desde el punto de vista de la posición del propio enunciador en relación con el enunciado que está produciendo.

En este modelo gramatical, se manifiestan como ejes estructuradores del metalenguaje natural que constituye la gramática parámetros de la interacción como el tratamiento de la información por parte del enunciador, su actitud respecto a lo que dice y respecto a su interlocutor y, por último, el grado de referencia al mundo extralingüístico. En nuestra hipótesis, tales parámetros comunicativos configurarían las instrucciones procedimentales de estos operadores lingüísticos (Cf. Solís, 2012: 11) que la IA debería ser entrenada a considerar para crear intercambios naturales y que resulten válidos para la didáctica de ELE y el autoaprendizaje.

En concreto, la investigación pretende poner de relieve cómo la gestión de los marcadores del discurso representa un ensayo crucial para evaluar la calidad comunicativa de los sistemas de inteligencia artificial conversacional.

Por lo tanto, se pretende explorar sistemáticamente no sólo la frecuencia de uso de los marcadores en los diálogos generados por IA, sino también su ubicación, congruencia pragmática y capacidad para reproducir matices de cortesía, acuerdo, desacuerdo, vacilación o confirmación que constituyen la riqueza del discurso humano.

Aclarar las diferencias entre los distintos operadores que nos ofrece la lengua española para expresar afirmación, consecuencia, persistencia en el tiempo etc. favorece su adquisición por parte de los estudiantes de español como lengua extranjera, sin embargo, la IA presenta explicaciones y usos que, sin adecuadas intervenciones, resultan poco funcionales para la enseñanza y aún más, para el autoaprendizaje, como, por ejemplo, para los marcadores de consecuencia:

- “Entonces” → consecuencia lógica directa: no almorzó → debe tener hambre.
- “Pues” → consecuencia explicativa: no comió → causa justificada.
- “O sea” → en este caso, como marcador de consecuencia lógica final (no comió → ahora está mareado), no como reformulación

Como puede apreciarse, las definiciones de los operadores analizados, proporcionadas por el sistema IA utilizado, no consiguen dar cuenta de la operación metalingüística de estos elementos y, por lo tanto, de su función pragmática. Es más, al ser sistemas que se basan en corpus en línea o en obras que pueden encontrarse en la red, se limitan a presentar las mismas definiciones que normalmente se encuentran en todos los libros de gramática y en las obras que tratan los marcadores del discurso, lo que conlleva que presenten las

mismas problemáticas. En nuestro estudio, para alcanzar un resultado efectivo, nos proponemos interactuar con la IA de manera que tenga en cuenta las características de cada operador en el enunciado que está formulando. Por consiguiente, es fundamental seguir unas pautas clave como puede ser la de elaborar prompts adecuados para que el sistema responda de forma efectiva.

El objetivo final es proponer directrices concretas para el diseño de modelos lingüísticos que integren anotaciones pragmáticas de diálogos auténticos y representativos en su proceso de entrenamiento.

Esta perspectiva interdisciplinar pretende hacer avanzar la investigación hacia sistemas más sensibles a los aspectos contextuales y relacionales del lenguaje. En efecto, cada vez es más evidente que la capacidad de un asistente virtual para inspirar confianza y fomentar la colaboración con el usuario pasa también por microseñales lingüísticas que, aunque aparentemente secundarias, son fundamentales para la experiencia de interacción.

En conclusión, la propuesta pretende estimular un amplio y profundo debate científico sobre la necesidad de incluir sistemáticamente la dimensión discursiva y pragmática como componente estructural e indispensable en el desarrollo de modelos de IA conversacional, con el fin de superar una visión parcial y reduccionista de la competencia lingüística que corre el riesgo de descuidar la complejidad interaccional propia de las auténticas comunicaciones humanas.

El auspicio es que este trabajo pueda abrir un espacio de reflexión sobre los desafíos que plantea la generación automática del habla en contextos reales, y subrayar la importancia de los marcadores del discurso como elementos clave para lograr una IA verdaderamente interactiva y natural.

Bibliografía

CHIU, T. K., XIA, Q., ZHOU, X., CHAI, C. S., & CHENG, M. (2023), «Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education», *Computers and Education: Artificial Intelligence*, 4, art. 100118.

COSENTINO, I., & SANTAMATO PERONI, A. A. (2024), *La pratica della traduzione spagnolo-italiano: un nuovo approccio metodologico*, Roma, UNINT University Press.

COUPER-KUHLEN, E., & SELTING, M. (2018), *Interactional Linguistics: Studying Language in Social Interaction*, Cambridge, Cambridge University Press.

GAVIÑO RODRÍGUEZ, V. (2008), *Español coloquial. Pragmática de lo cotidiano*, Cádiz, Servicio de Publicaciones de la Universidad de Cádiz.

MATTE BON, F. (1997), «Criterios para el análisis de la lengua desde la perspectiva de la comunicación», en *Curso de Lengua Española III de la Carrera de Humanidades de la Universitat Oberta de Catalunya*, Barcelona, Universitat Oberta de Catalunya (publicación interna).

MATTE BON, F. (1999), «¿Cómo debe ser una gramática que aspire a generar autonomía y adquisición?», en LOSADA ALDREY, M.ª C., MÁRQUEZ CANEDA, J. F., & JIMÉNEZ JULIÁ, T. E. (coords.), *Español como lengua extranjera, enfoque comunicativo y gramática. Actas del IX Congreso Internacional de ASELE, Santiago de Compostela, 23–26 de septiembre de 1998*, Santiago de Compostela, Servicio de Publicaciones de la Universidad de Santiago de Compostela, pp. 57–79.

MATTE BON, F. (2000 [1992]), *Gramática comunicativa del español. De la lengua a la idea*, Madrid, Edelsa, 2 vols.

SOLÍS GARCÍA, I. (2012), *Por supuesto et alii… Tomas de posición en la afirmación*, Nápoles, Pisanti.

El Siglo de Oro en la prensa periódica (SILEM III): primera aproximación cuantitativa a los textos de la Biblioteca Digital Prensa y Canon

Laura Hernández-Lorenzo
Universidad de Sevilla

Palabras clave: Siglo de Oro; prensa decimonónica; Análisis de corpus; Topic Modelling; Análisis de sentimiento
Keywords: Golden Age; 19th-century press; Corpus analysis; Topic modeling; Sentiment analysis

La institución del Siglo de Oro en la prensa periódica (1801-1868): Proyecto SILEM III

Este estudio se enmarca dentro de las investigaciones llevadas a cabo en el Proyecto de investigación La institución del "Siglo de Oro". Procesos de construcción en la prensa periódica (1801-1868). SILEM III, dirigido por Mercedes Comellas. Se trata de un proyecto interuniversitario e interdisciplinario de las universidades de Sevilla y Córdoba, financiado por el Plan Nacional de I+D+i, que continúa la línea de investigación en la conformación del canon áureo de los anteriores proyectos SILEM y del grupo de investigación PASO. En esta ocasión el foco de atención es el papel de la prensa periódica del siglo XIX en la construcción de un ideario nacional en el que estaban presentes los autores áureos, institucionalizando el Siglo de Oro (Comellas, 2024). Para ello, se combinan métodos y enfoques como la historiografía, la historia literaria, o las Humanidades Digitales. En lo referente a esta última, con el objetivo de localizar y difundir las evidencias en los materiales periodísticos de la construcción e institucionalización del Siglo de Oro, el equipo del proyecto ha realizado un rastreo de referencias al periodo áureo, y a sus autores y obras, en un amplísimo corpus de prensa periódica decimonónica. Los resultados están siendo difundidos mediante transcripción y etiquetado semántico XML-TEI de textos en la Biblioteca Digital Prensa y Canon, que cuenta en la actualidad con más de 270 entradas. La cantidad de textos reunidos supone un aliciente para el empleo de métodos cuantitativos y de minería de textos de Humanidades Digitales que permitan extraer una visión global de estos y del tratamiento que se realiza en ellos de los autores y obras áureas.

Preparación del corpus

Este trabajo se ha llevado a cabo mediante el análisis de los textos que constituyen la Biblioteca Digital Prensa y Canon, preparada por el equipo de SILEM III. Esta se encuentra en proceso de ampliación y, en el momento de comenzar este trabajo (junio de 2025), contaba con 277 textos publicados. Estos fueron extraídos como archivos independientes en formato de texto plano y limpiados de metadatos e información ajena a los textos en sí. Asimismo, se descartaron dos textos por ser excesivamente breves (menos de 50 palabras). El corpus resultante contiene 275 textos y un total de 353,619 palabras tokens y 28,895 palabras tipo.

Metodología

La metodología utilizada se corresponde con los enfoques cuantitativos y de minería de textos, una de las líneas más destacadas dentro de las Humanidades Digitales (Hoover, 2008; Jockers & Underwood, 2016). Más concretamente, se combina el análisis con herramientas de corpus a través de Voyant tools (Sinclair & Rockwell, 2016), con las metodologías más sofisticadas del Topic Modelling y del Análisis de sentimiento. Voyant tools nos permitirá realizar una primera exploración del corpus a través de los términos y colocaciones más frecuentes del mismo, complementado con el análisis temático de Topic Modelling y con el análisis de la positividad y negatividad del lenguaje presente en los textos analizados a través del Análisis de sentimientos.

En el caso de Topic Modelling, se ha aplicado el algoritmo LDA, a través del paquete topicmodels (Grün & Hornik, 2011). La elección de esta implementación se debe a que está bien documentada, permite trabajar con los paquetes tidyverse, desarrollados para realizar análisis de minería de textos de forma rápida y eficiente, y cuenta con funciones para crear gráficos y evaluar automáticamente la coherencia de los tópicos obtenidos. Para el Análisis de sentimientos, se ha empleado el paquete 'syuzhet', que ha dado ya buenos resultados con textos en español (Isasi, 2021).

Resultados obtenidos

Primer análisis cuantitativo con Voyant tools

En primer lugar, se llevó a cabo un análisis cuantitativo con Voyant tools para hacer una primera exploración de los textos del corpus a través fundamentalmente de las palabras y las colocaciones más frecuentes, una vez eliminadas las stopwords . En el primer caso, se generó una nube de palabras mediante la herramienta Cirrus, en la que los términos más frecuentes aparecen con un tamaño mayor (véase la Figura 1). Esta información visual fue comprobada mediante la observación de la lista de palabras más frecuentes ofrecida por la herramienta Términos.

Figura 1. Nube de palabras con los textos del corpus.

Como puede apreciarse en la Figura 1, entre los primeros puestos se encuentran los siguientes autores áureos: Lope (430 ocurrencias), Cervantes (298), Calderón (260) y Quevedo (203).

También ha sido de gran interés el resultado obtenido con las colocaciones, un método de Lingüística computacional en el que se recuperan las palabras que tienden con más probabilidad a aparecer juntas en ciertos contextos (Baker et al., 2006, pp. 36-37). Las diez primeras colocaciones, obtenidas mediante la herramienta del mismo nombre, se encuentran recogidas en la Tabla 1.

	Término	**Contexto del término**	**Frecuencia absoluta**
1	lope	vega	164
2	siglo	xvi	75
3	literatura	española	70
4	siglo	xvii	59
5	siglo	xv	43
6	lope	calderón	40
7	literatura	historia	39
8	poesía	española	33
9	siglo	oro	28
10	poesía	lírica	28

Tabla 1. Diez primeras colocaciones obtenidas

Como puede apreciarse, entre las colocaciones más frecuentes se encuentran nombres de autores áureos como “lope-vega”, que hace referencia al dramaturgo, y “lope-calderón”, que muestra cómo ambos autores aparecen nombrados juntos con cierta frecuencia. Además, aparecen sintagmas relacionados con la época tratada (siglos XV, XVI y XVII), y lo que es aún más

interesante, encontramos el marbete “Siglo de Oro” como una de las colocaciones más frecuentes.

Topic Modelling

En segundo lugar, se realizó un análisis de Topic Modelling con el objetivo de explorar los principales ejes temáticos de los textos de la Biblioteca Digital. Puesto que no hay un número de tópicos predefinido, se han entrenado modelos de 10, 15, 20 y 25 tópicos. Asimismo, estos modelos se han entrenado tanto sobre el corpus sin lematizar como lematizado a través de Freeling (Padró & Stanilovsky, 2012). Seguidamente, se realizó una evaluación de los textos, tanto manualmente como a través de medidas de coherencia de tópicos (Mimno et al., 2011). Los mejores resultados fueron los obtenidos con el modelo de 20 tópicos sobre los textos lematizados (véase la Tabla 2).

Topic	Valor de Coherencia de Tópicos
1	-169.16228
2	-125.02811
3	-166.39404
4	-125.21372
5	-126.44931
6	-83.93241
7	-137.83220
8	-125.90416
9	-98.50346
10	-126.76636
11	-92.11150
12	-133.48740
13	-82.16397
14	-106.20614
15	-99.64114
16	-148.81405
17	-101.17806
18	-115.46980
19	-155.95658
20	-115.16893

Tabla 2. Valores de Coherencia de Tópicos obtenidos para 20 tópicos con el corpus lematizado

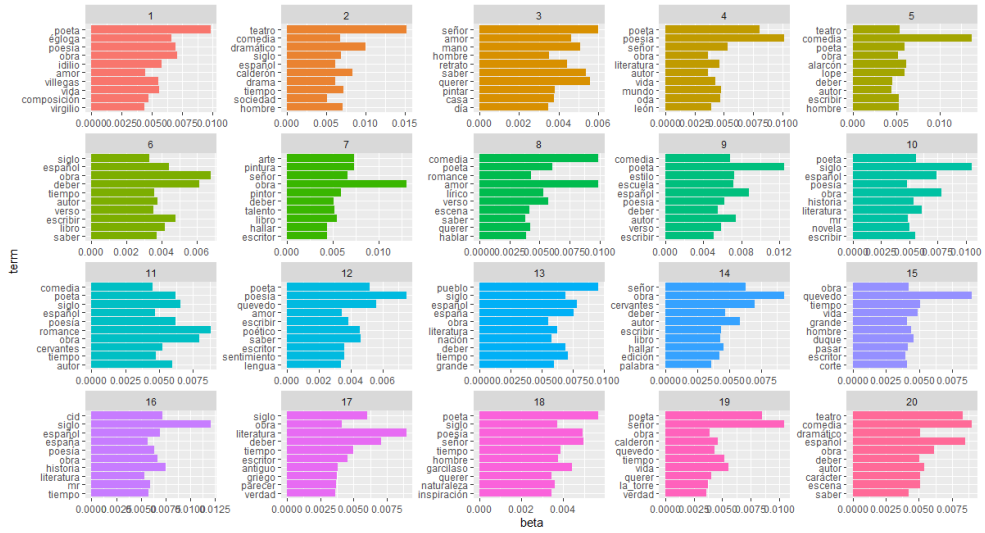


Figura 2. Resultados del análisis de Topic Modelling con el corpus lematizado, entrenado sobre 20 tópicos

Como puede apreciarse en la Figura 2, se han obtenido tres tópicos que se centran en el género dramático (2, 5 y 20), cuatro relacionados con el género lírico (1, 4, 12 y 18), tres que abordan la relación entre Literatura e Historia (10, 13 y 16), cinco que giran en torno a la escritura y al oficio de escritor (3, 6, 7, 9 y 17) y dos sobre Cervantes (11 y 14).

Análisis de sentimientos

Por último, se ha realizado una primera aproximación al Análisis de sentimientos para tratar de averiguar si en los textos de prensa que mencionan obras y autores áureos predominan los valores positivos o negativos. La media de sentimientos positivos obtenida para el corpus completo ha sido de 0.06727 frente a una media de sentimientos negativos de 0.03727, por lo que puede concluirse que predominan los valores positivos (Figura 3).

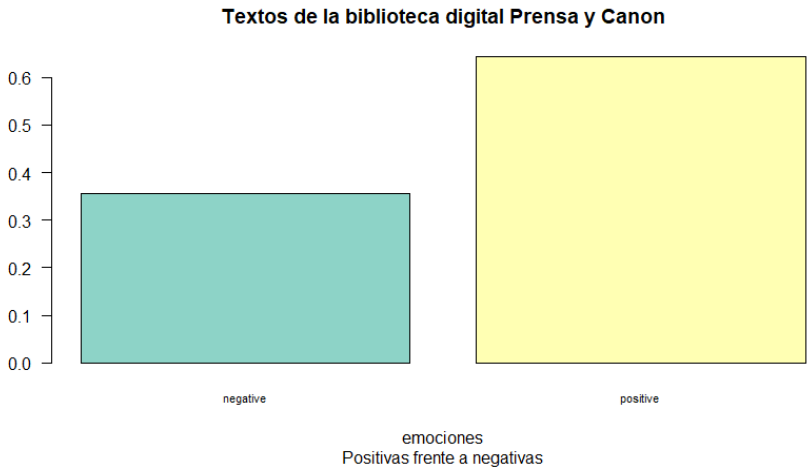


Figura 3. Gráfico de barras que representa las medias obtenidas para las emociones positivas y negativas

Conclusiones

En esta primera aproximación con métodos cuantitativos a los textos de la Biblioteca Digital Prensa y Canon, se ha extraído información sobre los términos y expresiones más frecuentes (entre los que se encuentran algunos destacados autores áureos, así como el marbete “Siglo de Oro”), se han explorado los principales ejes temáticos, relacionados con algunos subgéneros literarios como el lírico o el dramático, así como con Cervantes, la relación entre Literatura e Historia y el oficio de escritor; por último, el Análisis de sentimientos ha permitido corroborar que en estos textos de prensa que tratan del periodo literario áureo predominan los valores positivos. Ahora bien, este es un trabajo en progreso, por lo que será necesario profundizar en estas cuestiones y realizar nuevos análisis antes de alcanzar conclusiones definitivas.

Bibliografía

BAKER, P., MCENERY, A., & HARDIE, A. (2006), *A Glossary of Corpus Linguistics*, Edinburgh University Press.

COMELLAS, M. (2024), «Nación, historia y canon literario en la prensa decimonónica. El “Siglo de Oro” visto por los románticos», *Cahiers de civilisation espagnole contemporaine*, Spécial 5. https://doi.org/10.4000/1349h

GRÜN, B., & HORNIK, K. (2011), «topicmodels: An R package for fitting topic models», *Journal of Statistical Software*, 40(1), art. 1. https://doi.org/10.18637/jss.v040.i13

HOOVER, D. L. (2008), «Quantitative analysis and literary studies», en R. SIEMENS & S. SCHREIBMAN (eds.), *A Companion to Digital Literary Studies*, Blackwell, pp. 517–533.

ISASI, J. (2021), «Análisis de sentimientos en R con “syuzhet”», *The Programming Historian en español*, 5. https://doi.org/10.46430/phes0051

JOCKERS, M. L., & UNDERWOOD, T. (2016), «Text-mining the humanities», en S. SCHREIBMAN, R. SIEMENS, & J. UNSWORTH (eds.), *A New Companion to Digital Humanities*, Wiley-Blackwell, pp. 291–306.

MIMNO, D., WALLACH, H., TALLEY, E., LEENDERS, M., & MCCALLUM, A. (2011), «Optimizing semantic coherence in topic models», *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 262–272. https://aclanthology.org/D11-1024

PADRÓ, L., & STANILOVSKY, E. (2012), «FreeLing 3.0: Towards wider multilinguality», *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, ELRA.

SINCLAIR, S., & ROCKWELL, G. (2016), «Text analyzing and visualization: Making the meaning count», en S. SCHREIBMAN, R. SIEMENS, & J. UNSWORTH (eds.), *A New Companion to Digital Humanities*, Wiley-Blackwell, pp. 274–290.

Metodología de identificación de entidades mediante PLN en el corpus biográfico de DICAT[1]

Ariadna Fuertes Seder
Verónica Romero Gómez
Universitat de València
Luis de la Fuente Valentín
Universidad Internacional de La Rioja

Palabras clave: Reconocimiento de entidades (NER); biografías de actores áureos; procesamiento automático; topónimos
Keywords: Entity recognition (NER); biographies of golden actors; automatic processing; place names

La investigación humanística sobre el teatro clásico español se enfrenta hoy al reto de aunar e interpretar los grandes volúmenes de información biográfica, generada en décadas previas, a través de nuevas metodologías digitales. El auge de las Humanidades Digitales (HD) ha provocado una profunda transformación en la forma en que los investigadores abordan el estudio de grandes corpus textuales. En este contexto, y en el marco de su integración en el sistema federado de bases de datos del proyecto ASODAT, el proyecto DICAT (Diccionario biográfico de actores del teatro clásico español) es una base de datos clave para el conocimiento de la actividad teatral de la época. Publicado por primera vez en 2008 en formato DVD (Ferrer Valls et al., 2008) y ahora disponible en abierto en red (Ferrer Valls et al., 2025), constituye una de las fuentes más amplias y críticas de información sobre profesionales del teatro español entre mediados del siglo XVI y comienzos del XVIII. DICAT reúne más de 5.000 entradas biográficas, que están siendo incrementadas y actualizadas constantemente, y están redactadas con un lenguaje formal y contemporáneo, lo que facilita su procesamiento automático. En este trabajo presentamos una propuesta metodológica inicial para aplicar técnicas de reconocimiento de entidades nombradas (NER[2]) sobre dicho corpus (que consta de unas 4.000.000 de palabras), empezando por la identificación de topónimos y empleando modelos de lenguaje basados en arquitectura Transformer —principalmente BETO (Cañete et al.,2020) y RoBERTa-BNE (Biblioteca Nacional de España & PlanTL,2021)— integrados en el ecosistema spaCy (Explosion AI, 2023). El

[1] Con la ayuda de los proyectos con referencia PID2022-136431NB-C61, PID2022-136431NB-C62, PID2022-136431NB-C65 y PID2022-136431NB-C66 (MCIU/AEI/FEDER, EU).
[2] Referencia al acrónimo en inglés *Named Entity Recognition*

objetivo de esta comunicación es presentar un primer ejemplo realizado sobre el actor y autor de comedias Andrés de Claramonte, de lo que se podría conseguir a largo plazo al mostrar, a partir de una casuística concreta, cómo la tecnología lingüística de última generación puede usarse para beneficiar la investigación humanística ayudando a enriquecer semánticamente los datos y así facilitar la apertura de nuevas vías de investigación sobre la movilidad de los actores y directores, de las redes profesionales y los espacios de representación.

El reconocimiento de entidades nombradas se ha convertido en una tarea fundamental dentro del Procesamiento de Lenguaje Natural (PLN), con aplicaciones que abarcan desde la extracción de información hasta la construcción de grafos de conocimiento (Ehrmann et al.,2021). En el ámbito de las HD, ya se ha mostrado su potencial en combinación con teorías de red para analizar personajes literarios, prosopografías históricas y correspondencias intelectuales (Barros, 2020). Sin embargo, la mayoría de estos trabajos se han centrado en el inglés o en entornos textuales del siglo XX, dejando relativamente inexplorado el patrimonio documental hispánico. La incorporación de modelos preentrenados en español —como BETO, RoBERTa-BNE o BERTIN (de la Rosa et al., 2022)— abre la puerta a superar esta brecha, proporcionando embeddings contextuales que captan matices semánticos y sintácticos propios del castellano. Como trabajo futuro nos planteamos vincular automáticamente menciones de lugares con coordenadas geográficas, apoyándonos en servicios de georreferenciación como GeoNames o OpenStreetMap, facilitando visualmente el análisis espacio-temporal.

Para el caso concreto de DICAT, Badía (Badía, 2021) ya esbozó propuestas metodológicas donde proponía la visualización de redes como una vía prometedora para estudiar las relaciones entre actores y compañías y así hacer visibles estructuras sociales, vínculos profesionales y núcleos de poder simbólico en el ecosistema teatral. Sin embargo, aquella propuesta se basaba todavía en un etiquetado manual de las entidades. La incorporación de técnicas de PLN representará un paso decisivo hacia esa dirección, ya que permite escalar la extracción de relaciones sin depender exclusivamente del etiquetado manual o del conocimiento experto puntual, abriendo así el camino hacia una explotación sistemática y replicable del corpus. El trabajo que presentamos pretende, haciendo uso de ese conocimiento, automatizar esa fase inicial de detección de topónimos, aumentando la escala y la replicabilidad del análisis.

Hay que tener en cuenta que DICAT fue concebido con el objetivo de compilar información dispersa sobre profesionales del teatro español del Siglo de Oro y cada entrada biográfica sigue una estructura cronológica en la que se detallan los eventos más relevantes de la vida del sujeto —nacimiento, contrataciones, representaciones, viajes, vínculos con compañías, etc.—. Su migración a MySQL durante la segunda fase del proyecto ASODAT ha permitido exponer los datos a formatos abiertos. Del total del corpus, aproximadamente el 82% de los registros incluyen menciones explícitas a ciudades, villas o espacios teatrales. El carácter contemporáneo y académico de la redacción implica que

se emplean nombres canónicos de personas y lugares (e.g., "Sevilla", "Madrid", "Écija"), con escasas variaciones ortográficas para el caso de los lugares, lo que reduce la carga de normalización previa, pero con una gran variabilidad para el caso de las personas, motivo por el cual empezamos las pruebas abordando el caso de los topónimos que ya de por sí conlleva su complejidad. Además, la redacción de muchas noticias de cada biografía se hizo repitiendo ciertos patrones sintácticos para facilitar al lector la identificación de la información esencial de cada noticia, lo que ayuda al proceso de automatización que aquí proponemos.

En términos técnicos, el trabajo plantea desafíos significativos llevados a cabo combinado herramientas de PLN (como *spaCy*), clasificadores entrenados con anotaciones específicas y estrategias híbridas de desambiguación basadas en reglas y aprendizaje automático.

En la metodología de procesamiento de lenguaje natural, el flujo de trabajo se inicia con la extracción de los textos desde MySQL y su conversión a UTF-8. Se aplica una tokenización respetuosa con las reglas de la Real Academia Española, seguida de una segmentación por oraciones mediante el modelo de dependencia de spaCy para español. Dado que los textos son actuales, no se requiere una normalización histórica intensiva; sin embargo, se procede a la eliminación de marcas tipográficas y a la corrección automática de abreviaturas (p. ej., "S." → "San"), con el fin de optimizar el reconocimiento de entidades.

Una vez tenemos los textos preparados, se han realizado las pruebas usando dos diferentes modelos para llevar a cabo la tarea de reconocimiento de las entidades nombradas (NER), los cuales han sido seleccionados por que fueron entrenados exclusivamente en español y por la información obtenida de la bibliografía:

- ***BETO*** (Cañete et al.,2020), es una versión de BERT entrenada sobre Wikipedia y C4-es, que ha demostrado un F1 superior al 90 % en el corpus CoNLL-2002.
- ***RoBERTa-BNE*** (Biblioteca Nacional de España & PlanTL,2021), entrenado con 570 GB de textos históricos y modernos, cuya fine-tuning para NER (Capitel) alcanza valores de F1 comparables a BETO, pero con mejor cobertura léxica en dominios literarios.

Además, ambos modelos se integran en spaCy mediante la clase `spacy-transformers`, lo que permite aprovechar la infraestructura de tuberías, anotadores y visualizadores de la librería (Explosion AI, 2023). La categoría de entidad `LOC` se puede refinar a subcategorías internas (`CITY`, `REGION`, `COUNTRY`) para facilitar en un futuro la georreferenciación, mientras que se ha creado una entidad `TIT` que se emplea para almacenar los títulos de las obras que se han identificado en el texto y que por tanto están relacionados con la compañía teatral o actor que se está analizando, y que en nuestro caso de estudio es Andrés de Claramonte, lo cual nos va a permitir establecer una relación por años de los lugares y títulos de las obras. Por otra parte, se tratará de ampliar la relación de los topónimos con acciones llevadas a cabo en dichos

lugares como pueden ser firmas de contratos o conciertos para representaciones y que serán identificadas por entidades `VERB`.

Para evaluar el sistema, se ha construido un conjunto de validación manual, anotado por especialistas en teatro áureo. La métrica principal será F1 macro para las clases `LOC` y `TIT. Se comparará el rendimiento de BETO y RoBERTa-BNE. Además del análisis cuantitativo, se realizará una inspección cualitativa de los casos de error, prestando especial atención a topónimos con polisemia (p. ej., «Granada» como ciudad frente a «Reino de Granada») y nombres propios o apodos que también contienen nombres de lugares («Miguel Granada», «la Comadre de Logroño»). El objetivo es testear la capacidad de estas herramientas para lograr unos resultados fiables y útiles para los especialistas en teatro áureo, y perfilar sus posibilidades para un uso sistemático que abarque el conjunto de la información contenida en DICAT.

Si finalmente los resultados resultan favorables, dicha combinación de NER y georreferenciación representaría una estrategia eficaz para enriquecer las biografías de DICAT y abrir nuevas líneas de investigación en el teatro clásico español. Al aprovechar un lenguaje contemporáneo y bien normativizado, se minimizan los obstáculos típicos de la lengua histórica y se maximizan los beneficios de las herramientas PLN de última generación. Este enfoque demuestra la vocación transformadora de las Humanidades Digitales, que no se limita a la preservación del patrimonio, sino que aspira a su relectura crítica a través de metodologías computacionales. No se trata solo de conocer mejor los datos ya existentes, sino de activar un proceso de acción y transformación del conocimiento humanístico, que democratice el acceso a la información, explore grandes cantidades de datos ya disponibles desde perspectivas que solo son factibles mediante acercamientos digitales, estimule nuevas preguntas y fomente modelos colaborativos de análisis e interpretación. En última instancia, nuestro objetivo para el futuro es transformar una base de datos biográfica en una herramienta dinámica de lectura relacional y crítica de la historia del teatro español, poniendo de manifiesto el compromiso cívico de las Humanidades Digitales con la recuperación, la comprensión y la reapropiación del patrimonio cultural.

Bibliografía

BADÍA, J. (2021), «La transformación de DICAT: el potencial de la visualización digital de redes para el teatro clásico español», en *Lope en Red. Bases de datos y tecnologías digitales aplicadas al teatro clásico*, Reichenberger, pp. 35–55. https://doi.org/10.5209/tret.70303

BARROS GARCÍA, B. B. (2020), «El texto literario hecho datos: F. M. Dostoievski en el marco de las Humanidades digitales y los enfoques cuantitativos», *Revista de Teoría de la Literatura y Literatura Comparada*, n.º 23, pp. 53–77. https://doi.org/10.1344/452f.2020.23.3

BIBLIOTECA NACIONAL DE ESPAÑA & PLANTL (2021), *RoBERTa-base-bne* [modelo preentrenado en español], Hugging Face. https://huggingface.co/PlanTL-GOB-ES/roberta-base-bne

CAÑETE, J., CHAPERON, G., FUENTES, R., & PÉREZ, J. (2023), «Spanish pre-trained BERT models and evaluation data», *arXiv:2308.02976*. https://arxiv.org/abs/2308.02976

DE LA ROSA, J., PONFERRADA, E. G., VILLEGAS, P., GONZÁLEZ DE PRADO SALAS, P., ROMERO, M., & GRANDURY, M. (2022), «BERTIN: Efficient pretraining of a Spanish language model using perplexity sampling», *arXiv:2207.06814*. https://arxiv.org/abs/2207.06814

EHRMANN, M., HAMDI, A., PONTES, E. L., ROMANELLO, M., & DOUCET, A. (2021), «Named Entity Recognition and Classification on Historical Documents: A survey», *ACM Computing Surveys*, 56(2), art. 27, pp. 1–47. https://dl.acm.org/doi/10.1145/3604931

EXPLOSION AI (2023), *spaCy: Industrial-strength Natural Language Processing in Python*, recurso en línea. https://spacy.io

FERRER VALLS, T., et al. (2008), *Diccionario biográfico de actores del teatro clásico español (DICAT)* [DVD + libro], Reichenberger.

FERRER VALLS, T., et al. (2025), *Diccionario biográfico de actores del teatro clásico español (DICAT)*, recurso en línea. https://dicat.uv.es

Anotación y detección de las emociones en cartas de mujeres del siglo XVI en español: los retos actuales

Albina Sarymsakova
Patricia Martín-Rodilla
Instituto de Estudios Gallegos Padre Sarmiento – IEGPS-CSIC-XuGa

Palabras clave: Reconocimiento de emociones; historia; anotación; mujeres; cartas
Keywords: Emotion recognition; history; annotation; women; letters

Introducción

La minería de emociones es un campo que ha captado el interés de los investigadores en la última década (Liu et al., 2010; Zhao, Liu, & Xu, 2016; Chatterjee, Narahari, Joshi, & Agrawal, 2019; Wang, Ma, Yu, Xia, & Cambria, 2024) y se basa en enfoques establecidos para la extracción de características textuales, como el análisis semántico y las técnicas de inteligencia artificial, que permiten la creación de sistemas basados en perfiles emocionales (Koptyra, Ngo, Radliński, & Kocoń, 2023) desarrollados principalmente en psicología durante las décadas de 1970 y 1980 (Ekman, 1992) para su aplicación en los ámbitos como el marketing, los servicios de salud y las redes sociales, ayudando a investigadores y organizaciones a comprender los aspectos emocionales, tomar decisiones informadas y mejorar los servicios. Hasta ahora, el análisis de emociones ha mostrado resultados prometedores, especialmente con datos en inglés (Maks & Vossen, 2011), debido a la amplia disponibilidad de recursos en este idioma. Sin embargo, son pocos los estudios que han explorado la dimensión emocional en otros tipos de fuentes textuales o en diferentes idiomas.

En este contexto, nos centramos en una aportación metodológica significativa, ya que contribuye a la minería de emociones en fuentes históricas mediante el desarrollo de un sistema de etiquetación de emociones para este dominio específico. Dicho sistema no solo pretende facilitar y precisar la anotación de emociones de los textos históricos, sino también mejorar la detección automática de emociones en estos textos usando los grandes modelos de lenguaje (LLM).

En esta línea, nuestro estudio se enfoca en identificar los avances recientes y los desafíos en la anotación y la detección de emociones con LLM dentro de la investigación histórica, enmarcada en el campo de las Humanidades Digitales.

Para ello, seleccionamos un estudio de caso centrado en la anotación de emociones en un corpus epistolar de autoría femenina del siglo XVI, recopilado en el marco del proyecto CARTER@S. Este estudio representa un avance en varios aspectos clave de nuestra investigación: por un lado, en el desarrollo y evaluación de una metodología adecuada para la anotación de textos históricos; y por otro, en la evaluación del desempeño de los grandes modelos de lenguaje en esta misma tarea.

Con el fin de avanzar en estos aspectos, se realizaron las siguientes tareas:

1. Anotación manual de un corpus epistolar del siglo XVI escrito por mujeres compuesto por 20 cartas.
2. Cálculo del acuerdo entre anotadores para validar la calidad de la anotación y para ajustar el sistema y las instrucciones relativos al etiquetado de las emociones.
3. Evaluación de métodos y herramientas de LLMs existentes, así como de su eficacia en la detección automática de emociones usando nuestro corpus epistolar anotado.

Metodología

En esta aproximación inicial para la creación del sistema de anotación de emociones para el dominio textual histórico hemos seleccionado el corpus de 20 cartas escritas o recibidas por las mujeres del siglo XVI, procedentes del proyecto CARTER@S, como ya lo habíamos mencionado previamente. La motivación tras la selección de esta base de datos está condicionada por la necesidad del desarrollo de las herramientas para la investigación en el ámbito de la Humanidades Digitales que permitan validar o profundizar en las hipótesis de los trabajos desarrollados en los ámbitos interdisciplinares que abarcan los estudios históricos, filológicos, antropológicos y lingüísticos.

El equipo de los anotadores consta de los especialistas y expertos en distintos ámbitos, procedentes de dos centros de investigación (Instituto de Estudios Gallegos Padre Sarmiento y Centro de Ciencias Humanas y Sociales, ambos del Consejo Superior de Investigaciones Científicas). En total contamos con 10 anotadores, expertos en historia, lingüística, ciencias de computación, documentalistas y filólogos. En comparación con enfoques basados exclusivamente en anotaciones de expertos lingüistas, el uso de este enfoque para la anotación de emociones en el corpus de correspondencia histórica nos ha permitido desarrollar un sistema inicial robusto. Este sistema considera la complejidad de la transmisión emocional desde una perspectiva histórica, lingüística y computacional, lo que facilita su futura aplicación en otros textos históricos.

El proceso de anotación se llevó a cabo de forma individual; es decir, cada anotador recibió el conjunto de cartas sin etiquetas y las anotó manualmente. El conjunto de etiquetas iniciales incluye la clasificación emocional de Ekman (1992) junto con dos categorías nuevas introducidas por nosotros (esperanza y

nostalgia), condicionadas por el estudio previos del contenido textual de las cartas. Hasta donde sabemos, éste es el primer intento de introducir la esperanza y la nostalgia como categorías de emoción dentro de un sistema de anotación, en particular para textos epistolares históricos.

Tras la anotación de nuestro corpus, precedemos al cálculo del acuerdo entre anotadores mediante el coeficiente Kappa de Cohen, usando la propuesta de Artstein R. & Poesio M. (2008). Empleamos el sistema de anotación multi etiqueta (Muhammad et. al., 2025) que permite detectar, si se dan los casos, más de una emoción por el enunciado, lo que, en consecuencia, permite captar la riqueza y la complejidad de las emociones que contiene el corpus epistolar analizado.

Por último, realizamos los experimentos con LLM existentes, como GPT y DeepSeek, para la tarea de predicción de emociones a partir de las cartas históricas. Para ello empleamos la técnica de Zero-shot prompting propuesta por Palatucci et. al (2009).

Resultados

Nuestro proyecto se distingue por las siguientes aportaciones clave:

1. Hemos desarrollado un sistema robusto para la anotación de textos históricos, respaldado por un alto porcentaje de acuerdo entre anotadores, a pesar de la diversidad de sus perfiles, como se muestra en la Tabla 1 . A partir de los resultados del acuerdo entre anotadores por categorías emocionales, hemos refinado y mejorado las instrucciones para las futuras anotaciones.
2. Al aplicar el sistema de anotación multietiqueta, pudimos identificar algunos patrones preliminares. Por ejemplo, la combinación de ira y tristeza suele expresar una estrategia comunicativa de reproche, mientras que la combinación de ira y miedo tiende a construir una premisa de amenaza. Aunque estos patrones aún deben validarse en relación con el contenido completo de nuestro corpus, los consideramos provisionales, ya que contribuyen a una mejor comprensión de las emociones en contexto y, por tanto, facilitan su anotación.

	Ira	Asco	Miedo	Alegría	Tristeza	Sorpresa	Nostalgia	Promedio
IEGPS	82.87	97.45	90.51	70.83	66.44	94.21	96.53	**85.55**
CCSH	60.88	67.82	82.64	65.51	53.94	85.42	76.16	**70.34**

Tabla 1. Porcentaje del acuerdo entre anotadores del Instituto de Estudios Gallegos Padre Sarmiento y el Centro de Ciencias Humanas y Sociales.

3. Los resultados preliminares con LLM indican que los errores cometidos por estos modelos están, en gran medida, relacionados con desplazamientos semánticos, cambio en el significado de las unidades léxicas condicionado por la variación diacrónica, en el contenido léxico

específico de este dominio textual. Las pruebas también muestran que, entre los LLM evaluados, DeepSeek-v-3 es el que presenta mayor precisión en la tarea de anotación de textos históricos, utilizando la técnica de evaluación zero-shot.

En cuanto a las limitaciones de nuestro trabajo, una de las principales dificultades ha sido la recopilación de datos textuales históricos, ya que muy pocas fuentes están digitalizadas y, entre ellas, una proporción aún menor es de acceso libre. Además, el uso de grandes modelos de lenguaje implica otra limitación significativa, ya que la mayoría de estos modelos operan como cajas negras, lo que dificulta la interpretación de sus predicciones. Entre las futuras líneas de investigación se contempla la ampliación de nuestro conjunto de datos, incorporando una mayor cantidad de textos históricos y extendiéndolo hacia un enfoque multilingüe. Asimismo, se prevé la exploración de modelos con arquitecturas más complejas, que puedan mejorar el rendimiento en la extracción y clasificación de características textuales.

Agradecimientos

Este proyecto está financiado por el proyecto Momentum del Consejo Superior de Investigaciones Científicas (CSIC), que procede del Mecanismo de Recuperación y Resiliencia de la Unión Europea-Next Generation, en el marco de la Invitación General de la entidad pública empresarial Red.es para participar en los programas de atracción y retención del talento dentro de la Inversión 4 del Componente 19 del Plan de Recuperación, Transformación y Resiliencia. Los autores están sumamente agradecidos a Cristian David López Ushakov, Miguel García-Fernández, Javier Martínez Escobar, Magdalena Caso-López, Yara Mostazo Fernández, Sergio Bravo Sánchez, Teresa Abejón Peña, y Judith Farré, por su apoyo técnico, teórico, y su contribución a la anotación del corpus.

Bibliografía

ARTSTEIN, R., & POESIO, M. (2008), «Inter-coder agreement for computational linguistics», *Computational Linguistics*, 34(4), pp. 555–596.

CHATTERJEE, A., NARAHARI, K. N., JOSHI, M., & AGRAWAL, P. (2019), «SemEval-2019 Task 3: EmoContext—contextual emotion detection in text», en *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*, pp. 39–48.

EKMAN, P. (1992), «An argument for basic emotions», *Cognition & Emotion*, 6(3–4), pp. 169–200.

KOPTYRA, B., NGO, A., RADLIŃSKI, Ł., & KOCOŃ, J. (2023), «CLARIN-EMO: Training emotion recognition models using human annotation and ChatGPT», en *International Conference on Computational Science (ICCS)*, pp. 365–379.

LIU, B., et al. (2010), «Sentiment analysis and subjectivity», en *Handbook of Natural Language Processing* (2.ª ed.), pp. 627–666.

MAKS, I., & VOSSEN, P. (2011), «A verb lexicon model for deep sentiment analysis and opinion mining applications», en *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2011)*, pp. 10–18.

MUHAMMAD, S. H., et al. (2025), «Brighter: Bridging the gap in human-annotated textual emotion recognition datasets for 28 languages», preprint, arXiv:2502.11926.

PALATUCCI, M., POMERLEAU, D., HINTON, G. E., & MITCHELL, T. M. (2009), «Zero-shot learning with semantic output codes», *Advances in Neural Information Processing Systems*, 22.

WANG, F., MA, H., YU, J., XIA, R., & CAMBRIA, E. (2024), «SemEval-2024 Task 3: Multimodal emotion cause analysis in conversations», preprint, arXiv:2405.13049.

ZHAO, J., LIU, K., & XU, L. (2016), *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*.

Escritura, afectos y legitimación: redes sociales de autoras españolas (1890-1920)

Gema Duarte Abós
Universitat Rovira i Virgili

Palabras clave: Epistolarios; escritoras españolas; exilio; redes sociales; visualización de redes sociales
Keywords: Letters; Spanish women writers; exile; social media; social media visualization

La presente comunicación aborda el estudio de las redes sociales de escritoras españolas nacidas entre 1890 y 1920, en un contexto de profundas transformaciones marcadas por la emergencia del feminismo, la institucionalización de la cultura de masas y la fractura de la Guerra Civil. Partimos de la hipótesis de que las relaciones afectivas, profesionales y literarias que las autoras establecieron, documentadas en cartas, memorias y autobiografías, funcionaron como mecanismos de sostenimiento, validación y continuidad de la escritura femenina frente a los discursos patriarcales de exclusión.

Se propone una metodología mixta combinando el análisis cualitativo de epistolarios personales con el Análisis de Redes Sociales (ARS), utilizando herramientas como Gephi para representar gráficamente las conexiones entre autoras. Los resultados muestran la existencia de nodos clave (como Carmen Conde o María Teresa León), la persistencia de vínculos entre exilio e interior, y la importancia de relaciones —amistosas, epistolares o literarias— en la sostenibilidad de los proyectos autorales. Los grafos y gráficos cuantitativos que visualizan dichas redes permiten reubicar a las escritoras en el centro del sistema literario. Nos centraremos en las relaciones de algunas de las protagonistas dado que se trata de una investigación en curso.

Introducción y definición del problema

El papel jugado por las escritoras españolas nacidas entre finales del siglo XIX y principios del XX ha sufrido históricamente los efectos de una canonización excluyente, una dispersión de fuentes y una narrativa de silencios que ha invisibilizado la dimensión relacional de su escritura. Frente al archivo tradicional, los textos personales —cartas, diarios, memorias, autobiografías— ofrecen una alternativa no solo documental, sino epistemológica. Estos testimonios permiten reconstruir las redes de afecto, colaboración y legitimación que hicieron posible

la persistencia de la escritura femenina en contextos hostiles, como la dictadura franquista o el exilio.

El problema de investigación reside en cómo articular metodológicamente la información dispersa contenida en esos documentos para trazar un mapa de relaciones que visibilice las estrategias de subsistencia, visibilidad y consolidación de una autoría femenina moderna.

Marco teórico

Desde una perspectiva de estudios de género y sociología de la literatura, se parte de una concepción relacional de la autoría. El sujeto escritor no se define únicamente por su obra publicada, sino por los vínculos que construye y que lo sostienen simbólicamente. La red, entendida como espacio de reconocimiento, de validación y de circulación de capital simbólico (Bourdieu, 1995), es especialmente útil para estudiar a las autoras que no accedieron a las estructuras oficiales de legitimación literaria.

En línea con los trabajos del Proyecto BIESES - Bibliografía de Escritoras Españolas, liderado por María Martos y Nieves Baranda de la UNED, y las propuestas de Franco Moretti (2011), entendemos que el Análisis de Redes Sociales permite traducir relaciones textuales y paratextuales en estructuras visuales que revelan patrones de asociación, distribución de poder y formas de agencia autoral.

Metodología

Se ha adoptado una metodología mixta basada en tres fases:

- **Fase documental:** vaciado de fuentes primarias (epistolarios, memorias, autobiografías, dedicatorias, colaboraciones en prensa y revistas) de 30 autoras nacidas entre 1890 y 1920.
- **Fase analítica:** codificación de relaciones (personales, afectivas, editoriales, críticas) y categorización de agentes (escritoras, editoras, amigas, familiares, figuras culturales).
- **Fase visual y cuantitativa:** representación de redes mediante matrices de nodos y aristas en Gephi, aplicando medidas de centralidad (grado, intermediación) y detección de comunidades (algoritmos de modularidad).

Resultados

Redes personales y profesionales

Los grafos generados permiten identificar varios nodos de alta centralidad —como Carmen Conde, Consuelo Berges o María Teresa León— que ejercen funciones de mediación y articulación entre grupos. La presencia de redes mixtas (exilio-interior) revela la continuidad de vínculos más allá de la fractura bélica. Se destacan también comunidades afines, como la de mujeres vinculadas a la

Residencia de Señoritas, el Lyceum Club Femenino o la revista *Ínsula* en los años 40.

Tipologías de relaciones

Las relaciones de amistad, colaboración literaria o sororidad crítica aparecen como fundamentales para la supervivencia simbólica y material de las autoras. Las redes no responden a estructuras jerárquicas sino rizomáticas, fluidas y transversales.

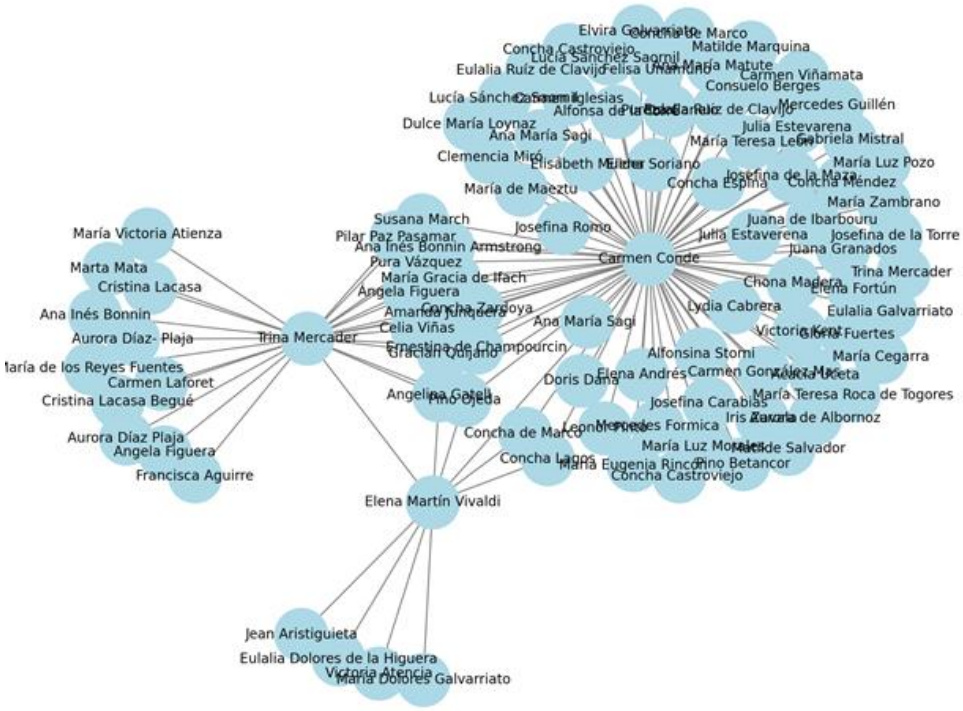
Visualizaciones

Un ejemplo ilustrativo es el de relaciones recuperadas a través de epistolarios entre Carmen Conde, Trina Mercader y Elena Martín Vivaldi, de manera particular y como parte del proyecto de investigación más ambicioso.

Tabla de métricas comparativas

Autora	Grado	Intermediación	Cercanía	PageRank
Carmen Conde	77	0.906	0.860	0.332
Trina Mercader	26	0.242	0.582	0.102
Elena Martín Vivaldi	10	0.089	0.529	0.040

El análisis de la red epistolar compuesta por Carmen Conde, Trina Mercader y Elena Martín Vivaldi permite observar las diferencias en sus posiciones estructurales dentro de la constelación de escritoras republicanas y del exilio. A través del uso combinado de las métricas de grado, intermediación, cercanía y PageRank, se configura un paisaje de relaciones que trasciende lo anecdótico y ofrece claves de lectura sobre la función histórica de estas autoras como nodos culturales. La primera dimensión analizada es el grado, que indica el número de conexiones directas de cada autora dentro de esta subred. Carmen Conde domina esta métrica de forma abrumadora, con 77 vínculos directos, frente a los 26 de Trina Mercader y los 10 de Elena Martín Vivaldi. Este indicador refleja su capacidad de interlocución directa con otras autoras, lo que la sitúa como el epicentro de una red de alta densidad, propia de quienes desempeñan funciones de articulación y acogida intelectual.



Más relevante aún es la centralidad de intermediación, que mide cuántos caminos entre dos nodos pasan por una autora determinada. Aquí también Carmen Conde presenta un valor extremadamente alto (0.906), lo que confirma su papel de puente estructural. La gran mayoría de relaciones indirectas dentro del grafo pasan por ella, lo que implica que fue una figura clave para la transmisión de información, vínculos afectivos y legitimaciones cruzadas entre escritoras.

Primeros análisis de resultados

El uso del ARS ha permitido visibilizar las relaciones ya conocidas e inferir la existencia de redes de cuidado y apoyo informales, pero determinantes. Las escritoras aparecen no como figuras aisladas o excepcionalmente insertas en el sistema, sino como parte de un entramado denso, afectivo e intelectual, que cuestiona los relatos individualistas de la autoría y propone un modelo colaborativo y coral de la creación literaria.

Además, se ha confirmado que los espacios periféricos o no oficiales (cartas privadas, dedicatorias, colaboraciones en prensa) constituyen verdaderos acerbos de información para la consolidación de la escritura femenina.

Una red densa y poliédrica: estructura general y dinámica epistolar

El grafo construido a partir de 251 escritoras y 485 vínculos epistolares revela una estructura de alta densidad: las conexiones son múltiples. En esta trama relacional, tupida y resistente, circulan ideas, afectos, complicidades y

estrategias de supervivencia ante el silencio impuesto por la Guerra Civil, el exilio, la censura y la posguerra. Lo notable es que, pese a los ello, sostuvieron un tejido comunicativo que operó como espacio de continuidad intelectual y emocional. La carta, como género y como práctica, permitió prolongar proyectos truncados, trazar genealogías posibles, y construir una comunidad literaria transnacional a contracorriente del canon dominante.

Nodos centrales: mujeres que sostienen la red

El análisis de centralidades confirma algo más que una distribución desigual de vínculos: revela la existencia de figuras que, desde su posición relacional, actúan como sostenedoras simbólicas y operativas de la red. Cabe destacar nombres como Carmen Conde emerge como vértice clave, articulando relaciones entre autoras del exilio y del interior; Consuelo Berges, María Teresa León o Gabriela Mistral por su papel estructural: su grado de intermediación indica que muchas otras, con menor visibilidad, accedían a la red a través de ellas; o Trina Mercader o Elisabeth Mulder, con relaciones diversas tanto en términos temáticos como generacionales. Muchas fueron verdaderas arquitectas de esta red: fundaron revistas, coordinaron editoriales, organizaron eventos, y —sobre todo— sostuvieron una memoria literaria femenina cuando otras formas de visibilidad estaban clausuradas.

Clústeres temáticos y generacionales: la promesa de Louvain

Aunque aún no se ha podido aplicar todavía el algoritmo de Louvain, el grafo sugiere la existencia de comunidades determinadas por afinidades ideológicas, estéticas y geográficas. En su momento, permitirá verificarlas, delimitarlas, colorear trayectorias y, sobre todo, estudiar las tensiones y solidaridades internas que configuran esta red.

Implicaciones históricas y simbólicas: la carta como tejido de memoria

La carta fue mucho más que un medio de comunicación: funcionó como tecnología afectiva, archivo íntimo y espacio de legitimación literaria. En ese sentido, cumplió tres funciones clave:

- **Legitimación literaria**: Las escritoras se citaban, se leían, se recomendaban. Se tejía así una crítica literaria informal, que escapaba a los circuitos académicos o editoriales tradicionales.
- **Cuidado y sororidad**: Las cartas eran consuelo, memoria compartida, acompañamiento.
- **Archivo alternativo**: Hoy constituyen fuentes primarias insustituibles. Permiten reconstruir una historia de la literatura escrita por mujeres que no es adyacente al canon, sino paralela, resistente y profunda.

Conclusiones

Además de mostrar el proceso de transformación de las fuentes y las posibilidades de análisis de un corpus digital, se evidencian los diferentes caminos de análisis de estas. Igualmente, se aborda la riqueza de los archivos como repositorios imprescindibles para recuperar fuentes epistolares privadas y cotidianas y se reflexiona sobre el futuro de la conservación en el entorno digital. Con relación a la cuestión de las escritoras podemos afirmar que:

- Las redes de relaciones personales y profesionales fueron fundamentales para la legitimación simbólica de las escritoras nacidas entre 1890 y 1920, especialmente en contextos de represión o exilio.
- El análisis relacional permite reubicar a las autoras en el centro del sistema literario, desmontando su posición de marginalidad.
- Las herramientas del ARS, combinadas con la lectura cualitativa de fuentes personales, abren nuevas vías de estudio y valorización del legado literario femenino en la España contemporánea.
- Este enfoque favorece la construcción de genealogías femeninas no lineales, donde la historia literaria se entiende como un tejido de afectos, resistencias y alianzas.

Bibliografía

BOURDIEU, P. (1995), *Las reglas del arte. Génesis y estructura del género literario*, Anagrama.

COTARELO-ESTEBAN, L., & COTARELO ESTEBAN, L. (2022), «Mujeres intelectuales españolas y sus redes sociales en el exilio estadounidense», *Journal of Spanish Cultural Studies*, 23(1), pp. 13–33. https://doi.org/10.1080/14636204.2022.2033416

IMÍZCOZ BEUNZA, J. M., & ARROYO RUIZ, L. (s. f.), «Redes sociales y correspondencia epistolar del análisis cualitativo de las relaciones personales a la reconstrucción de redes egocentradas», *Redes: Revista hispana para el análisis de redes sociales*, 21, pp. 98–138. https://doi.org/10.5565/rev/redes.419

JUÁREZ LÓPEZ, A. (2021), «El editor como autor: prácticas ecdóticas en textos epistolares», *Signa. Revista de la Asociación Española de Semiótica*, 30, pp. 501–531.

MORETTI, F. (2007), *Graphs, Maps, Trees: Abstract Models for Literary History*, Verso.

Progetto Mambrino: Tabula Imaginaria

Antonio Azaustre Lago
Stefano Bazzaco
Stefano Neri
Patricia Sánchez Migallón
Universidad de Verona

Tabula Imaginaria es un recurso de la Biblioteca Digital del Proyecto Mambrino de la Universidad de Verona desarrollado en el marco del Proyecto de Excelencia del Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona 2023/2027 *Inclusive Humanities* (sub project: Glocal Fiction in 16th Century Coord. Anna Bognolo) y del Proyecto Nacional PRIN 2022 *Chivalric Spaces* (prot. 2022HT3XYP P.I. Annalisa Perrotta).

El objeto de estudio del Proyecto Mambrino es un vasto corpus de novelas en prosa en italiano derivadas de los libros de caballerías ibéricos en forma de traducciones y, sobre todo, continuaciones e imitaciones originales. Debido a la extensión del corpus y a sus características materiales y literarias, la edición científica digital ha sido desde el principio la forma ideal elegida por el equipo del Proyecto Mambrino para editar, representar y estudiar de forma comparativa y sistemática estos textos. La Biblioteca Digital alberga actualmente las ediciones científicas digitales de seis libros pertenecientes al ciclo italiano del *Amadís de Gaula* y se va a extender a otros textos del mismo ciclo y del ciclo del *Palmerín* hasta llegar, en los próximos años a la edición completa del corpus. Cada edición permite visualizar la reproducción del original y su transcripción y cuenta con un aparato de recursos que permiten explorar el texto singularmente y en el conjunto del corpus, como por ejemplo las sinópsis argumentales de cada capítulo, el índice de los personajes, el índice de los motivos y el índice de los lugares narrativos. Dentro de este último índice semántico se encuentran registrados todos los topónimos que aparecen en la narración, cada uno vinculado con sus recurrencias en el texto y fichados de forma descriptiva para que el usuario pueda colocarlos en el entramado argumental. El recurso *Tabula Imaginaria* quiere reunir en una visión de conjunto todos los lugares indexados y representarlos gráficamente a través de dos opciones básicas: en un mapa actual que permite la geolocalización y en un mapa antiguo representativo de los conocimientos geográficos de la época. El poster pretende exponer el estado de la cuestión sobre la realización de este proyecto.

A partir de 2024 se ha llevado a cabo una larga labor de investigación sobre las diferentes aplicaciones de SIG disponibles y finalmente, se ha optado por dos visualizaciones con las herramientas Google MyMaps y Leaflet. En el primer

caso, se han mapeado todos los lugares en la aplicación Google MyMaps, creando una capa en el mapa para cada obra. De esta forma, se pueden visualizar los marcadores en el mapamundi actual, y a la vez se consiente descargar los datos relativos a las coordenadas geográficas específicas en archivos KML, fomentando la interoperabilidad con proyectos afines. La segunda visualización con Leaflet, por otra parte, consta de una librería en Java para introducir coordenadas en una imagen y visualizar el resultado del mapeado a través de un documento HTML. Esta opción se fundamenta en la idea de integrar en unos mapas antiguos los datos geográficos nombrados en las obras de interés. Para visualizar los espacios marcados en las ediciones digitales de entrada se eligió el mapa de Ptholomeus impreso en 1507 por Martin Waldseemüller, pero se prevé la integración con otros mapas. Ambas visualizaciones son compatibles con el contenido de las ediciones digitales y respetan los criterios de accesibilidad y sostenibilidad promovidos por el proyecto.

Con esta metodología, el proyecto contribuye a la circularidad de la investigación gracias al empleo de herramientas y datos pre existentes, asegurando su almacenamiento a largo plazo y su reutilización futura, dos conceptos presentes como ejes principales en el Data Management Plan europeo.

Bibliografía

BAZZACO, S., & BOGNOLO, A. (2024), «Editar libros de caballerías en la era digital: la Biblioteca Digital del Proyecto Mambrino», en *Editar el Siglo de Oro en la era digital*, Universitat Autònoma de Barcelona – Servei de Publicacions, pp. 19–47.

BAZZACO, S. (2024), «Revolucionar el acceso al patrimonio librario: los sistemas de HTR entre Humanidades Digitales y ciencia de la información», *Philologia Hispalensis*, 38(2), pp. 59–77.

DAMMANN, F., & KREMER, D. (eds.) (2024), *Geographical Research in the Digital Humanities: Spatial Concepts, Approaches and Methods,* Transcript Verlag. Recurso en línea: http://www.jstor.org/stable/jj.13083323.

DRUCKER, J. (2011), «Humanities Approaches to Graphical Display», *Digital Humanities Quarterly*, 5.

LOSADA PALENZUELA, J. L. (2019), «Mapeado digital de lugares en la novela bizantina española», *Artnodes*, 23, pp. 72–78. Recurso en línea: http://dx.doi.org/10.7238/a.v0i23.3222.

THOMAS, L. (2013), «Cartographic and Literary Intersections: Digital Literary Cartographies, Digital Humanities, and Libraries and Archives», *Journal of Map & Geography Libraries,* 9(3), pp. 335–349. Recurso en línea: https://doi.org/10.1080/15420353.2013.823901.

Visualización de patrones de uso del vocabulario 'mobiliario y enseres' en AMERLEX mediante Voyage Tools para acciones de transferencia

Mª Teresa Cáceres Lorenzo
José-Ignacio Salas-Cáceres
Universidad de Las Palmas de Gran Canaria – IATEX

Con el propósito de realizar acciones de transferencia sobre del uso de los americanismos desde una perspectiva histórica, llevamos a cabo esta investigación dentro del Proyecto de Humanidades Digitales AMERLEX (I+D+i PID2022-138801NB-I00) y (MICIU/AEI/10.13039/501100011033/ y FEDER, UE). Este proyecto de Humanidades Digitales busca construir una base de datos con un buscador eficiente, que no solo muestre ejemplos empíricos de estas palabras, sino que también desvele las tendencias léxicas que configuran el español americano. Para conseguir este fin, se planteó la necesidad de recoger léxico de una selección de obras de los siglos XVI al XVIII, además de asegurar cumplir el objetivo de identificar las áreas léxicas que en esta etapa presentan mayor número de americanismos y que varían en las dos centurias analizadas.

En el proyecto AMERLEX, las acciones de comunicación y difusión son fundamentales para la valorización y transferencia de nuestros resultados. Para lograrlo, utilizamos una variedad de formatos como vídeos, infografías y fichas tecnológicas, y los distribuimos a través de canales estratégicos como redes sociales y sitios web. Este enfoque no solo mejora la visibilidad de la actividad de transferencia de la universidad, sino que también genera un mayor impacto al asegurar que nuestros hallazgos lleguen a quienes más pueden aprovecharlos.

En esta comunicación se presentan los resultados de investigación obtenidos tras la identificación y análisis de ejemplos de términos relacionados con el campo semántico `mobiliario y enseres´ que se utilizan en los documentos áureos seleccionados en AMERLEX (https://amerlex.iatext.ulpgc.es/). La elección de este vocabulario especializado se fundamente en la evidencia de contar con un lexicón que llega a sumar más de100 lexías que permita identificar y comprender el uso de esas voces en más de 30 autores clásicos en los estudios diacrónicos hispánicos que escribieron sobre la América colonial. Ahora bien, uno de los objetivos del proyecto es encontrar tendencias en el uso del vocabulario específico que se puedan visualizar mediante los datos recogidos. El marco teórico publicado hasta el momento sobre cómo se configuró el léxico del español americano habla de la necesidad de realizar estudios que ofrezcan una visión

panorámica en distintos periodos con el propósito de dar respuesta a cuestiones como la preferencia por determinadas voces en un contexto en el que los autores disponían de una gran variedad de términos. A modo de ejemplo ilustrativo para denominar a una *hamaca* como objeto novedoso de los indígenas, los autores podían utilizar el marinerismo chinchorro o los vocablos románicos *catre*, *cama*, *litera* o *yacija*.

Ante esto se elige la herramienta *Voyant Tools* (http://voyant-tools.org/) por su capacidad de ofrecer de manera fácil varios recursos idóneos para responder a las siguientes preguntas de investigación con respecto a las designaciones de utensilios, muebles, instrumentos necesarios o convenientes en una casa: a) ¿qué lexías fueron las más usadas por los autores seleccionados? y b) ¿se mantienen algunas voces en los dos siglos?

La metodología de acceso a *Voyant Tools* se realizó con la subida a la herramienta de textos que provenían de AMERLEX, separados por centuria. Luego se analizaron los 28 recursos según su capacidad de ofrecer respuestas a las preguntas de investigación, hasta llegar a la herramienta de visualización *Bubbles* que nos permite visualizar la tendencia que se buscaba como resultado de la investigación para la acción de transferencia (Figura 1 y 2)

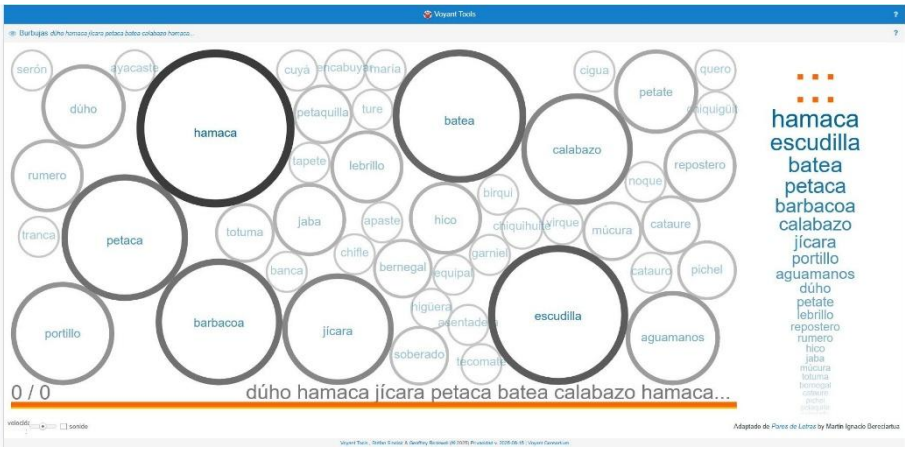


Figura 1. Tendencia del vocabulario de `mobiliario y enseres´ (siglo XVI)

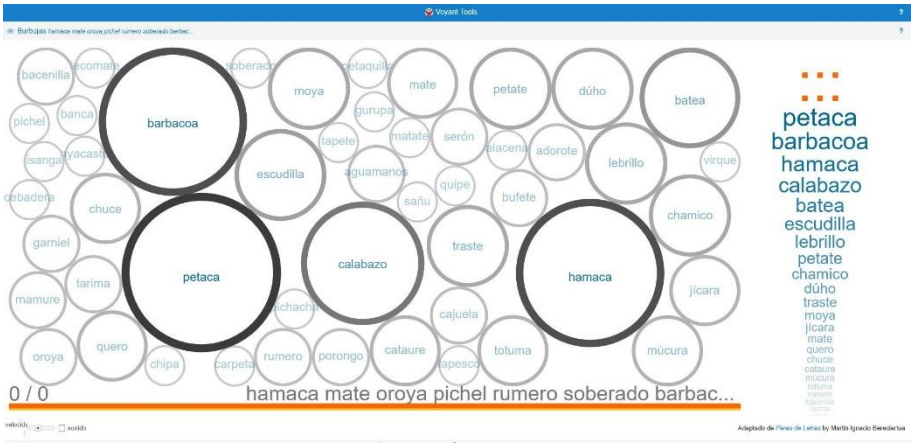


Figura 2. Tendencia del vocabulario de `mobiliario y enseres´ (siglo XVII)

A través de la observación de ambas imágenes es posible contestar a las preguntas propuestas y a otras que surjan de los participantes en la acción de transferencia.

Bibliografía

CÁCERES-LORENZO, M. T. (2025), «Antecedentes del léxico náutico americano en textos no especializados de los siglos XVI y XVII», *Rilce. Revista de Filología Hispánica*, 41(1), pp. 165–196. Recurso en línea: https://doi.org/10.15581/008.41.1.165-96.

SAMPSEL, L. J. (2018), «Voyant Tools», *Music Reference Services Quarterly*, 21(3), pp. 153–157. Recurso en línea: https://doi.org/10.1080/10588167.2018.1496754.

SMITH, A. M. (2024), «Sinclair, S., & Rockwell, G. 2016. Voyant Tools. http://voyant-tools.org/», *Journal of Web Librarianship*, 18(1), pp. 35–37. Recurso en línea: https://doi.org/10.1080/19322909.2024.2327877.

El uso de LogosLink para la unificación de datos discursivos y etnográficos

Beatriz Calderón-Cerrato
Paula Sánchez-Carrera
César González-Pérez

Palabras clave: LogosLink, análisis del discurso, IAT/ML, etnografía
Keywords: LogosLink, discourse analysis, IAT/ML, ethnography

El discurso es una de las principales herramientas del ser humano de construcción y propagación de fenómenos sociales. Mediante la comunicación, se trasladan las ideas que conviven alrededor de un fenómeno, por lo que el análisis del discurso se convierte en una disciplina adecuada para captar y comprender de manera profunda tales discursos y comprender los fenómenos sociales aparejados. La realización de estos estudios requiere un conocimiento más o menos profundo sobre el contexto en el que se desenvuelve el discurso, necesitando obtener información sobre quiénes lo producen, ante quién y en qué circunstancias. Esta información permite desambiguar significados e interpretar el discurso de manera más fidedigna. Una de las soluciones a este respecto pasa por una investigación multidisciplinar, incluyendo técnicas propias de la etnografía. Mediante la propuesta etnográfica del trabajo en/desde la propia comunidad es posible conocer las dinámicas y las relaciones que se establecen en ella, así como explicitar aquello que no se menciona, que queda oculto/ado, pudiendo así añadir al análisis discursivo una valiosa información contextual. El trabajo de campo, *in situ*, del etnógrafo/a lo/la sitúa en una situación metodológica como es «la observación participante», la cual, «implica adentrarnos profundamente en situaciones sociales y mantener un papel activo, así como una reflexión permanente. Estar atento a los detalles, sucesos, eventos e interacciones» (Hernández et al., 2014, p. 399). Esta idea de *estar presente* –de observar– se convierte en un elemento de gran relevancia discursiva, puesto que contribuye a una mayor profundización y ampliación de los comportamientos verbales y/o conductuales, que en muchas ocasiones no tienen lugar durante la realización de entrevistas (Díaz de Rada, 2011).

Sin embargo, a menudo el contexto es difícil de captar y ordenar en relación con los datos discursivos que se obtienen, debido a diferentes problemas. Por una parte, es necesario unificar los diferentes datos recogidos que permiten relacionar los textos con aquello que se ha observado en el campo. Por otra parte, en el proceso de análisis discursivo resulta necesario dar a conocer de qué manera el contexto está influyendo en la interpretación de determinadas

declaraciones textuales (Widdowson, 2004). Así, el problema observado en ambas disciplinas tiene que ver con la unificación de la información discursiva y contextual, así como la transparencia durante todo el proceso analítico.

La solución que proponemos es el uso de la herramienta de software LogosLink, basada en la metodología de análisis del discurso IAT/ML (www.iatml.org) (González-Pérez *et al*., 2024). De acuerdo con esta metodología, el componente contextual es imprescindible a lo largo de varias de las tareas que la conforman. Así, IAT/ML parte de un modelo contextual, que recoge los temas, posturas y agentes que están siendo estudiados y que caracterizan el contexto social de los discursos. Sobre esta descripción del contexto se añade información de otros tipos que describe los distintos modelos discursivos que se empleen (Figura 1).

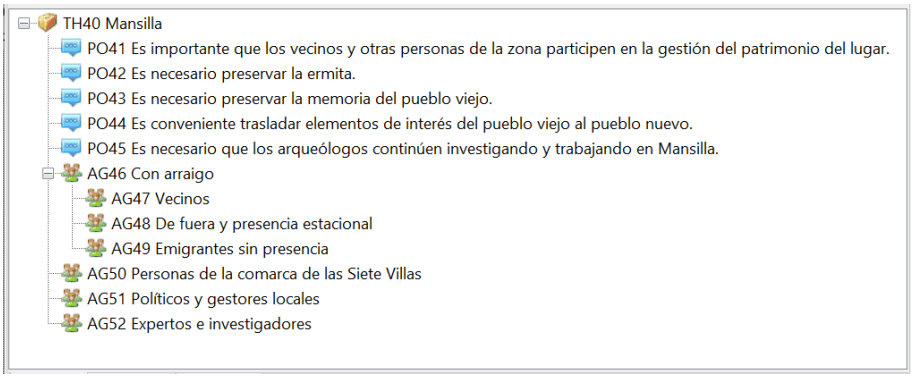


Figura 1. Modelo de contexto de un proyecto de análisis de discurso en LogosLink.

Uno de los pilares de IAT/ML es el análisis ontológico, mediante el que se representan las entidades que aparecen aludidas en los textos y las relaciones que se establecen entre ellas. Esta tarea tiene concomitancias con el desarrollo de sociogramas en etnografía. LogosLink permite unificar cada entidad del sociograma o modelo ontológico al texto mediante el uso de denotaciones, pudiendo así mantener vinculados texto y contexto (Figura 2).

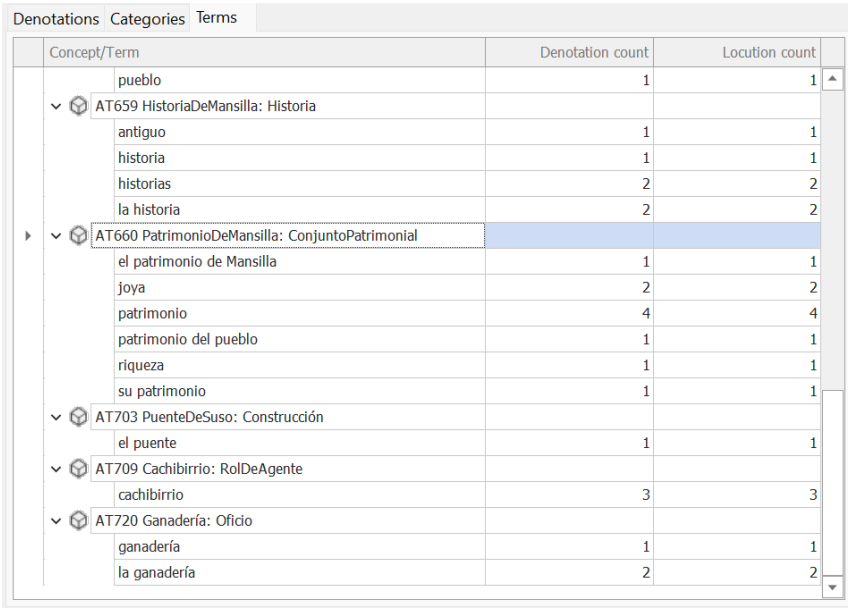


Figura 2. Analítica de denotaciones en LogosLink.

Asimismo, IAT/ML incluye también el análisis agencial, mediante el que se exploran los estados mentales de los hablantes (como creencias, deseos e intenciones) plasmados en el texto a través de preguntas como *¿Qué rol juega cada agente?* o *¿Qué está permitido o prohibido para cada agente?* Es difícil conocer esta información sin atender al contexto, por lo que LogosLink incluye formas específicas que permiten documentar el contexto al que se alude para responder este tipo de preguntas (Figura 3).

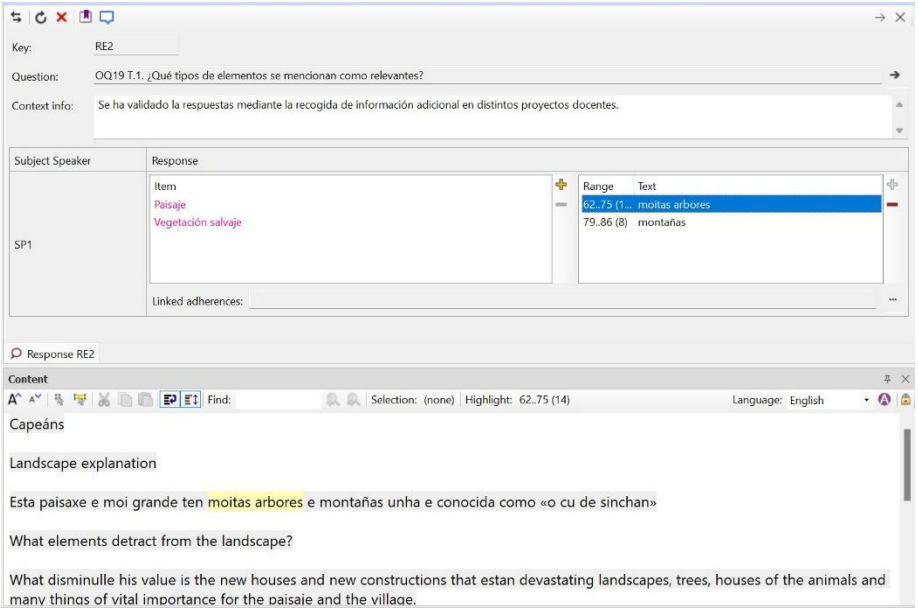


Figura 3. Información contextual (campo “Context info”) relativa al análisis agencial en LogosLink.

Una vez que los elementos ontológicos y agenciales se han conectado con el contexto de estas maneras, es posible realizar analíticas automáticas sobre el

conjunto de la información, para determinar, por ejemplo, el grado de adhesión a diferentes posturas por parte de diferentes agentes, pudiendo incorporar de manera cualitativa si los apoyos mostrados son más o menos fuertes (Figura 4).

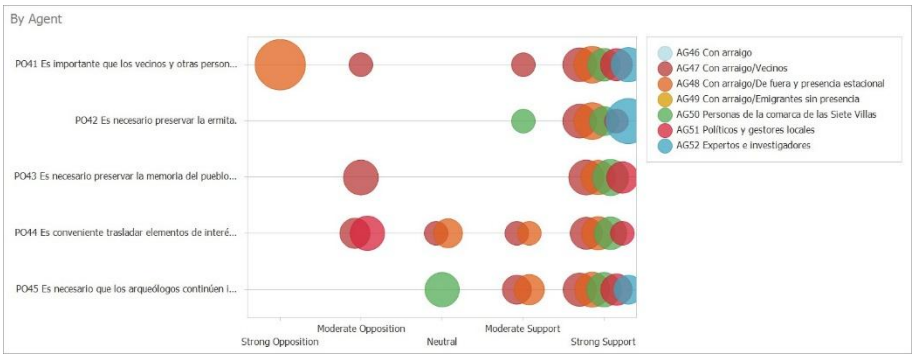


Figura 4. Analítica de adhesión a posturas en LogosLink.

De este modo, LogosLink funciona como herramienta centralizadora de información discursiva y etnográfica, facilitando los estudios interdisciplinares entre estos campos.

Bibliografía

HERNÁNDEZ, R., FERNÁNDEZ, C., & BAPTISTA, M. P. (2014), *Metodología de la investigación*, México D.F., McGraw Hill Education.

DÍAZ DE RADA, Á. (2011), *El taller del etnógrafo: Materiales y herramientas de investigación en etnografía*, Universidad Nacional de Educación a Distancia (UNED).

WIDDOWSON, H. (2004), *Text, Context, Pretext. Critical issues in discourse analysis*, Blackwell Publishing.

GONZÁLEZ-PÉREZ, C., PEREIRA-FARIÑA, M., CALDERÓN-CERRATO, B., & MARTÍN-RODILLA, P. (2024), «IAT/ML: A Metamodel and Modelling Approach for Discourse Analysis», *Software and Systems Modelling*, recurso en línea: https://doi.org/10.1007/s10270-024-01208-7

Metodología de Impresión 3D para recintos amurallados: aplicación y evaluación en Toledo y Algeciras

Daniel Castilla Rodríguez
Alfredo Miguel Maximiano Castillejo

Palabras clave: Metodología digital; impresión 3D; patrimonio urbano; reconstrucción virtual; simulación 3D; inteligencia artificial; Toledo; Algeciras.
Keywords: Digital methodology; 3D printing; urban heritage; virtual reconstruction; 3D simulation; artificial intelligence; Toledo; Algeciras.

Este trabajo presenta una metodología técnico-analítica para el modelado y la impresión 3D de estructuras defensivas urbanas, con aplicación directa en el ámbito del patrimonio histórico-arquitectónico. La propuesta parte de una necesidad común en el campo de las Humanidades Digitales: construir modelos reconstructivos que sean no solo visualmente verosímiles, sino también técnicamente reproducibles, conceptualmente trazables y metodológicamente críticos, de acuerdo con los planteamientos de Demetrescu & Ferdani (2021).

La metodología se estructura en cinco fases iterativas, interconectadas en un circuito de evaluación empírica. Estas son: (1) recopilación crítica de documentación histórica y arqueológica, (2) modelado tridimensional con software libre (Blender), (3) simulación técnica avanzada, (4) fabricación física mediante impresión FDM (filamento fundido) y (5) evaluación con herramientas de inteligencia artificial para detección de errores geométricos y estructurales. Esta secuencia permite abordar la reconstrucción del patrimonio desde un enfoque no lineal, en el que cada fase retroalimenta a las demás, dando lugar a un sistema evolutivo de aprendizaje técnico.

El modelado se realiza sobre bases cartográficas históricas geo-referenciadas, integrando análisis estratigráfico, proporciones defensivas y alineaciones urbanas específicas. A partir de esta base, el entorno digital permite recrear volúmenes y geometrías con criterios transparentes, documentando cada decisión morfológica. Esta información no solo queda registrada en el propio archivo de modelado, sino que se incorpora a un esquema de metadatos que acompaña al modelo final.

Una vez definido el modelo digital, se somete a una fase de simulación técnica tridimensional, en la que se estudia el comportamiento estructural del objeto en entornos de *slicing* como Ultimaker Cura y otros motores de previsualización. Aquí se analizan factores como la generación de soportes, la eficiencia del patrón de relleno, el comportamiento térmico y las trayectorias de extrusión. En este punto se incorpora además un sistema de asistencia basada

en IA, que permite detectar automáticamente errores topológicos, zonas de tensión estructural o duplicaciones de malla, optimizando así tanto la fidelidad visual como la integridad física del modelo.

La impresión se lleva a cabo con parámetros ajustados en función del comportamiento previsto: tipo de relleno (*gyroid*, cúbico, lineal), orientación adaptativa, altura de capa según la morfología, uso o no de soportes y material de impresión. Estos criterios han sido estandarizados mediante un conjunto de heurísticas técnicas sistematizadas en un *checklist* replicable en otros contextos patrimoniales (Oropallo & Piegl, 2016).

Una innovación metodológica destacada es la implementación de un sistema de carcasas intercambiables, que permite representar diferentes fases históricas sobre una misma base topográfica impresa. Esto habilita una lectura diacrónica del espacio urbano, optimiza el uso de material y facilita la comparación estructural entre épocas, en sintonía con enfoques sostenibles como los propuestos por Neumüller et al. (2014).

La metodología ha sido aplicada a los recintos amurallados de Toledo (Toletum–Tulaytula) y Algeciras (Iulia Traducta–Al-Yazirat al-Jadra), seleccionados por su relevancia arqueológica, diversidad cronológica y complejidad topográfica. Esta aplicación ha permitido identificar cómo las condiciones urbanas y defensivas específicas de cada caso influyen directamente en los criterios técnicos del modelado e impresión. En el caso de Toledo, el relieve escarpado ha requerido segmentaciones verticales complejas; en Algeciras, la regularidad del trazado ha facilitado configuraciones más eficientes.

Más allá de su dimensión técnica, esta metodología ofrece una vía sólida para la investigación académica, la formación universitaria, la divulgación museográfica y la mediación patrimonial. Su uso de software libre, bajo coste material y compatibilidad con equipos accesibles (como impresoras Ender-3 o K1C) la convierten en una herramienta transferible, adaptable a distintos entornos y escalas.

En definitiva, esta propuesta contribuye al debate metodológico de las Humanidades Digitales mediante un enfoque crítico, empírico y evolutivo de la reconstrucción virtual. Al poner el énfasis no solo en el resultado, sino en el proceso técnico y documental que lo sustenta, se reivindica una forma de modelado 3D que no es meramente visual, sino verificable, documentada y abierta al escrutinio científico.

Bibliografía

DEMETRESCU, E., & FERDANI, D. (2021), «From field archaeology to virtual reconstruction: A five steps method using the extended matrix», *Applied Sciences*, 11(11), art. 5206. Recurso en línea: https://doi.org/10.3390/app11115206.

KENNY, G. (2012), «An introduction to Moustakas's heuristic method», *Nurse Researcher*, 19(3), pp. 6–11. Recurso en línea: http://dx.doi.org/10.7748/nr2012.04.19.3.6.c9052.

NEUMÜLLER, M., REICHINGER, A., RIST, F., & KERN, C. (2014), «3D printing for cultural heritage: Preservation, accessibility, research and education», en M. IOANNIDES & E. QUAK (eds.), *3D Research Challenges in Cultural Heritage*, Springer, pp. 119–134. Recurso en línea: https://doi.org/10.1007/978-3-662-44630-0_9.

OROPALLO, W., & PIEGL, L. A. (2016), «Ten challenges in 3D printing», *Engineering with Computers*, 32, pp. 135–148. Recurso en línea: https://doi.org/10.1007/s00366-015-0407-0.

Buscador de recursos TEISEL: de la ciencia abierta a las buenas prácticas en la investigación sobre segundas lenguas

Mar Cruz Piñol
Universitat de Barcelona
Sophie McBride
Universitat Politècnica de València
Carmen López Ferrero
Universitat Pompeu Fabra

Palabras clave: Humanidades digitales; buenas prácticas; segundas lenguas; ciencia abierta; buscador de recursos digitales
Keywords: Digital humanities; best practices; second languages; open science; digital resource search engine

Esta propuesta parte de la necesidad que tenemos los investigadores de acceder a experiencias previas de buenas prácticas en el uso de tecnologías para la investigación en L2, así como de entender cuál ha sido el proceso de elaboración de esas herramientas y de contar con una valoración crítica contrastada de su alcance. Hasta dónde llega nuestro conocimiento, no existía un recurso que diera respuesta a esta necesidad, por lo que hace unos meses nos planteamos incorporar en la revista TEISEL un nuevo sistema de navegación hipertextual que conectara las publicaciones de TEISEL con las mencionadas necesidades, de acuerdo con los tres tipos de trabajos que acoge la revista: 1) investigaciones en las que se emplean tecnologías, 2) creación de herramientas y 3) reseñas sobre libros y portales relacionados con el uso de tecnologías para la investigación.

Con el objetivo de lograr esta navegación hipertextual se ha diseñado una base de datos que indexa las herramientas digitales citadas en las publicaciones de TEISEL: https://revistes.ub.edu/index.php/teisel/SEARCH Esta base de datos permite la búsqueda por palabras clave, por el nombre de los recursos o por el DOI de los artículos y, gracias a un buscador, se posibilita que los investigadores accedan de manera rápida y eficiente a consideraciones metodológicas sobre herramientas, *software*, corpus y otros recursos relevantes en el ámbito de la investigación en segundas lenguas. En consecuencia, este buscador facilita la localización y el acceso a los recursos mencionados en las publicaciones de la revista, promoviendo la reutilización de los recursos analizados y la colaboración entre investigadores.

La metodología empleada incluyó, en una primera fase, la revisión

exhaustiva de los artículos publicados en TEISEL, la extracción (manual y automatizada; en el segundo caso, asistida por inteligencia artificial) de referencias a recursos y la implementación de una interfaz (con la ayuda de Flourish) que permite consultas eficientes. Tras esa primera fase, actualmente hemos incorporado en las pautas de la revista la petición de que los autores, junto con la versión final de sus contribuciones, entreguen un listado de todas las herramientas que han analizado en sus trabajos, acompañada cada una de dos o tres palabras clave que la definan. Se trata, pues, de una información, a modo de metadatos y en el contexto de la ciencia abierta, que enriquece cada una de las publicaciones, el conjunto de la revista TEISEL y la investigación futura en general.

El resultado, tras cerrar el volumen 4 (primavera de 2025), es una base de datos de más de 100 recursos que se han citado en 30 trabajos publicados en TEISEL. En el póster que proponemos, se presentará el diseño, el desarrollo y la diversidad de lecturas que brinda el buscador de recursos de TEISEL (https://revistes.ub.edu/index.php/teisel/SEARCH).

Figuras, tablas, mapas o gráficas

Figura 1. Buscador de recursos del sitio web de la revista TEISEL. Fuente: Universitat de Barcelona (s.f.).

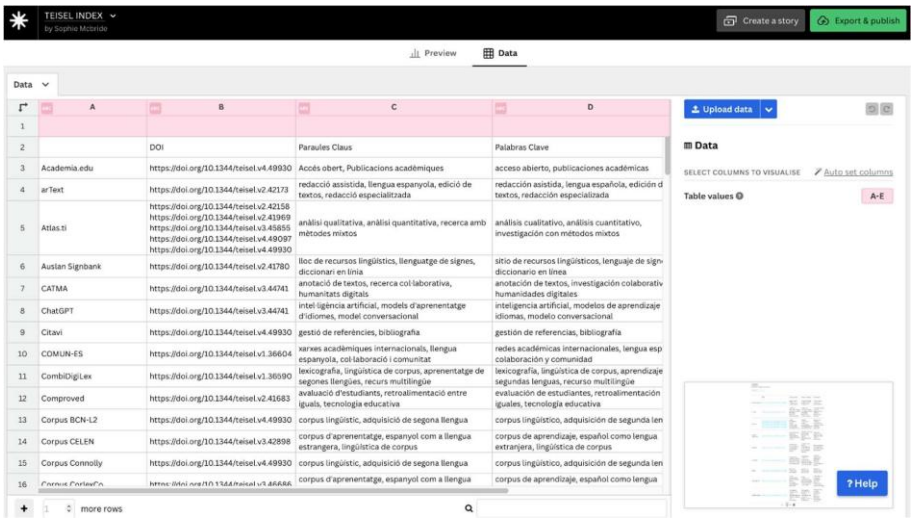


Figura 2. Visualización de datos realizada con la herramienta Flourish (https://flourish.studio).

Bibliografía

ALLAN, R. N. (2009), *Virtual Research Environments*, Chandos.

BARNETT, R. (2000), «University Knowledge in an Age of Supercomplexity», *Higher Education*, 40(4), pp. 409–422. Recurso en línea: https://doi.org/10.1023/A:1004159513741.

BEERMANN, D., & MIHAYLOV, P. (2011), «e-Research for Linguists», en *5th ACL HLT Workshop*, ACL, pp. 24–32. Recurso en línea: https://dl.acm.org/doi/10.5555/2107636.2107640.

CRUZ PIÑOL, M. (ed.) (2021), *e-research y español LE/L2*, Routledge.

CRUZ PIÑOL, M., et al. (2025), «Revista TEISEL», en *I Jornada sobre Humanitats Digitals de la UB*. Recurso en línea: https://hdl.handle.net/2445/220431.

FLOURISH (s. f.), *Flourish*, recurso en línea: https://flourish.studio/.

GOLDER, S. A., & HUBERMAN, B. A. (2006), «Usage Patterns of Collaborative Tagging Systems», *Journal of Information Science*, 32(2), pp. 198–208. Recurso en línea: https://doi.org/10.1177/0165551506062337.

PEÑA CERVEL, M. S. (2010), «Herramientas para la investigación lingüística», en *Teoría lingüística: métodos, herramientas y paradigmas*, Ramón Areces, pp. 51–107.

PEÑA-LÓPEZ, I. (2008), «The Personal Research Portal», *Technology Innovation Management Review*. Recurso en línea: https://www.timreview.ca/article/123.

TEIRA, C., & POLO, N. (2021), «Digitalización y recursos para la investigación en Lingüística», *RSEL*, 51(1), pp. 157–176. Recurso en línea: http://dx.doi.org/10.31810/RSEL.51.1.9.

WILLIAMS, F., et al. (1988), *Research Methods and the New Media*, The Free Press/Macmillan. Recurso en línea: https://archive.org/details/researchmethodsn0000will.

La dieta ideal en el más allá: análisis iconográfico y digitalización de tumbas egipcias

Inmaculada Delage González
Beatriz Garrido Ramos
Antonio Criado Martín
Juan Francisco Romero Collazos
UNIR
Esteban García Viñas
IAPH
Myriam Seco Álvarez
US
Daniel Becerra Fernández
UCM

Palabras clave: Alimentación egipcia; iconografía; fotogrametría; base de datos; Humanidades Digitales
Keywords: Egyptian food; iconography; photogrammetry; database; Digital Humanities

Esta propuesta de póster expone los primeros avances del proyecto de investigación *"Alimentación en el Antiguo Egipto como Paradigma de la dieta Mediterránea" (DEAMED)*, financiado en el marco de la convocatoria I+D de Jóvenes Talentos de la Universidad Internacional de La Rioja (UNIR). Su objetivo es reconstruir el concepto de "dieta ideal" en el Antiguo Egipto a través del análisis iconográfico de diversas escenas funerarias de las tumbas privadas de Luxor (TT, *Theban Tomb*) del Reino Nuevo (ca. 1550–1070 a.C.), y contrastarlo con los principios nutricionales contemporáneos, especialmente con el modelo de la dieta mediterránea ideal.

Para la consecución de este fin, DEAMED se inscribe en el campo de las Humanidades Digitales, al combinar el uso de tecnologías de documentación visual y trabajar los datos a partir de metodologías y estándares novedosos que difieren de los tradicionales, con el objetivo de generar conocimiento científico accesible, reproducible y de interés tanto para el ámbito académico como para la sociedad. El equipo investigador está conformado por especialistas en historia, arqueología, nutrición, biología y humanidades digitales, lo que permite un abordaje transversal y complementario del objeto de estudio.

En la primera fase del proyecto, en la que nos encontramos actualmente, el trabajo se ha centrado en la recogida y organización sistemática de datos fotográficos y fotogramétricos procedentes de un corpus de treinta tumbas, así

como en la digitalización de las obras. En cada una de ellas se ha documentado la representación de alimentos en las mesas de ofrendas, mesas concebidas simbólicamente para acoger los víveres que el difunto llevaría al más allá y que, por lo tanto, reflejan una idealización de cómo deseaba alimentarse por la eternidad.

El objetivo de este primer estadio es realizar un análisis iconográfico detallado de los elementos presentes en estas escenas. Cada componente alimenticio representado (panes, frutas, aves, pescado, cerveza, etc.) es descrito y codificado según criterios taxonómicos, simbólicos y nutricionales. Este proceso permite descomponer las representaciones en unidades analíticas que se integran en una base de datos digital. El software seleccionado es Omeka, una herramienta diseñada específicamente para colecciones digitales y exposiciones virtuales, la cual permite estructurar y estandarizar la información gracias al uso de metadatos Dublin Core. En este entorno se ha comenzado una base de datos estructurada que incluye información sobre cada objeto identificado: su naturaleza, localización en el interior de la tumba, contexto arqueológico y geolocalización dentro de la necrópolis, fuentes documentales y gráficas asociadas y, cuando sea posible, enlaces a modelos tridimensionales de las tumbas disponibles en plataformas como Sketchfab. Esta integración de fuentes heterogéneas busca favorecer una experiencia de consulta rica, contextualizada y abierta.

Los objetivos de esta propuesta de digitalización son múltiples. En primer lugar, facilitar el trabajo colaborativo del equipo mediante una sistematización rigurosa y estandarizada de la información. En segundo lugar, ofrecer a la comunidad científica un corpus estructurado y reutilizable que pueda servir como base para futuras investigaciones sobre la alimentación, el ritual funerario o la iconografía egipcia. Y en tercer lugar, permitir que esta información sea accesible también al público general interesado en el patrimonio y en la historia de la alimentación, fomentando así la divulgación del conocimiento.

En la presente propuesta se muestra cómo se ha aplicado esta metodología en dos tumbas, TT40 y TT277, de cronología y estructura diversas. Estas han servido como casos piloto para poner a prueba la eficacia de la base de datos y confirmar que se cumple el primer objetivo de la investigación: la sistematización rigurosa y accesible de la información. A partir de este punto, los siguientes pasos consisten en continuar con el volcado completo de datos de las restantes tumbas para avanzar hacia una fase de producción científica que combine el potencial de las Humanidades Digitales con el análisis histórico, arqueológico, iconográfico y nutricional.

Aunque el proyecto se encuentra aún en una fase temprana, sus primeros resultados permiten vislumbrar el potencial transformador de las Humanidades Digitales en el estudio de sociedades del pasado. La combinación de documentación visual de alta calidad, análisis iconográfico sistemático y herramientas digitales de código abierto constituye una base sólida para futuras

etapas del proyecto, que incluirán el análisis nutricional de los alimentos identificados y su comparación con los principios de una dieta saludable actual.

Figuras, tablas, mapas o gráficas

Figura 1. Recogida de datos en la TT277

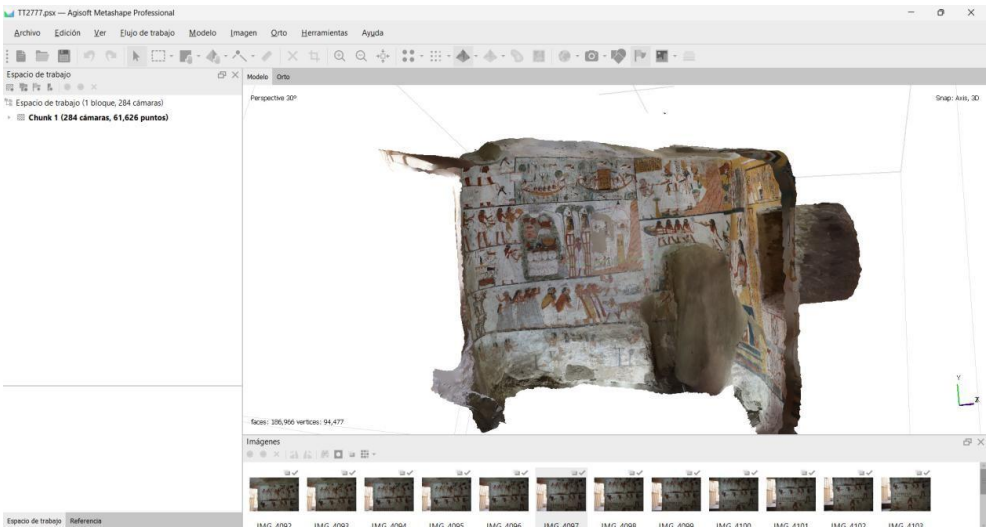


Figura 2. Fotogrametría de la TT277

Figura 3. Detalle de la mesa de ofrendas analizada en la TT277

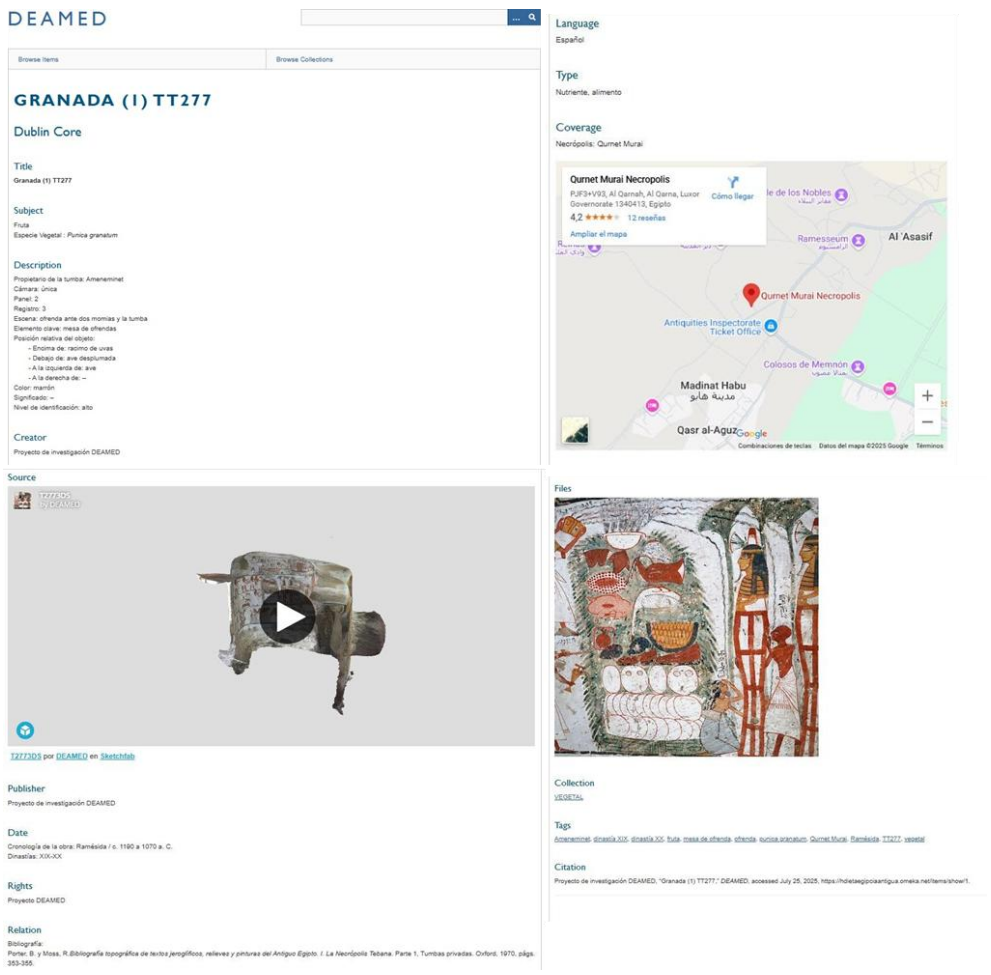


Figura 4. Ejemplo de entrada en Omeka de la granada en la mesa de ofrendas de la TT277

Bibliografía

ALEGRE GARCÍA, S. (2013), *El arte en el antiguo Egipto. Claves para su interpretación*, Cuenca, Alderabán.

GARRIDO-RAMOS, B. (2023), «Propuesta de Metadatos Dublin Core para esculturas egipcias: Creación de una colección digital en Omeka», *Revista Anales de la Universidad de Cuenca*, 62, pp. 35–49.

GARRIDO-RAMOS, B. (2017), «Visualización de datos y clasificación. Iconclass: Un estudio de caso desde la perspectiva de las Humanidades Digitales», *UCOARTE*, 10, pp. 9–18.

IKRAM, S. (1995), *Choice Cuts. Meat Production in Ancient Egypt*, *Orientalia Lovaniensia Analecta*, 69.

MÉNDEZ-MARTÍNEZ, J. Á. (2015), «Definición de Humanidades Digitales», *ArtyHum. Monográfico de Humanidades Digitales*, 1.

PORTER, B., & MOSS, R. L. B. (1960), *Topographical Bibliography of Ancient Egyptian Hieroglyphic Texts, Reliefs, and Paintings*, Vol. I, Oxford, Griffith Institute/Ashmolean Museum/Oxford University Press.

Rasgos fonéticos y léxicos del español canario en las cartas de Rosalía Franchi (1820-1821)

Miguel Antonio Di Novella[1]
Universidad de La Laguna
Universidad de Las Palmas de Gran Canaria

Palabras clave: Español canario; Epistolografía; Léxico diferencial; Humanidades digitales; Siglo XIX
Keywords: Canarian Spanish; Epistolography; Differential lexicon; Digital humanities; 19th century

La presente propuesta analiza un corpus compuesto por 180 cartas manuscritas autoría de Rosalía Franchi, noble tinerfeña del siglo XIX, dirigidas a su esposo Juan Máximo de Franchi durante los años 1820-1821. Estas cartas, accesibles a través de la plataforma *Corpus Documental de las Islas Canarias* (CORDICan), constituyen un testimonio del español canario decimonónico, al registrar tanto rasgos fonéticos como léxicos propios de esta variedad dialectal en un contexto íntimo y no normativizado.

La investigación se plantea como una aproximación sociolingüística en el que se articula el análisis tradicional filológico (centrado en fenómenos de variación fonética y léxica) con herramientas propias de las Humanidades Digitales. En este sentido, se hace uso de entornos digitales como Voyant Tools (véase fig. 1) para explorar cuantitativamente el corpus y la biblioteca spaCy para su etiquetado morfosintáctico automatizado. Además, se integra el *Diccionario Histórico del Español de Canarias* (DHECan) como fuente lexicográfica, de manera que se verifique la validez y frecuencia de los canarismos documentados.

[1] Estudiante del programa de doctorado en Territorio y Sociedad. Evolución Histórica en un Espacio Tricontinental (África, América y Europa)

Figura 1. Nube de palabras del corpus de cartas de Rosalía Franchi, realizado con Voyant Tools.

El estudio identifica fenómenos fonéticos relevantes: seseo, aspiración y pérdida de /s/, yeísmo, betacismo, confusión de j/g, neutralización entre r/l y omisión de la h. Estos rasgos se manifiestan en la escritura de la autora mediante grafías no normativas, que revelan la oralidad subyacente. Junto a estas formas, se detectan casos de hipercorrección (como *ermocísimo* por *hermosísimo*) que evidencian inseguridad ortográfica influida por la norma.

En el plano léxico, se documentan voces canarias como *papa, millo, bubango, medianero* o *limetón*. Estas formas se clasifican por campos semánticos (alimentación, trabajo, relaciones sociales) y se contrastan con las entradas del DHECan. Su frecuencia de uso, así como la coocurrencia con otros términos del campo léxico, se visualiza mediante gráficos generados a partir del corpus.

Desde el punto de vista verbal, se observa una preferencia clara por el uso del pretérito perfecto simple (como *recibí*, *dije*, *vino*) frente al perfecto compuesto (*he dicho*, *ha venido*), lo que concuerda con patrones característicos del español atlántico (véase fig. 2). Este rasgo, no obstante, se encuentra matizado por la formación madrileña de la autora, generando una hibridación lingüística de notable interés.

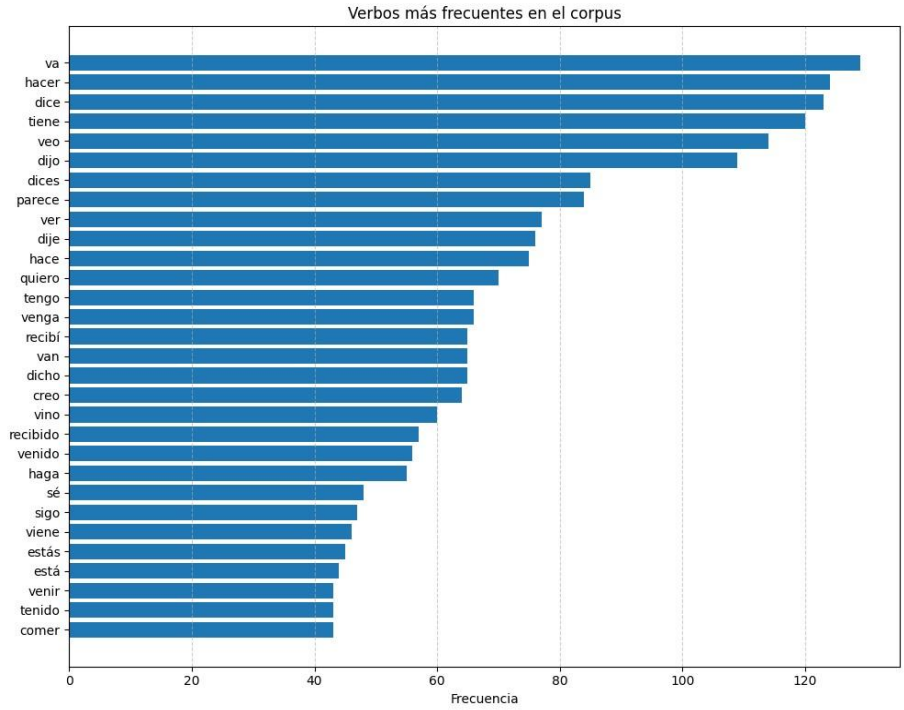


Figura 2. Verbos más frecuentes en el corpus.

Asimismo, se constata una alta presencia de diminutivos afectivos y marcadores discursivos como *pues* (véase fig. 3), lo cual refuerza la dimensión emocional e interpersonal de la correspondencia.

Figura 3. Uso de «pues».

Desde una perspectiva sociolingüística, estas cartas permiten no solo describir usos dialectales, sino también interpretar cómo la lengua escrita de Rosalía de Franchi refleja una identidad tensionada entre la pertenencia al entorno isleño y el deseo de proyectar una imagen culta y diferenciada. En una de las cartas, la autora afirma explícitamente que no es “isleña”, lo que revela

una voluntad consciente de distanciamiento lingüístico y social. Este componente identitario dota al corpus de un valor añadido para los estudios de género, representación social y configuración de normas.

En definitiva, este estudio demuestra que la correspondencia privada es una fuente privilegiada para el análisis de la variación lingüística, permitiendo observar fenómenos de contacto entre norma y habla vernácula. La autora, como hablante no normativa, pero alfabetizada, constituye un ejemplo paradigmático del español canario del XIX y de cómo este se proyecta en la escritura. Los resultados obtenidos abren la posibilidad de futuras investigaciones comparativas con otros corpus insulares o con correspondencia femenina de la misma época.

Bibliografía

CORBELLA DÍAZ, D., & CORRALES ZUMBADO, C. (2004), «Primeros testimonios e impresiones sobre el habla canaria», *Anuario de Estudios Atlánticos*, 50, pp. 71–120. Recurso en línea: https://revistas.grancanaria.com/index.php/aea/article/view/837.

CORBELLA DÍAZ, D., & VIÑA BRITO, A. (2019), «Corpus Documental de las Islas Canarias: Un nuevo reto en humanidades digitales», *Anuario de Estudios Atlánticos*, 65, pp. 1–17. Recurso en línea: https://revistas.grancanaria.com/index.php/aea/article/view/10283/9694.

CORRALES ZUMBADO, C., & CORBELLA DÍAZ, D. (2013), *Diccionario histórico del español de Canarias* [DHECAN], San Cristóbal de La Laguna, Instituto de Estudios Canarios. Recurso en línea: https://webfrl.rae.es/dhecan.html.

LLERENA CASTRO, M. (1959), *Las sibilantes en la época de la colonización de Tenerife (estudiadas en las Datas)*, San Cristóbal de La Laguna, Universidad de La Laguna. Tesina dirigida por Álvaro Galmés. Recurso en línea: https://riull.ull.es/xmlui/handle/915/23900.

TRAPERO, M. (2018), *Diccionario de Toponimia de Canarias: los guanchismos*, Santa Cruz de Tenerife, Idea.

Preservación y digitalización, un viaje de 150 años por el Archivo Histórico de la Sociedad del Gran Teatro del Liceo de Barcelona

Ana Escañuela Barradas
Universidad Autónoma de Barcelona

Palabras clave: Archivo Histórico de la Sociedad del Gran Teatro del Liceo; digitalización; documentación teatral; historia de la ópera; patrimonio cultural
Keywords: Historical Archive of the Gran Teatro del Liceo Society; digitization; theatrical documentation; history of opera; cultural heritage

Problema de investigación

El Archivo Histórico de la Sociedad del Gran Teatro del Liceo incluye más de 50.000 documentos generados durante la actividad continua del teatro desde 1837 hasta 1999. Durante estos 150 años en los cuales la Sociedad gestionó el Liceo, se produjo una gran cantidad de documentación muy diversa, como la derivada de la gestión del propio teatro, que organizaba temporadas de ópera y ballets y una importante cantidad de representaciones teatrales y de conciertos extraordinarios. Además, el Liceo fue la sede de una intensa actividad social, en la que destacan los bailes de máscaras anuales, organizados por la propia Sociedad. El fondo abarca momentos clave, como la inauguración del teatro en 1847, el incendio y reconstrucción de 1861 o la guerra civil española.

Este archivo proporciona una visión extraordinaria de la historia de un teatro de ópera y su impacto en la cultura de Barcelona. Como documentación única, debía digitalizarse para su preservación, su difusión y su acceso por parte de la comunidad investigadora y de toda la sociedad.

Metodología

El año 2011, la Sociedad del Gran Teatro del Liceo y la UAB firmaron un convenio para abrir al público el Archivo.

La Biblioteca de Humanidades forma parte fundamental en la realización de este proyecto, que se plasma de la siguiente manera: en una primera etapa, colaborando en la selección de la documentación y coordinando la posterior digitalización, que realiza una empresa externa. En una segunda etapa, con la clasificación, catalogación y publicación de la documentación en el Depósito Digital de Documentación de la UAB (DDD).

La documentación se organiza en diez grandes series:

- Administración del teatro (documentación administrativa, contable, jurídica y de contratación)
- Direcció del teatro
- Documentación técnica (proyectos arquitectónicos y de diseño)
- Escenografías (fotografías y dibujos en acuarela)
- Carteles
- Fotografías
- Libretos (principalmente del siglo XIX)
- Partituras (muchas, inéditas)
- Programas de mano (temporadas de ópera, conciertos, ballets, actividades extraordinarias)

Los diferentes materiales se relacionan entre ellos: a partir de una escenografía aislada de una ópera concreta se puede acceder al resto de documentación relacionada con esta obra (programas de mano, fotografías, partituras, etc.).

Resultados obtenidos

Este proyecto destaca por su valor interdisciplinario y la colaboración entre el personal investigador y técnico de la UAB y la Sociedad. El archivo se encuentra en constante crecimiento gracias a la constante digitalización y la incorporación de nuevos materiales. Su consulta en línea permite una comprensión profunda de la evolución artística, técnica y social de uno de los teatros de ópera más importantes del mundo.

La c**onsulta** del Archivo en el DDD es la siguiente:

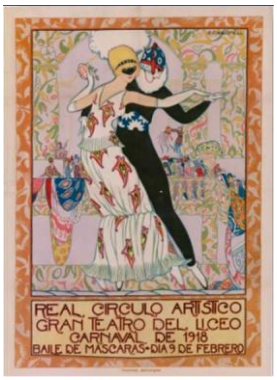
- Entre 2023 y 2025: 811.192 consultas, con 580.266 descargas
- Año 2024: 100.927 consultas, con 48.771 descargas

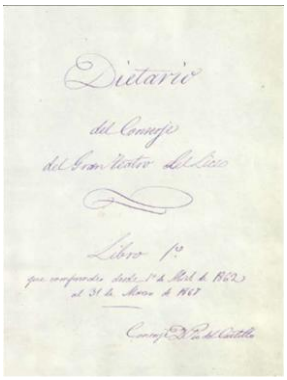
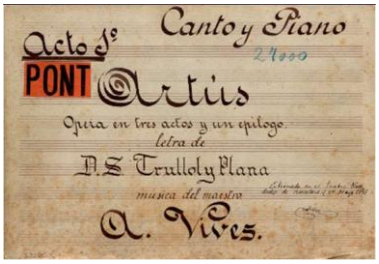
Datos disponibles en https://ddd.uab.cat/usage.py?c=societatliceu&report=usage&ln=es .

La **difusión** del Archivo se hace principalmente a través de:

- DDD de la UAB (https://ddd.uab.cat/collection/societatliceu?ln=es)
- Web (https://www.bib.uab.cat/human/arxiusocietatliceu/publiques/indexcas.php), que presenta y difunde los fondos del Archivo
- Difusión en redes sociales y en la web del Servicio de Bibliotecas de la UAB
- Presentación de trabajos y resultados en la reunión anual de la Sociedad del Gran Teatro del Liceo
- Participación en la Red de Humanidades Digitales de la UAB (https://www.uab.cat/web/proyectos-en-curso-1345837647802.html)

Figuras, tablas, mapas o gráficas





Bibliografía

CLAVÉ, J. A. (2023), *L'Aplec del Remei: sarsuela de costums catalanes bilingüe en dos actes* (ed. crítica de F. Cortès i Mir; edición del libreto de M. Comas Lamarca; trad. de L. Poltrony).

DAUFÍ MUÑOZ, A. (2022), *De la transgressió carnavalesca al silenci de la postguerra. Estudi musical, social, cultural i econòmic dels balls de màscares al Gran Teatre del Liceu (1848–1936)*, recurso en línea.

SOCIETAT DEL GRAN TEATRE DEL LICEU (2022), *Un recorregut a través del seu Arxiu Històric* (coord. F. Cortès), Barcelona, Huygens Editorial.

De Diccionarios históricos a Ontologías: El "Abub" y la Organología del S. XIX en LexiMus

Eugenia Gallego Cañellas
María Ordiñana Gil
Universidad Internacional de Valencia
Gorka Rubiales
Musikene: Centro Superior de Música del País Vasco

Palabras clave: Ontologías musicales; lexicografía musical; datos enlazados; organología; siglo XIX
Keywords: Musical ontologies; musical lexicography; linked data; organology; 19th century

Los diccionarios musicales históricos constituyen fuentes esenciales para reconstruir el universo organológico del siglo XIX. Diccionarios como el de Felipe Pedrell (1894), Carlos José Melcior (1859), Antonio Fargas y Soler (1852) o José Parada y Barreto (1868) recopilan una vasta cantidad de términos referidos a instrumentos musicales procedentes de diversas culturas, lugares y épocas. Este estudio, enmarcado en el proyecto LexiMus (Universidad de La Rioja; PID2022-139589NB-C31-C32-C33), cuyo objetivo es crear una ontología musical del siglo XIX, presenta un proceso metodológico para la identificación, extracción y catalogación de instrumentos musicales incluidos en dichas fuentes. Este proceso implica: 1) la generación de una base de datos estructurada con descriptores detallados, esenciales para la construcción de la ontología y su aplicabilidad en sistemas de catalogación; y 2) la validación rigurosa por parte de un musicólogo experto. Para su desarrollo se han combinado herramientas de inteligencia artificial (IA), programación y validaciones humanas expertas. Algunos procedimientos clave han sido:

- **Extracción asistida por IA:**
  Mediante un prompt desarrollado en Python con apoyo de una API de IA, se identificaron términos instrumentales iniciales. Esta etapa se basó en criterios establecidos por LexiMus, como la mención explícita, descripción organológica y contexto de uso musical.
- **Organización estructurada:**
  Los datos extraídos se almacenaron un archivo CSV con campos predefinidos como: Término; Lexema Principal; Nombres Alternativos/Sinónimos; Descripción Detallada;

Origen/Época/Construcción; Notas Adicionales; Clasificación (según la clasificación Hornbostel-Sachs, inferida); Clasificación descriptiva.

- **Revisión Musicológica en dos fases:**
  1. **Validación cruzada: Cada entrada fue cotejada exhaustivamente con las fuentes originales para** garantizar precisión y la correcta aplicación de los criterios de LexiMus.
  2. **Análisis especializado, que incluye:**
     - **Confirmación de identidad instrumental:** Evaluación experta para discernir inequívocamente si el lexema describe un instrumento musical, crucial para la coherencia de la ontología.
     - **Clasificación Hornbostel-Sachs (HS):** Asignación de categorías estándar para estructurar jerarquías y relaciones equivalentes en la ontología.
     - **Interpretación de clasificación y descriptores para H-S y la Ontología:** Identificación y validación de características cuantificables y cualitativas (n.º de cuerdas, orificios, materiales, dimensiones, mecanismo sonoro). Estos datos son vitales tanto para la granularidad de la ontología como para la aplicabilidad a sistemas de catalogación como el HS.
     - **Contextualización histórico-cultural:** Integración de información sobre origen, época y uso, enriqueciendo las entidades y relaciones en la ontología.

Para ilustrar en detalle cada paso del proceso, se mostrará el estudio del término “Abub”. Este caso resulta especialmente complejo al tratarse de un instrumento referido en fuentes bíblicas y de escasa presencia en los fondos museísticos. La descripción del instrumento presenta una notable variabilidad en las diferentes fuentes asignándole características diversas, a menudo contradictorias y de poca coherencia histórica. La ambigüedad terminológica lo convierte en un ejemplo paradigmático de los retos asociados a la taxonomía musical histórica y a los desafíos de la ontologización. Como se demostrará, la intervención experta resulta indispensable para asegurar la precisión y la coherencia en la construcción de una ontología musical fiable y útil.

Bibliografía

DE SCHRYVER, G.-M., & JOFFE, D. (2023), The end of lexicography, welcome to the machine: on how ChatGPT can already take over all of the dictionary maker’s tasks, recurso en línea: http://hdl.handle.net/1854/LU-01GWSGV0M8HYZVNXV8SSZ0VXBG.

FINESINGER, S. B. (1926), «Musical instruments in OT», *Hebrew Union College Annual*, 3, pp. 21–76. Recurso en línea: http://www.jstor.org/stable/23502521.

GHIRARDINI, C. (2008), «Filippo Bonanni's Gabinetto armonico and the Antiquarians' Writings on Musical Instruments», *Music in Art*, 33(1/2), pp. 168–234. Recurso en línea: http://www.jstor.org/stable/41818565.

KOLOZALI, S., BARTHET, M., FAZEKAS, G., & SANDLER, M. B. (2011), «Knowledge representation issues in musical instrument ontology design», en *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 465–470. Recurso en línea: https://ismir2011.ismir.net/papers/PS3-19.pdf.

POLIFONIA-PROJECT (s. f.), Music Instrument Ontology [Repositorio de GitHub], recurso en línea: https://github.com/polifonia-project/music-instrument-ontology.

Una ventana digital a la Historia Natural: Virtualización y divulgación científica accesible desde el Museu de la Universitat de València de Historia Natural.

Anna García-Forner
José Antonio Villena Gómez
Natalia Conejero-Ortega
Teresa Sáez Máñez
Sergio Montagud Alario

Palabras clave: Museología Digital; Digitalización 3D; Accesibilidad; Divulgación Científica; Patrimonio Natural y Científico
Keywords: Digital Museology; 3D Digitization; Accessibility; Scientific Dissemination; Natural and Scientific Heritage

El Museu de la Universitat de València d'Història Natural (MUVHN), como espacio dedicado a la ciencia, la cultura y la divulgación, se enfrentó al reto de garantizar el acceso universal a sus colecciones y actividades durante la pandemia de COVID-19.

Para dar respuesta a esta necesidad, nació el proyecto "*Una ventana digital a la Historia Natural",* con el objetivo de abrir el museo a la sociedad a través de la virtualización de sus espacios y la creación de recursos digitales. El proyecto, que se llevó a cabo con el apoyo de las convocatorias de ayudas de la Fundación Española para la Ciencia y la Tecnología (FECYT), pone un énfasis especial en la inclusión, asegurando que personas con diversidad funcional también puedan disfrutar del patrimonio natural y científico que custodia el MUVHN.

El proyecto se desarrolló a través de cuatro actividades interrelacionadas, pensadas para llevarse a cabo tanto en el museo como desde casa, gracias a los recursos digitales.

1. Qué nos cuentan los picos? Un acercamiento al mundo de las aves a través de sus cráneos y picos. Gracias a técnicas como la fotogrametría y el escaneado 3D, se creó una colección de modelos digitales cuya disponibilidad se realizó a través de la plataforma Sketchfab (bajo licencia CC). A partir de ellos, se imprimieron réplicas en 3D con materiales biodegradables (PLA), que se utilizaron, y se siguen utilizando, en talleres didácticos para explorar cómo la forma del pico está relacionada con la dieta y funcionalidad de cada especie.

2. La fonoteca de aves del MUVHN. Los sonidos también cuentan historias. Por eso, el museo incorporó una *fonoteca digital* de *cantos de aves* vinculada a las especies presentes en su exposición. Se adquirieron grabaciones de calidad

profesional, que ahora pueden escucharse directamente desde la web del MUVHN o durante la visita virtual en 360º, y también durante la visita presencial a través de códigos QR asociados a las aves correspondientes, dinamizando la exposición y permitiendo una experiencia, un mejor reconocimiento y aportando más “Naturalidad” a los ejemplares expuestos.

3. La colección 3D de fósiles del MUVHN. Se seleccionaron y digitalizaron diversos ejemplares de la colección paleontológica del museo mediante fotogrametría y escaneado 3D con luz estructurada. Los modelos resultantes se publicaron en la plataforma Sketchfab, con acceso libre para su visualización y descarga (CC). A partir de estos modelos, se crearon *puzles tridimensionales* impresos en PLA y diseñados para uso en talleres didácticos, donde los participantes reconstruyen fósiles con sus propias manos. Esta experiencia fomenta el aprendizaje táctil y visual y ayuda a identificar formas y estructuras de diferentes especies extintas.

4. Visita virtual 360º del MUVHN. El corazón del proyecto fue la creación de una visita virtual en 360° que permite recorrer el museo desde cualquier ordenador y en cualquier lugar del mundo. Se digitalizaron los principales espacios y ejemplares del MUVHN, integrando puntos interactivos a lo largo de todo el recorrido. Estos puntos actúan como puertas de acceso a todos los recursos digitales desarrollados en el proyecto: modelos 3D, audios, vídeos, fichas didácticas y mucho más. Además, la visita puede disfrutarse desde cualquier dispositivo, ya sea un ordenador, una tableta o un móvil, adaptada a diferentes sistemas operativos y sin necesidad de presencialidad en el museo.

Los resultados obtenidos demuestran el éxito del proyecto en sus cuatro líneas de acción. Las colecciones digitales 3D generadas con un alto valor científico y educativo, accesibles en abierto bajo licencia CC en Sketchfab, han permitido posicionar al MUVHN entre los museos internacionales de referencia. Las reproducciones tanto de cráneos de aves como de fósiles permiten no sólo la comprensión de la fauna (Actual y extinta), sino que también la hacen accesible a un mayor número de público por la manipulación, reconocimiento y estructuración de formas. La fonoteca de aves aporta un componente sensorial que da vida a la colección histórica y patrimonial de aves y permite integrar a dicha visita a público se encuentra en situación de discapacidad visual. La visita virtual 360°permite la salida del MUVHN de su espacio físico para llegar, a través de la tecnología, a cualquier rincón del mundo, salvando las barreras físicas y creando expectativas de visitas presenciales. Como museo universitario, el MUVHN garantiza la continuidad y el mantenimiento de estos recursos digitales, manteniendo su compromiso con la ciencia rigurosa y de calidad, pero a la vez al servicio de la sociedad, abierta, inclusiva y para todos.

Este proyecto se ha llevado a cabo con la colaboración de la FECYT , FCT-20-16150.

Cartografía bio-bibliográfica de un campo literario: el caso de la República Democrática Alemana

Jörn Kreutel
Berliner Hochschule für Technik
Axelle Lecroq
Berliner Hochschule für Technik – Universität Leipzig
Thomas Möbius
Humboldt-Universität zu Berlin

El proyecto FLFDDR[1] - acrónimo alemán de “Plataforma de investigación del campo literario de la República Democrática Alemana (RDA)” - tiene, a largo plazo, el objetivo de crear una base de datos bio-bibliográficos que represente las biografías y obras de todas las autoras y autores de obras literarias de la RDA, incluyendo todos los géneros y niveles literarios. Como actores de este campo literario [2] se identificaron un total de aproximadamente 3.400 personas que estaban presentes como autores de publicaciones independientes y/o como contribuidores a antologías literarias, incluidos escritores que publicaron en revistas “undergound” y otros medios no oficiales.

Como proyecto piloto se desarrollaron un modelo de datos biográficos y la infraestructura técnica que facilita la colección, el almacenamiento y el análisis de datos.[3] Se coleccionaron, como muestra inicial y ampliamente, los datos biográficos de 344 estudiantes que participaron en un programa de estudios en el “Instituto de literatura ‘Johannes R. Becher”’ (IfL) en Leipzig. El IfL fue fundado en el 1955 con el fin de institucionalizar la educación de autores conforme con el programa cultural del sistema socialista. Sin embargo, del entorno del IfL salieron autores que, como conjunto, pueden ser considerado representativos para todo el campo literario de la RDA. Como fuente de datos bibliográficos se importó una bibliografía completa, según su propia ambición, de la literatura de la RDA. Los datos de esta bibliografía fueron vinculados con la lista de los estudiantes del IfL y el resto de los escritores identificados como actores del campo literario de la RDA.

El póster que vamos a presentar en el congreso tiene la intención de exponer unos aspectos básicos del proyecto con respecto al modelo de datos biográficos y a las taxonomías y generalizaciones que se aplicaron para poder llevar a cabo determinados análisis cuantitativos. Mediante el análisis de correspondencias múltiples (ACM) vamos a demostrar, como contribución

[1] https://www.ddr-literatur.de/

original del póster, la construcción de un campo social sobre determinados atributos biográficos que se pueden considerar indicadores de diferentes tipos de “capital” en el sentido de Bourdieu, tomando en cuenta las circunstancias específicas de la RDA. Actualmente estamos investigando, colocando seleccionados atributos bibliográficos en este campo, cuáles son los efectos del IfL con respecto a la formación prevista de sus estudiantes como autores y si existen afinidades entre géneros y otros aspectos de la obra literaria, por una parte, y la posición de autores en el campo social, por otra. Estos análisis y la visualización de los resultados facilitan el acceso explorativo – como “lectura a distancia” – al conjunto de los datos coleccionados en nuestro proyecto y pueden dirigir la “lectura cercana”, basada en las biografías y las obras concretas de los autores.

La presentación del proyecto en el contexto del HDH 2025 tiene la intención de crear vínculos con las investigadoras e investigadores hispánicos en nuestro campo común de actividades. Vamos a demostrar cómo se pueden relacionar datos biográficos y bibliográficos para llegar a una perspectiva amplia de la producción cultural y de la formación de autores de obras culturales bajo un régimen de dirección y represión política, las similitudes del cual con las condiciones en España durante el franquismo fueron investigadas muy recientemente. El póster presentará el estado actual de nuestra investigación al que hayamos llegado en el momento de la entrega final.

Bibliografía

BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN (ed.) (2021), *Literatur in der SBZ/DDR. Bibliographische Annalen 1945–1990*, Berlin/Boston, De Gruyter Akademie Forschung.

BOURDIEU, P. (1992), *Les règles de l’art: genèse et structure du champ littéraire*, Paris, Éditions du Seuil.

KREUTEL, J., MÖBIUS, T., DAHLKE, B., & MARTUS, S. (2023), «Forschungsplattform Literarisches Feld DDR – Ein Werkstattbericht zur prosopographischen Erfassung von Schriftsteller:innen in der DDR», en H. ALBRECHT, M. FARRENKOPF, H. MAIER, & T. MEYER (eds.), *Historische Biographik und kritische Prosopographie als Instrumente in den Geschichtswissenschaften*, De Gruyter, pp. 141–165.

MÖBIUS, T. (2025), *Korpus Autor*innen in der DDR*, Working paper, recurso en línea: https://ddr-literatur.de/daily-research/korpus-autorinnen-in-der-ddr/.

VOLLMEYER, J. (2024), «Memoria suprimida: literatura bajo represión política», *1616: Anuario de Literatura Comparada*, 14, pp. 71–102.

ARStimuli: tecnologías digitales para el estudio de las obras de arte

Carme López Calderón
Universidad de Santiago de Compostela

Palabras clave: Base de datos relacional; Iconclass; Mapeado; Plástica barroca; Fuentes impresas
Keywords: Relational database; Iconclass; Mapping; Baroque art; Printed sources

El objetivo de este póster es presentar ARStimuli, un proyecto que todavía se encuentra en fase de desarrollo al haber sido propuesto recientemente para financiación al amparo del Programa de Consolidación Investigadora (Actuación CNS2024-154733 financiada por MICIU/AEI /10.13039/501100011033).

ARStimuli es un proyecto de Humanidades Digitales que empleará herramientas digitales y técnicas de computación tanto en su metodología de investigación como en su plan de difusión. En concreto, busca demostrar cómo las tecnologías digitales pueden desempeñar un papel clave para la Historia del arte no solo a nivel de divulgación, sino primeramente a nivel de investigación, bien porque permiten afianzar o invalidar ciertas tesis, o bien porque abren nuevas líneas de trabajo[1].

Para ello, y teniendo en cuenta los dos años previstos inicialmente para el proyecto, ARStimuli utilizará los relieves y las pinturas del barroco gallego como estudio de caso: un conjunto de, al menos, 500 obras que ya han sido objeto de numerosos estudios[2], pero que todavía (a) carecen de un estudio sistemático de las fuentes grabadas en que se basan, a diferencia de lo que sucede en otras geografías[3]; (b) no han sido analizadas a la luz de las nuevas tecnologías.

En concreto, cuatro son las herramientas digitales que ARStimuli propone aplicar a estos ejemplos para poder desarrollar análisis cualitativos a partir de datos cuantitativos: (1) una base de datos relacional, que permitirá registrar

[1] Hasta donde tenemos constancia, el único proyecto planteado desde la Historia del Arte que propone usar las nuevas tecnologías en un sentido similar al nuestro es ARCA, *Arte Colonial Americano* (Borja Gómez, 2023), si bien a nuestro juicio presenta el problema de la no estandarización de los contenidos iconográficos de acuerdo a un sistema más extendido, que por nuestra parte proponemos solventar a través del empleo del Iconclass.

[2] A las obras de referencia para el estudio del Barroco en Galicia, como pueden ser las dirigidas por Pulido (1999 y 2004), cabe sumar las numerosas tesis de doctorado y tesis de licenciatura defendidas en la Universidad de Santiago que en su mayoría permanecen inéditas.

[3] El ejemplo más significativo es PESSCA, *Project on the Engraved Sources of the Spanish Colonial Art* (Ojeda, 2005-2025). Igualmente, para el caso andaluz existe la monografía de Navarrete Prieto (1998).

información sobre las obras de arte, las fuentes grabadas, las localizaciones, los artífices y los promotores; (2) el online Iconclass Browser, que permitirá indexar el contenido de las imágenes de acuerdo con este sistema de clasificación, caracterizado por ofrecer un vocabulario estandarizado conformado por códigos alfanuméricos y correlatos textuales insertos en una estructura jerárquica; (3) el mapa, que permitirá localizar geográficamente las obras y será visible a través de (4) la web pública, que permitirá visualizar los metadatos enlazados y asociados a cada obra, así como hacer búsquedas cruzadas.

Estas cuatro herramientas serán un fin en sí mismo, porque permitirán difundir la plástica barroca gallega, pero ante todo funcionarán como un medio para la investigación, en relación con los objetivos específicos del proyecto, resumidos en su acrónimo (*Artworks' Study, Index, Mapping, and comprehensive Understanding by Linking Information*), y las preguntas de partida, que atañen a la distribución geográfica, contenidos, referentes, artífices y promotores de las obras.

Asimismo, conviene destacar que se trata de un proyecto fácilmente escalable a otro tipo de obras, cronologías y geografías, sentando por tanto las bases para futuras investigaciones que podrán ser tanto más ricas y novedosas, cuantas más obras de arte se incorporen a la base de datos.

Bibliografía

BORJA GÓMEZ, J. H. (2023), *ARCA (Arte Colonial Americano) (Principal)*, dataset, Zenodo, recurso en línea: https://doi.org/10.5281/zenodo.10864546.

NAVARRETE PRIETO, B. (1998), *Pintura andaluza del siglo XVII y sus fuentes grabadas*, Fundación Arte Hispánico.

OJEDA, A. (2005–2025), Project for the Engraved Sources of Spanish Colonial Art (PESSCA), sitio web, recurso en línea: https://colonialart.org.

PULIDO NOVOA, A. (dir.) (1999), *Artistas Galegos. Pintores. Ata o Romanticismo*, Nova Galicia Edicións.

PULIDO NOVOA, A. (dir.) (2004), *Artistas Galegos. Escultores. Séculos XVI e XVII*, Nova Galicia Edicións.

La biblioteca universitaria como apoyo a la docencia en innovación educativa y humanidades digitales

María del Don Martínez Aliaga[1]
Elena Viqueira Lagares[2]
Elisa Pastor Planells[3]
Eva Montilla Navas[4]
Universitat de València –
Biblioteca d'Humanitats «Joan Reglà»

Palabras clave: Epistolarios; escritoras españolas; exilio; redes sociales; visualización de redes sociales
Keywords: Letters; Spanish women writers; exile; social media; social media visualization

Descripción del problema

Las bibliotecas universitarias deben seguir desempeñando un papel clave en la integración de las iniciativas de Innovación Educativa y las Humanidades Digitales, consolidando su rol tradicional como centros de acceso al conocimiento, conservación y apoyo a la docencia y la investigación.

En el último congreso IMPACTA REBIUN, celebrado en Girona los días 29 y 30 de mayo de 2025, sobre microcredenciales, varios ponentes señalaron que, en los últimos años, se ha trabajado intensamente en el ámbito del apoyo a la investigación, con resultados muy positivos, pero se ha descuidado ligeramente el apoyo a la docencia, un ámbito en el que las bibliotecas siempre hemos estado muy implicadas.

Desde el Servei de Biblioteques i Documentació, y especialmente desde la Biblioteca d'Humanitats "Joan Reglà", llevamos tiempo colaborando con el profesorado de diferentes departamentos y facultades, para seguir trabajando y apoyando las tareas docentes, también en el ámbito de la innovación educativa y las humanidades digitales, aunque no de forma estructurada con acciones diseñadas a corto y largo plazo.

Actualmente, los resultados son satisfactorios, hemos participado en numerosos proyectos, con buena acogida por parte del profesorado y con una participación muy entusiasta del alumnado, que ve su trabajo reflejado más allá

[1] Ayudante de biblioteca, sección de Información y Referencia.
[2] Ayudante de biblioteca, sección de Apoyo a la Investigación.
[3] Ayudante de biblioteca, sección de Información y Referencia.
[4] Coordinadora de la Sección de Información y Referencia.

del aula.

Los resultados de estos proyectos están disponibles en el canal de YouTube de bibliotequesuv, accesibles a la sociedad en forma de exposiciones virtuales en la página web de la Biblioteca d'Humanitats: https://ir.uv.es/humanbib/exposicions

Actualmente, nos enfrentamos a varios problemas:

1. Dificultad para delimitar funciones, alcance y recursos en HD.
2. Difusión insuficiente de estas colaboraciones entre todo el profesorado. La mayoría de los docentes que contactan con nosotras, han oído hablar de estas colaboraciones a un profesor o profesora que ha trabajado con la sección de información. No por haber visitado la página web de la biblioteca.
3. Falta de un catálogo estructurado de actividades y servicios que, desde las bibliotecas, podemos ofrecer en apoyo a la docencia, especialmente, en el ámbito de la innovación educativa y humanidades digitales.
4. Desconexión con otros servicios y proyectos universitarios.

Metodología

El Servei de Biblioteques i Documentació ha integrado, desde hace unos años, una apuesta consciente y responsable por la mejora continua de instalaciones, procesos y servicios. En 2025, el servicio ha conseguido el prestigioso sello de calidad EFQM-500. Este logro, nos anima a continuar mejorando, progresando e innovando.

Nuestro Plan Director 2024-2028 dedica su línea estratégica 2 al estudio y la mejora del Apoyo a la Docencia y el Aprendizaje desde las bibliotecas.

Una de las iniciativas clave dentro de esta línea es el impulso de las humanidades digitales, integradas en el contexto institucional. Desde el SBD, y especialmente desde la Biblioteca d'Humanitats, llevamos unos años colaborando con el profesorado, en la visualización de trabajos multimedia realizados por el alumnado. Muchos de estos trabajos, forman parte de proyectos de innovación educativa y visibilizan, autores y autoras a través de listas de reproducción en el canal de bibliotequesuv, redes sociales institucionales y exposiciones virtuales, donde los proyectos quedan organizados y accesibles.

El objetivo de este grupo de trabajo se centra en definir con claridad nuestro papel en estas colaboraciones, para ofrecer aquellos servicios que el profesorado necesite.

Nuestros primeros pasos, con el objetivo de ofrecer un catálogo detallado de los servicios y abrirnos a nuevos proyectos en el futuro, son:

1. Revisión de los proyectos llevados a cabo.
2. Identificación de buenas prácticas y modelos de referencia.
3. Formación de los miembros del grupo.

Resultados

Propuesta de definición operativa de las humanidades digitales para las bibliotecas de la UV:

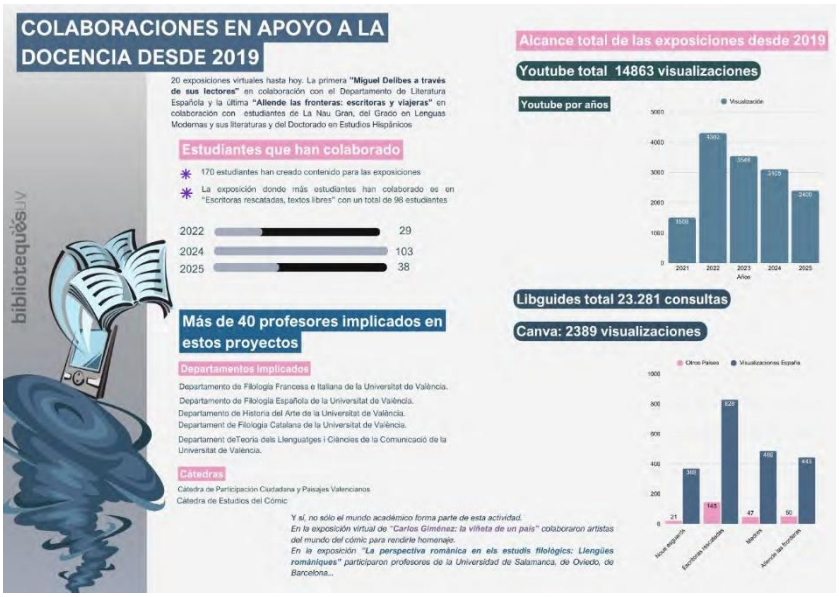
- Acceso a recursos digitales. Las suscripciones principales están centralizada en la sección Gestión de Recursos de Información del SBD.

  https://uv-es.libguides.com/az/databases?s=130163

  - Bases de datos en el área de humanidades:
  - Repositorio Roderic: https://roderic.uv.es/home
  - Incorporación de la IA en Trobes y bases de datos suscritas
- Preservación y Digitalización. Trabajo realizado en Biblioteca Histórica
- Capacitación y Alfabetización Digital:
  - Formación a la carta (Sección de Información y Referencia en diferentes niveles de enseñanza)
  - Formación y asesoramiento en Investigación (Sección de Apoyo a la Investigación y Portal de Producción Científica): https://produccioncientifica.uv.es/
- Colaboración en Proyectos de Investigación:
  - Participación de la Biblioteca d'Humanitats en Memory Novel Lab: https://mnlab.uv.es/
- Colaboración con proyectos de Innovación Docente y Educativa en la Biblioteca d'Humanitats. Recogidos, hasta el momento, en la web de Exposiciones y Eventos de la biblioteca: https://ir.uv.es/humanbib/exposicions

Muestra de los proyectos llevados a cabo

- Estadísticas de los trabajos de colaboración



- Colaboraciones con el profesorado del Departamento de Filología Española
  - Proyectos de investigación
    - Memory Novels LAB
      Dirección base de datos: José Martínez Rubio; Luz C. Souto Larios.
      El proyecto Memory Novels LAB tiene como objetivo la creación del primer Laboratorio de literatura de la memoria sobre la Guerra Civil española y el franquismo
    - Escritoras rescatadas, textos libres (marzo 2024)
      Su objetivo es visibilizar y promover la lectura de autoras que escribieron durante el franquismo o los primeros años de democracia.
      Este trabajo parte de la investigación “Rojas sombras en la periferia del canon. La novela femenina frente al franquismo” (Souto Larios, Luz C. 2024). Disponible en: h ttps://hdl.handle.net/10550/104356
  - Proyectos de innovación educativa
    - DTD_iLAB: Aprendizaje colaborativo por proyectos interdisciplinares (tercera fase). Proyecto de innovación Educativa Consolidado de la UV
    - Subjetividades en crisis en la literatura español contemporánea (1914-1975). (CIGE/2023/74) (2024-2025)
    - Memory Novels LAB. Laboratorio de literatura de la memoria.
    - *Allende las fronteras, escritoras viajeras* (junio 2025)
      Resultado del trabajo colectivo de estudiantes de La Nau Gran, del Grado en Lenguas Modernas y sus Literaturas y del Doctorado en Estudios Hispánicos, bajo la supervisión de la profesora Rita Rodríguez Varela.
    - *Madres, el portento quebrantando las sombras. La maternidad en la producción cultural hispánica* (marzo 2025)
      Surge a partir del debate y la reflexión mantenida en la asignatura “Literatura, teatro y productos culturales: proyección en la sociedad actual”, en el marco del Máster en Estudios Hispánicos Avanzados: Aplicaciones e Investigación impartido por la profesora Luz C. Souto
    - *Voces y letras contra la violencia: imaginarios literarios y creación en videoart* (marzo 2022)
      Proyecto de innovación docente, coordinado por la profesora Josefa Badía, que trata de fomentar el pensamiento crítico, y la contribución, desde los estudios literarios, a la reflexión sobre la violencia como forma que atenta contra la dignidad del ser humano.

Parte del trabajo en videoarte de estudiantes de la Facultad de Filología, Traducción y Comunicación, en el marco de la asignatura Lecturas Españolas Contemporáneas, impartida por la profesora Luz C. Souto.

- Difusión de recursos en Humanidades Digitales
  - Lista institucional de Youtube bibliotequesuv
  - Trabajos de fin de curso, realizados durante 2021/2022 y 2023/2024 por los estudiantes del Máster Universitario de Estudios Hispánicos Avanzados: aplicaciones e investigación
- Universidad y sociedad
  - *Filosofía de la incomunicación* (octubre 2024)
    Fernando Reati y Paula Simón nos ofrecen en su libro *Filosofía de la incomunicación. Las cartas clandestinas de la Unidad Penitenciaria 1 (Córdoba, 1976-1979)*. Trabajo realizado con la coordinación de la profesora Luz C. Souto.
  - *Miguel Delibes a través de sus lectores* (diciembre 2020)
    Con motivo del aniversario de su nacimiento, la biblioteca quiso rendirle su homenaje, con la colaboración, entre otras aportaciones, de profesores y profesoras del departamento de Literatura Española.
  - *Benito Pérez Galdós y la deconstrucción de una sociedad* (diciembre 2019)
    Con motivo del centenario de su fallecimiento, la Biblioteca d'Humanitats quiso rendirle un homenaje con una exposición física de sus obras y una bibliografía confeccionada con ayuda del departamento de Filología Española
  - *Mujeres pioneras* (mayo 2018)
    Una bibliografía y una exposición física de libros sobre el tema, recopilados por la Biblioteca d'Humanitats

Bibliografía

FERRER VALLS, T. (2025), «Introduction: digital humanities», *Mètode Science Studies Journal*, 15.

GONZÁLEZ-GUTIÉRREZ, S., & MERCHÁN-SÁNCHEZ-JARA, J. (2022), «Humanidades digitales y ecosistema educativo: hacia una nueva estructura epistémica desde las didácticas digitales», *Anuario ThinkEPI*, 16, pp. 16–35.

LÁZARO, M., & DE-JUANAS, Á. (2024), «Humanidades digitales y nuevas metodologías de enseñanza de las humanidades», *Revista Iberoamericana de Educación*, 94(1), pp. 9–14.

MARTÍNEZ CANTÓN, C. I., et al. (eds.) (2024), *Las humanidades digitales en la enseñanza de las literaturas hispánicas: aplicaciones prácticas*, Berlin/Bruxelles/Chennai/Lausanne/New York/Oxford, Peter Lang.

URQUIJO, H., & ZUBIETA, M. (eds.) (2014), *Renovación en el Siglo de Oro: repertorio e instrumentos de investigación*, Compañía Nacional de Teatro Clásico.

Preservación en formato futuro de la memoria en riesgo: difusión y conservación digital de la tradición abaniquera de Aldaia

Marta Segovia Cobo
Universitat de València

Palabras clave: Abanicos; modelos 3D; puesta en valor; conservación
Keywords: Fans; 3D models; enhancement; conservation

El abanico constituyó un elemento significativo en los siglos XIX y XX del folclore valenciano. Fue en estas dos centurias cuando su producción aumentó y su elaboración tuvo que tecnificarse para responder a las exigencias de la demanda. Debido a este crecimiento, los fabricantes ampliaron su área de manufacturación a las poblaciones vecinas (Casaní Real, 2017).

Por ende, Aldaya se convirtió en uno de los centros especializados en la creación de varillajes asociados a Valencia. Con el transcurso del tiempo, los operarios fueron adquiriendo las habilidades necesarias para generar sus propios focos de confección y extender su actividad a la realización completa de los abanos. Fue de este modo como se inició la distinguida industria abaniquera aldaiense, que actualmente es la única operativa del sector (Casaní Real, 2017).

Las empresas en activo sobreviven gracias a la valorización del abanico como objeto artístico a través de la moda (Casaní Real, 2017). Sin embargo, esto contrasta con el escaso interés que genera el estudio y la investigación de su historia a nivel académico. El restringido conocimiento de su existencia, junto a las limitaciones económicas y geográficas que sufren las compañías artesanas locales para impulsar proyectos de divulgación, deriva en la lejanía de las piezas con el ámbito cultural profesional y el público general que lo consume. Asimismo, la ausencia de una conciencia sobre su valor provoca el desamparo de la preservación de los ventalles, pudiéndose generar un riesgo de pérdida.

Con el fin de hallar una solución a esta problemática se origina el presente proyecto. Frente a la dicha desinformación, se busca implementar el uso de las nuevas tecnologías con dos objetivos principales. En primer lugar, conseguir que la tradición abaniquera alcance una mayor audiencia potencial y, por lo tanto, se difunda su legado a un mayor número de personas. El éxito de trabajos llevados a cabo en el recorrido de la digitalización avala el claro balance positivo de la implicación de los elementos virtuales en trabajos de transmisión del conocimiento (Ruiz Torres, 2017). Por otra parte, se quiere erradicar la probabilidad de desaparición de la huella de los abanos con la copia digital como

herramienta de conservación. Dada la coyuntura de amenaza por el devenir del tiempo, se ha optado por la tecnología como el aliado perfecto para frenar su caída en el olvido y asegurar la preservación de su memoria (Merchán y Pérez, 2020).

El desarrollo del estudio se está ejecutando sobre una selección de piezas de la empresa familiar Burriel Abanicos, ubicada en Aldaya. Concretamente, se están empleando algunos de los artículos que hoy en día fabrican y, principalmente, la colección histórica que los dueños han ido acopiando.

El flujo de trabajo ha comenzado con la obtención de modelos 3D de cada uno de los ventalles. A tal efecto, teniendo en cuenta las necesidades de los negocios y la pretensión de darles una herramienta útil perdurable, se ha escogido como método de captura la fotogrametría, dada su gran asequibilidad (Didier Niquet y Mas-Barberà, 2018). Tras una ardua organización, la toma de datos se ha registrado en múltiples anillos para recoger el volumen total de los abanicos.

Una vez finalizado este primer paso, se está efectuando el procesado de la información recopilada en un software especializado, a través del cual se obtendrá los modelos de altos polígonos. Mientras que estos se mantendrán para la versión profesional, también se generarán otros *low poly* a través de un procedimiento de optimización de la geometría y las texturas para el público general. Su menor peso y su más sencilla distribución permiten democratizar el acceso a los conocimientos a compartir.

Por último, se creará una base de datos donde almacenar los objetos tridimensionales. A través de la red informática, se generará un espacio de fácil entrada para salvaguardar el legado aldaiense, realizándose un catálogo web donde acudir para consultar una fuente bibliográfica de primera calidad y que haga perdurar esta tradición en el tiempo. En definitiva, se obtiene un primer punto de partida que pone a los fabricantes de Aldaia y sus piezas artísticas en el mapa, generando una experiencia piloto que podría implementarse en todos los focos de la localidad para solucionar efectiva y plausivamente la situación de este ámbito poco explorado.

Bibliografía

CASANÍ REAL, M. C. (2017), *La fabricación de abanicos en Valencia en los siglos XIX y XX* [Tesis doctoral, Universitat de València], Repositorio de tesis de la Universitat de València. Recurso en línea: https://roderic.uv.es/items/7d3a9d9d-a089-4d2b-9ed4-a19f2d77acb7.

DIDIER NIQUET, N., & MAS-BARBERÀ, X. (2018), «El registro 3D como medio para el análisis y difusión del patrimonio escultórico. El caso de la escultura en cera del Écorché», *Ge-conservación*, 13, pp. 5–16. Recurso en línea: https://doi.org/10.37558/gec.v13i0.551.

MERCHÁN, M. J., & PÉREZ, E. (2020), «Conservation to overcome oblivion. New methods for the survival of lost heritage memory», *Ge-conservación*,

17(1), pp. 160–170. Recurso en línea: https://doi.org/10.37558/gec.v17i1.753.

RUIZ TORRES, D. (2017), «El uso de las tecnologías en la conservación, análisis y difusión del patrimonio cultural», en ACCIÓN CULTURAL ESPAÑOLA (ed.), *Anuario AC/E 2017 de Cultura Digital*, pp. 130–211. Recurso en línea: https://www.accioncultural.es/media/DefaultFiles/flipbook/Anuario2017/anuario2017.html#p=5.