

Encuentros y transformaciones

Las humanidades
digitales como propuesta
transdisciplinar

María Luisa Díez Platas

María Teresa Santa María

Nadia McGowan (coords.)



digitalia
humanística

Encuentros y transformaciones

**Las humanidades
digitales como propuesta
transdisciplinar**

María Luisa Díez Platas
María Teresa Santa María
Nadia McGowan (coords.)



UNIVERSIDADE DA CORUÑA

**ENCUENTROS Y TRANSFORMACIONES
LAS HUMANIDADES DIGITALES COMO
PROPUESTA TRANSDISCIPLINAR**

María Luisa Díez Platas, María Teresa Santa María,
Nadia McGowan (coords.)
A Coruña, 2024

Universidade da Coruña, Servizo de Publicacións

COLECCIÓN

Digitalia Humanística, n.º. 3

DIRECCIÓN DA COLECCIÓN

Carlota Fernández Travieso (Universidade da Coruña)

COMITÉ CIENTÍFICO

Mar Campos Souto (Universidade de Santiago de Compostela)

Fátima Díez Platas (Universidade de Santiago de Compostela)

Ana García Serrano (Universidad Nacional de Educación a Distancia)

Sagrario López Poza (Universidade da Coruña)

Stefano Neri (Università di Verona - Italia)

Nuria Rodríguez Ortega (Universidad de Málaga)

Amelia Sanz Cabrerizo (Universidad Complutense de Madrid)

Juan Luis Suárez (Western University - Canada)

Esta obra someteuse a revisión por pares simple cega,
con persoas avaliadoras externas á UDC

240 pp.

Índice: p. 5

Depósito Legal: C 1965-2024

ISBN (versión dixital): 978-84-9749-915-6

DOI (versión dixital): <https://doi.org/10.17979/spudc.9788497499156>

THEMA: UXA; UXJ; UN; UL

CDU: 004.5; 004.8 ; 004.9



Esta obra publícase baixo unha licenza Creative Commons

Atribución-NonComercial-CompartirIguual 4.0 Internacional (CC BY-NC-SA 4.0)

EDICIÓN

Servizo de Publicacións. Universidade da Coruña

© da edición: Universidade da Coruña | © dos textos: as autoras e os autores

ASESORAMENTO TÉCNICO DE DESEÑO

Eva Díaz Torres

FINANCIACIÓN

A colección *Digitalia Humanística* é impulsada polo grupo HISPANIA da Universidade da Coruña. Este volume foi parcialmente financiado grazas ás Axudas para a consolidación e estruturación de unidades de investigación competitivas (GPC) da Xunta de Galicia (referencia ED431B 2025/24).

Encuentros y transformaciones

Las humanidades
digitales como propuesta
transdisciplinar

María Luisa Díez Platas
María Teresa Santa María
Nadia McGowan (coords.)

ÍNDICE

[07] **PRÓLOGO**

María Luisa Díez, María Teresa Santa María,
Nadia McGowan

- [13] El sistema de bases de datos federadas ASODAT:
nuevos caminos para la investigación en teatro
clásico español

Teresa Ferrer Valls

- [39] El uso de las bases de datos y los SIG en arqueología.
Balance a través del análisis de encuestas, revisión
historiográfica y bibliometría

Leticia Tobalina Pulido, César González-Pérez

- [63] Diseño del acceso a un catálogo sobre la Antigua Grecia

Ana García Serrano, Fernando Chuquimarca Bolaños,
Francisco Paños Merino, César Fernández Sánchez

- [91] Sistema de etiquetado automático de catálogos
de exposiciones para la extracción de características
y entidades: primera aproximación

María Luisa Díez Platas, María Ortiz Tello,
Manuel García Lirio, Nuria Rodríguez Ortega

- [107] Comparativa de los sentimientos y las emociones
básicas más frecuentes en las novelas de Jane Austen
a través del análisis de sentimientos

Eduarne Sáenz

- [133] Análisis cuantitativo preliminar de la traducción inglesa
de la novela El diario de Satán, de Leonid Adreev

Patricia González Almarcha

- [163] Humanidades en el espacio: análisis y visualización

Nadège Rollet, José Antonio Berenguer-Sánchez,
Patricia Sánchez-García, Jan Thiele

- [191] Cuando el reto no es *descubrir* yacimientos:
integración semántica y caracterización histórica
de paisajes serranos con LiDAR

Sonia Medina Gordo

- [213] Visualización y reconocimiento de la memoria histórica
de las Capullanas/Cacicas indígenas desde el campo de
las Humanidades Digitales

Alicia Alvarado Escudero

PRÓLOGO

María Luisa Díez Platas

María Teresa Santa María

Nadia McGowan

Dyson, un físico y matemático inglés especializado en electrodinámica cuántica, escribió a finales de los noventa un libro en el que elucubraba sobre cómo podría ser el mundo de treinta años después. Entre sus muchas ideas, habla de dos tipos de revoluciones científicas diferentes, “las que son impulsadas por nuevas herramientas y las que lo son por nuevos conceptos” (Dyson, 1998: 49). Las diferencias que hay entre una y otra es la que se puede encontrar entre las aportaciones de Copérnico y Galileo. Ambos esenciales para la ciencia moderna pero completamente diferentes. Copérnico inició una revolución en cuanto a cómo se concebía el universo. Galileo trajo consigo la astronomía moderna a través de la invención del telescopio, una herramienta.

En la actualidad nos vemos inmersos en cambios técnicos constantes que abarcan desde la aparición de internet, los teléfonos móviles o la inteligencia artificial generativa. Estas tecnologías son aportaciones más similares a las de Galileo que a la

María Luisa Díez Platas, UNIR

ORCID: 0000-0002-9416-3147

María Teresa Santa María, UNIR

ORCID: 0000-0003-0281-7754

Nadia McGowan, UNIR

ORCID: 0000-0003-3584-4767

de Copérnico. Igual que un ordenador “no es un concepto, sino un instrumento que nos ayuda a pensar con claridad” (Dyson, 1998: 51), estos instrumentos nos ayudan a realizar tareas diversas, pero no aportan en sí mismas un cambio conceptual. Sin embargo, han surgido muchas voces buscando reflexiones en torno a todos estos cambios, que no se han realizado con la suficiente agilidad para dar respuesta a los nuevos retos que plantean estas herramientas.

En estos momentos de cambios constantes, las Humanidades Digitales son especialmente relevantes por su aportación en la combinación entre disciplinas.

Esta idea ya fue desarrollada por Antonio Rodríguez de las Heras, considerado uno de los padres de las Humanidades Digitales en España. En 2013 ya presentó una visión de las HD no como multidisciplinares ni interdisciplinares, sino transdisciplinares. Para él, las HD son el resultado del trabajo de cada persona desde su propia disciplina traspasando sus fronteras para llegar a un punto de encuentro común (Trujillo Sáez, 2013).

En el 2014, ahondando en esta misma idea, la revista Janus editó el Anexo 2, titulado “Humanidades digitales: una aproximación transdisciplinar” (Baraibar, Alvaro (editor), 2014). En este, se afirma que “las fronteras entre las distintas disciplinas y áreas del saber van quedando progresivamente más difusas e incluso más diluidas” y como “la confluencia de saberes o transdisciplinariedad brilla con especial fuerza y reclama una atención mayor si cabe en el campo de las Humanidades Digitales” (Baraibar Alvaro, 2014). Estas ideas son premonitorias de la evolución que sufrirían las Humanidades Digitales hasta llegar a su situación actual.

Es precisamente esta idea de la transdisciplinariedad como encuentro entre disciplinas y la generación de conocimiento nuevo, mayor que la suma de sus partes individuales, la que inspiró la temática del VI Congreso de las Humanidades Digitales 2023 “Encuentros y transformaciones: las Humanidades Digitales como propuesta transdisciplinar”.

A partir de los debates suscitados, se puede concluir que la naturaleza transdisciplinaria de la Humanidades Digitales es ya

una realidad, en la que, de acuerdo con este enfoque, tienen cabida proyectos e investigaciones en Humanidades Digitales que integran la Inteligencia Artificial, la representación 3D, el patrimonio, las bibliotecas digitales, el análisis de discurso y las bases de datos y sus aplicaciones.

En línea con estas tendencias en el estudio de las Humanidades Digitales, este volumen busca recopilar estudios cuya metodología u objetivos sean el resultado de la combinación de técnicas e investigaciones provenientes de distintas disciplinas, vinculando ciencias y letras, logrando así mejorar ambas.

Los capítulos de este libro incluyen tanto los resultados y conocimientos derivados de proyectos consolidados, con sus logros y perspectivas, así como nuevas propuestas que apuntan a la consecución de futuros objetivos, o el análisis del uso de las tecnologías en combinación con las humanidades. Todos ellos reflejan el carácter transdisciplinar de las Humanidades Digitales.

En primer lugar, en el libro se incluyen aportaciones relacionadas con bases de datos y tecnologías para las Humanidades Digitales. Los dos primeros capítulos abordan el uso de tecnologías en bases de datos y en sistemas de información geográfica, analizando su impacto en ámbitos de conocimiento tales como el teatro y la arqueología (“El sistema de bases de datos federadas ASODAT: nuevos caminos para la investigación en teatro clásico español”, “El uso de las bases de datos y los SIG en arqueología. Balance a través del análisis de encuestas, revisión historiográfica y bibliometría”). El tercer capítulo explora el empleo de herramientas digitales combinadas, tales como gestores de contenidos, herramientas de indexación de colecciones, plataformas digitales de acceso abierto y herramientas de análisis, que permiten optimizar las visualizaciones y el acceso a información histórica específica (“Diseño del acceso a un catálogo sobre la Antigua Grecia”, un análisis y uso de herramientas exhaustivo).

A continuación, se presentan dos capítulos que aplican técnicas de Inteligencia Artificial, centrándose en modelos de aprendizaje profundo para el procesamiento de lenguaje natural y

análisis de sentimientos. Estos son “Sistema de etiquetado automático de catálogos de exposiciones para la extracción de características y el reconocimiento de entidades nombradas: primeras etapas” (un trabajo en proceso) y “Comparativa de los sentimientos y las emociones básicas más frecuentes en las novelas de Jane Austen a través del análisis de sentimientos”. También se incluye un capítulo que utiliza técnicas de lingüística computacional para realizar un análisis cuantitativo de obras literarias (“Análisis cuantitativo preliminar de la traducción inglesa de la novela El diario de Satán, de Leonid Adreev”).

A continuación, se presentan dos capítulos que proponen, por una parte, el enriquecimiento de datos lingüísticos mediante análisis geográfico y su visualización, y, por otra, el enriquecimiento semántico de datos cuantitativos con la integración de conocimiento cualitativo. Estos artículos son “Humanidades en el espacio: análisis y visualización” y “Cuando el reto no es descubrir yacimientos: integración semántica y caracterización histórica de paisajes serranos con LiDAR”.

Por último, se incluye un capítulo centrado en la perspectiva histórica con el apoyo de las Humanidades Digitales, “Visualización y reconocimiento de la memoria histórica de las Capullanas/Cacicas indígenas desde el campo de las Humanidades Digitales”).

Este volumen es una muestra de los diversos ámbitos y tecnologías que convergen en la investigación en Humanidades Digitales. A través de los diferentes capítulos presentados, se evidencia no solo el impacto de la combinación de distintas disciplinas, sino el valor de la transdisciplinariedad como motor para la generación de conocimiento nuevo. Esta capacidad para trascender fronteras disciplinares pone de relieve el papel fundamental de las Humanidades Digitales en la construcción de respuestas innovadoras y reflexivas ante los desafíos tecnológicos y culturales de nuestro tiempo.

REFERENCIAS BIBLIOGRÁFICAS

- BARAIBAR, Álvaro (editor), *Humanidades Digitales: una aproximación transdisciplinar*, Janus [en línea], Anexo 2, 2014, < <https://www.janusdigital.es/anexo.htm?id=6>>, [consulta:10/10/2024].
- BARAIBAR, Álvaro, «Las Humanidades Digitales desde sus centros y periferias», en *Humanidades Digitales: una aproximación transdisciplinar*, Álvaro Baraibar (editor), Janus [en línea], Anexo 2, 2014, 7-15, <<https://www.janusdigital.es/anexos/contribucion.htm?id=42>>, [consulta: 10/10/2024].
- DYSON, Freeman, *Mundos futuros [Imagined Worlds]* (J. Ros Trans.), Barcelona: Grijalbo Mondadori, 1998.
- TRUJILLO SÁEZ, Fernando, *Humanidades digitales, una respuesta transdisciplinar: Antonio Rodríguez de las Heras en #cshdugr*. GrinUGR, 2013, <<https://grinugr.org/jornadas-cshd/humanidades-digitales-una-respuesta-transdisciplinar-antonio-rodriguez-de-las-heras-en-cshdugr/>>, [consulta:6/6/2024].

EL SISTEMA DE BASES DE DATOS FEDERADAS ASODAT: NUEVOS CAMINOS PARA LA INVESTIGACIÓN EN TEATRO CLÁSICO ESPAÑOL¹

Teresa Ferrer Valls

Resumen

Este artículo aborda la evolución del sistema ASODAT, una federación de bases de datos diseñadas para investigar el teatro clásico español desde el siglo XVI al XVII y ofrece ejemplos de las potencialidades de esta herramienta para la investigación. El sistema organiza información sobre la práctica escénica a partir de noticias sobre actores, compañías, representaciones y censuras. También a partir de los propios textos teatrales, tanto impresos como manuscritos, y sus características materiales o rasgos estilométricos, o sobre la función de la música. El sistema incluye la digitalización y marcación de muchos de estos textos y traducciones a otros idiomas. Se ponen de relieve en el capítulo las ventajas de la digitalización y de la marcación y relación de datos, que han permitido avances significativos. La herramienta ha contribuido a establecer cronologías precisas de representaciones y a profundizar en aspectos de cara a la autoría de obras, su adscripción a determinados géneros o el desarrollo de las compañías

- 1 Este artículo responde a la sugerencia de la Sociedad Internacional de Humanidades Digitales Hispánicas, y especialmente a las organizadoras de su VI Congreso, María Luisa Díez Platas, Elena Martínez Carro y Teresa Santa María (UNIR), para que llevase a cabo un balance de mi experiencia como humanista en relación con las nuevas tecnologías y las herramientas que hemos ido creando en estos años. Agradezco la invitación para su publicación.

Teresa Ferrer Valls, Universitat de València
m.teresa.ferrer@uv.es
ORCID: 0000-0002-5082-1179

profesionales. Sin embargo, también se ponen de relieve los desafíos asociados con el manejo de grandes volúmenes de información, como son la necesidad de mejorar los filtros de búsqueda y optimizar la presentación de resultados en un trabajo que está en progreso. El artículo evidencia cómo la tecnología digital transforma la investigación humanística, consolidando herramientas como ASODAT como referentes para la preservación y el análisis del patrimonio teatral. Finalmente, se subraya la importancia de la colaboración interdisciplinaria y el diseño de interfaces de acceso abierto para maximizar el impacto de estas iniciativas.

Palabras clave: bases de datos; teatro español; siglo de Oro.

Abstract

This article focuses on the evolution of the ASODAT system, a federation of databases designed to do research on sixteenth- and seventeenth-century Spanish theatre, and offers examples of this tool's potential. The system structures information on theatrical practices based on data about actors, professional companies, performances, and censorship. It is also based on the theatrical texts themselves, both prints and manuscripts, and their material characteristics or stylometric features, or on the role of music. The system includes the digitization and marking of many of these texts, as well as translations into other languages. This chapter highlights the advantages of the digitization, the tagging and linking of data, which have led to significant advances in the field. This tool has helped to establish precise chronologies of performances and to clarify aspects of the authorship of plays, their ascription to certain theatrical genres, or the development of professional companies. However, it also highlights the challenges associated with handling large volumes of information, such as the need to improve search filters and optimize the presentation of research results that are a work in progress. This paper demonstrates how digital technology is transforming humanistic research, as it has consolidated tools such as ASODAT as references for the preservation and analysis of theatrical heritage. Finally, this chapter underlines the importance of interdisciplinary collaboration and of designing open-access interfaces to maximize the impact of these initiatives.

Keywords: Databases; Spanish theater; Spanish Golden Age.

Fue a mediados de la década de 1990 cuando me interesé por la aplicación de nuevas tecnologías a la creación de instrumentos que permitiesen la organización de la información sobre teatro clásico español y sus prácticas escénicas, una información que, en el caso de España, es mucha, a pesar de toda la que se ha perdido, y que en aquel momento se encontraba dispersa en multitud de fuentes. Eso llevó a la publicación en 2008, después de casi quince años de trabajo, de lo que modestamente creo que es hoy una herramienta de referencia para los investigadores en teatro clásico, la base de datos *Diccionario biográfico de actores del teatro clásico español. DICAT* (2008), producto de un enorme esfuerzo colectivo². Transcurridos quince años desde su publicación, el soporte (un DVD) y el software sobre el cual se presentó aquella base de datos habían quedado obsoletos y, por otro lado, la aparición de información a lo largo de este tiempo nos hizo plantearnos la conveniencia de rediseñar la herramienta para ofrecerla en abierto y crear los fundamentos que permitiesen su ampliación y la preservación y puesta en valor de un patrimonio inmaterial que, en este caso, está constituido por numerosas noticias que dan cuenta de la actividad de compañías de actores en diferentes espacios y lugares.

El *Diccionario*, tal y como se acaba de publicar en acceso abierto en 2023 (<https://dicat.uv.es>), permite búsquedas que reproducen las que los usuarios de esta base de datos están habituados a hacer: por nombres y/o apellidos, por sexo, por fechas, por funciones (actor, actriz, apuntador, músico, director o directora de compañía) o por texto libre. El *Diccionario* ha mostrado ser muy útil en relación con la identificación de nombres de actores o de directores que figuran en repartos incluidos en manuscritos, lo que ha permitido en ocasiones no solo fijar cronológicamente la fecha y

2 Todos los proyectos a los que haré referencia hubiesen sido impensables sin la participación de muchos investigadores e investigadoras. No puedo mencionarlos a todos, pero vaya aquí mi agradecimiento. En estos momentos el grupo DICAT, que tomó su nombre de aquel primer proyecto, cuenta con la financiación para el actual proyecto del MICIU/AEI/10.13039/501100011033 y de FEDER, UE. Referencia: PID2022-136431NB-C62.

circunstancias de una puesta en escena, sino también apoyar propuestas de autoría de una obra o establecer la fecha de composición de un texto dramático, o apuntalar la fecha de una copia manuscrita. Pero el volumen de datos es bastante elevado y en ocasiones el resultado de las búsquedas (cuando no se limitan a la búsqueda de un nombre) requiere un esfuerzo de lectura que filtre mejor los resultados obtenidos. Por ejemplo, el *Diccionario* puede arrojar luz sobre preguntas de investigación que aclaren cuestiones sobre el propio funcionamiento de la profesión y de las compañías de actores. Lanzaré una pregunta cuyos resultados de antemano sé que resultan acotados por tratarse de un periodo inicial en el camino de la profesionalización y que, por lo tanto, son manejables y pueden servir de ejemplo extrapolable a otros periodos. La cuestión que planteo es: ¿qué podemos saber o deducir, a partir de las noticias documentales reunidas en DICAT entre 1540 y 1600, sobre el proceso de incorporación de músicos y su profesionalización como tales en el marco de la formación de compañías? Son muchas las noticias que tenemos a partir de los contratos conservados sobre la participación de músicos en compañías, especialmente en el siglo XVII³. Los propios textos dramáticos dan cuenta de la importancia de esa participación. La base de datos *Digital Música Poética* (DMP), que forma parte de la federación de bases de datos ASODAT, de la que luego trataré, tiene por objetivo registrar y analizar esas canciones, estableciendo patrones musicales susceptibles de ser comparados con tonos conservados en los cancioneros de la época.

Volviendo a la búsqueda a la que hacía referencia, si preguntamos a DICAT cuántas mujeres consta en la documentación que habrían ejercido como músicas entre 1540 y 1600, y una vez leído el resultado que arroja, tan solo dos mujeres aparecen como tales junto a su función como actrices en el año 1600, María Hurtado y María de la Peña.

3 Florez Asensio (2015) indagó sobre el papel de los músicos en compañías, pero centrándose en el siglo XVII, y con un corpus más acotado que el que se encuentra documentado en DICAT.

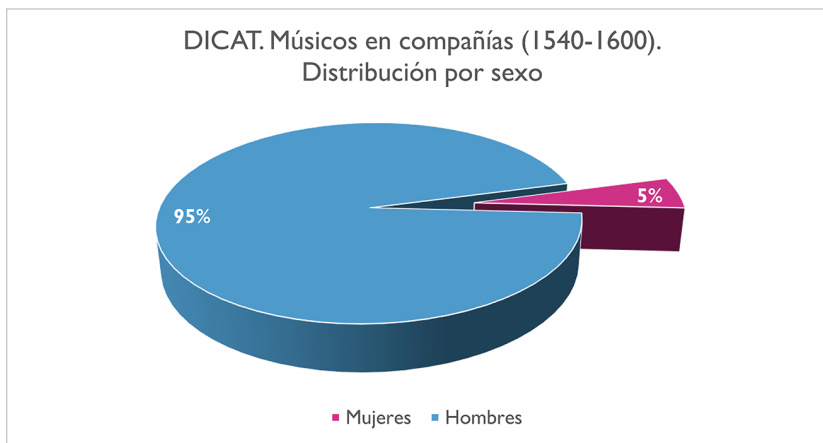


Imagen 1

Si la misma búsqueda la hacemos a partir de los hombres que consta que fueron músicos se arroja un resultado de 54 casos, aunque si los analizamos teniendo en cuenta solo aquellos que aparecen integrados como músicos en compañías, el resultado se reduce a 41, dado que algunos eran ministriles asalariados de municipios o catedrales o incluso en algún caso, como en los de Nicolás Navarro, Hernán Beranán, Antonio Beranán y Bartolomé Ginel, se trataba de ministriles de Su Majestad, que eran contratados puntualmente para participar en fiestas, como las del Corpus. En algún otro caso se documentan músicos asociados a casas de la nobleza como los de Juan Maldonado, Juan Sánchez Mudarra, Diego Núñez de Luna, Alonso de Morales y Marcos Sánchez, quienes fueron contratados en 1588 para servir como ministriles en la casa de don García Hurtado de Mendoza, recién nombrado virrey de Perú, tanto en España como en las Indias. El contrato, como se recoge en DICAT, estipulaba que no podrían tocar en ninguna fiesta sin permiso del virrey, lo que muestra que no debía ser infrecuente que los músicos trataran de obtener beneficios puntuales trabajando en fiestas. García Hurtado permaneció en Perú desde su llegada a comienzos de 1590 hasta 1596. Al menos uno de estos músicos de la casa del virrey, Juan Sánchez Mudarra, debió establecerse y per-

manecer en Lima tras su marcha, pues en 1603 era ministril de la catedral de la ciudad⁴.

El análisis de los datos que proporciona DICAT muestra que la incorporación de músicos a compañías estables se debió de llevar a cabo paulatinamente. Al parecer en un primer momento se trataba de colaboraciones puntuales, como la que los músicos Gaspar Díez y Francisco de la Vega declaraban en 1554 en relación con Lope de Rueda. El primero lo testimoniaba así: “cuando el dicho Lope de Rueda hace alguna comedia, si le llama y le paga bien su trabajo, tañe en la dicha comedia vihuela”. Pero ya a finales de siglo la relación que se establece en los contratos parece más duradera. Un ejemplo es el de José Esquivel, contratado en 1584 como músico para la compañía de Andrés de Angulo y que después sigue trabajando en las compañías de otros autores como Alonso de Cisneros o Tomás de la Fuente, hasta que se pierde su rastro documental en 1606. Las funciones que debía cumplir dentro de la compañía estarían relacionadas fundamentalmente con sus habilidades como músico, tal y como se detalla en un contrato de 1594, pues entre sus obligaciones figura: “buscar letras y romances y tonos nuevos para que se canten en las comedias”.

De la importancia que ya a fines de siglo XVI habían cobrado los músicos como parte importante de las compañías es muestra también el contrato que Gaspar de Porres firmó por duración de seis años con Francisco Villalta para formar parte de su compañía y “tañer y cantar y representar en aquella y dar tonos y letras a los demás músicos que anduvieren en dicha compañía”. Un contrato de una duración de seis años es excepcional en el contexto de la época, pues solían tener vigencia durante una temporada teatral o como mucho dos, en algunos casos. Un contrato de seis años manifiesta la importancia que Villalta tenía como músico. A pesar de las condiciones ventajosas que le ofrecía el contrato fir-

4 Pueden consultarse todos los datos que menciono aquí y los que siguen, así como las fuentes documentales de donde proceden, en DICAT (2023), en la entrada correspondiente a cada uno de los músicos mencionados.

mado con Porres, al año siguiente Villalta figura como miembro de la compañía de Alonso de Cisneros, otro de los grandes empresarios de la época. Los músicos se habían convertido ya en indispensables en este momento, pues tenían un papel muy relevante no solo en la interpretación de letras en las comedias, sino también en las aperturas y entreactos para la interpretación de los bailes, y era determinante su participación en las fiestas del Corpus, que tan importantes beneficios deparaban para las compañías, como sabemos. Un músico como Juan Vera recibía de Gaspar de Porres en 1596, por cantar “las letras que como tal músico se hubieran de cantar, y representar en entremeses”, 9 reales por representación, pero 9 ducados por participar en la fiesta del Corpus, una cantidad notablemente superior.

Las compañías italianas de la *commedia dell'arte*, como la del famoso Ganasa, probablemente se vieron obligadas a contratar músicos españoles para acompañar sus representaciones, que conocían bien los gustos musicales y las canciones que el público español sabría reconocer. En 1581 Ganasa contrata a los músicos Pedro de Salcedo y Antonio Laso para formar parte de la compañía tañendo con las guitarras y cantando tonadas.

Sin que pueda asegurarlo sin un estudio que abarque un periodo más amplio, la interpretación que brindan los datos que proporciona DICAT es la de que poco a poco surgieron profesionales que compaginaban habilidades musicales y actorales, algo que para una compañía debía de resultar más productivo a la hora de efectuar un contrato. Por otro lado, en el siglo XVII se produce un fenómeno que está bien documentado en los pueblos de los alrededores de Madrid, que es la representación en fiestas locales por parte de actores no profesionales, para las cuales se contrataban puntualmente a actrices y músicos, unos beneficios económicos complementarios para quienes formaban parte de compañía⁵.

5 Davis y Varey (2003), a la hora de revisar los protocolos del escribano Juan García de Albertos, cuya actividad se concentró en la región de Madrid entre 1634 y 1660, sacaron a la luz gran cantidad de documentos que testimonian esta actividad.

Evidentemente se trata de fragmentos de una realidad que debió de ser más compleja, pero que permiten extraer algunas conclusiones a partir de la reunión de datos. El ejemplo que he mostrado es abarcable por el número de datos que arroja la búsqueda, aunque sus conclusiones serán obviamente matizables en función del incremento de información que pueda darse en un futuro. Pero en el caso de otras indagaciones que queramos hacer sobre la base de datos, la cifra es tan elevada que exige un tratamiento previo para que el resultado que se ofrezca esté refinado y se haga más accesible para la interpretación de los investigadores. Por ejemplo, actualmente si hacemos una búsqueda por ciudades o poblaciones para hacer una determinada indagación, especialmente en aquellas, como es el caso de Madrid, sobre las que poseemos muchas noticias, el resultado es prácticamente inabarcable y no discrimina entre noticias de contratos, en los que figura la ciudad en la que se firma, y noticias de lugares en los que se representa, o lugares de residencia o de nacimiento de los firmantes, por ejemplo. Uno de los desafíos que tendremos que abordar en la siguiente fase del proyecto tiene que ver con la mejora de la posibilidad de realizar consultas más complejas a partir de DICAT. Pero para ello tendremos que someter los datos a un tratamiento previo de marcación, buscando herramientas que automaticen, al menos parcialmente, el proceso, dado el volumen de información, difícil de llevar a cabo en las circunstancias actuales de manera tradicional, con medios exclusivamente humanos. El objetivo es que las respuestas ofrezcan una discriminación entre datos que haga posible un estudio abarcador de los mismos por parte de los potenciales investigadores. Por ejemplo, si cruzamos los datos sobre los salarios de los miembros de las compañías, con su categoría profesional y función dentro de la misma, y con su evolución a lo largo de un determinado periodo, se puede arrojar luz sobre el funcionamiento económico de las compañías y la situación social de los actores, actrices y músicos, y facilitar su comparación en relación con otros gremios. Esto permitiría corroborar o desmentir, con acumulación suficiente de información, algunas ideas asumidas, como la de que

las primeras damas podían cobrar más que los galanes. ¿Hasta qué punto era esto una situación excepcional? ¿Cuál era la remuneración que recibían las mujeres frente a sus compañeros varones en relación con las funciones que cumplían dentro de las compañías? Si, como hemos visto, los músicos eran necesarios en las representaciones, ¿tiene esto un reflejo diferencial en sus salarios? ¿Y en el caso de las mujeres que actuaban como actrices, cantantes y bailarinas no solo en las representaciones, sino también en los entreactos de los espectáculos o en festividades, tiene esto un reconocimiento económico particular en su salario? Son preguntas cuya respuesta resultará más fiable en la medida en que seamos capaces de proporcionar datos suficientes y herramientas que faciliten su análisis.

Conscientes de esa necesidad de marcación previa, que permita obtener resultados de investigación más allá de noticias puntuales en relación con una obra, en su día, en el segundo proyecto iniciado por el grupo, se diseñó la base de datos CATCOM –esta ya desde el inicio en abierto–, incorporando un sistema de marcación por parte de los investigadores que permite no solo dar cuenta de noticias de representación vinculadas a años, sino extraer una serie de información que puede ser útil para plantear preguntas de investigación más complejas. Algo que con los medios técnicos con los que contábamos al iniciar DICAT, en los años 90, no había sido posible.

En sentido estricto CATCOM (<https://catcom.uv.es>) es un calendario de representaciones, que abarca desde 1.540 hasta 1.700 aproximadamente. En el momento de redactar estas páginas, la base de datos cuenta con 1.655 registros de títulos y 4.370 noticias de representación. Esperamos alcanzar alrededor de 2.500 registros de títulos de obras representadas en su última fase. Gracias a la marcación que desde el inicio realizamos los investigadores, se puede indagar en búsquedas por periodos, por lugares de representación, por directores de compañías o por tipología de obra. De momento, el mayor provecho sobre los datos se puede llevar a cabo de una manera interna, que deberemos pulir antes de

poder ofrecer la interfaz de búsqueda más refinada en abierto, aspecto que se implementará en la última fase. Plantearé algunas de las búsquedas que ya, con un elevado número de noticias acumuladas, podrán realizarse para llevar a cabo determinadas indagaciones.

Pongamos el caso de Calderón de la Barca. Como podemos ver en la imagen 2, el registro permite hacer búsquedas por fechas exactas o por periodos, por directores o directoras de compañía asociados a la representación de las obras de Calderón o por lugares de representación. También permite acotar la búsqueda definiendo si nos interesa realizarla sobre el conjunto de las noticias de representación de las obras dramáticas escritas por el autor o solo sobre autos, comedias o piezas breves.

Imagen 2

Pondré algún ejemplo de nuevo de una indagación que podría llevarse a cabo a partir de esta herramienta. Imaginemos que estamos interesados en los autos de Calderón para analizar en conjunto los datos que desde CATCOM podría arrojar una búsqueda a partir de las compañías que sabemos que representaron algunos de ellos y la circulación de los textos entre ellas. El listado resultante de esta búsqueda en estos momentos –recordemos que

se trata de bases de datos cuyos contenidos siguen aumentando—arroja una cifra de 73 autos sobre los que contamos con un total de 191 noticias de representación por parte de autores de comedias varones. Si nos interesa comparar con el número de autoras de comedias, es decir directoras de compañías, la cifra se reduce en dos tercios, pues existen en estos momentos 27 autos —entre los que se reparten un total de 44 noticias de representación—, que fueron puestos en escena por autoras de comedias.

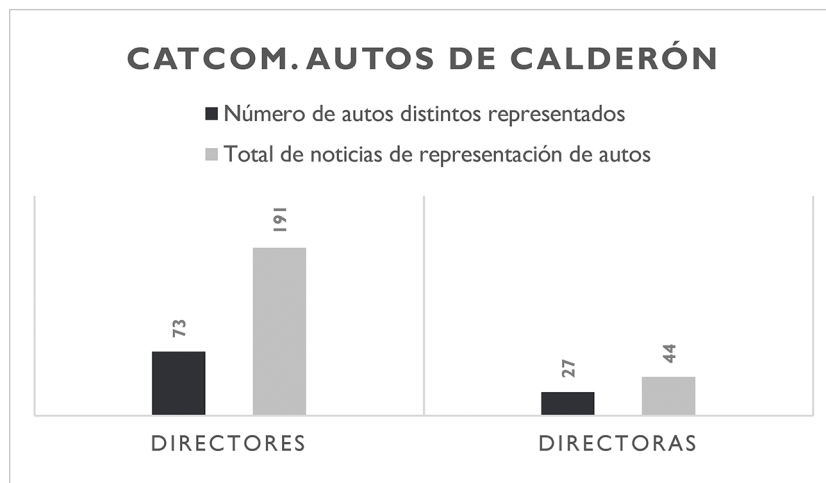


Imagen 3

Aunque menor que la de sus compañeros autores, la cifra puede parecer, incluso así, alta, dado el papel subordinado que la mujer tenía en la sociedad de la época. En realidad, si indagamos en la biografía de estas mujeres —cosa que se puede hacer por la vinculación establecida entre DICAT y CATCOM de manera interna y de manera externa a través de ASODAT—, tendremos que evaluar en qué medida se trataba de autoras que tenían licencia oficial para representar o no, o si se trataba de autoras que actuaban en nombre del marido autor por alguna circunstancia, lo que nos permitirá afinar las deducciones resultantes que podríamos hacer a partir de la cuantificación inicial.

Una indagación en la que introdujésemos como variable en la búsqueda por año los límites de 1600 y 1681, es decir los años de nacimiento y muerte de Calderón, nos permitiría saber cuántos autos escritos por Calderón cuentan con noticias ciertas de representación a lo largo de su vida, incluyendo alguno, especialmente del inicio de su carrera, cuyo texto no se conserva, como los titulados *Gradas de San Felipe*, *La casa del príncipe* o *San Juan y Corpus Cristi*. Hay que recordar que contar con noticias de representación de títulos cuyo texto no se ha conservado es algo frecuente, y contribuye a la ya de por sí compleja problemática del establecimiento fiable de autoría de muchos textos dramáticos.

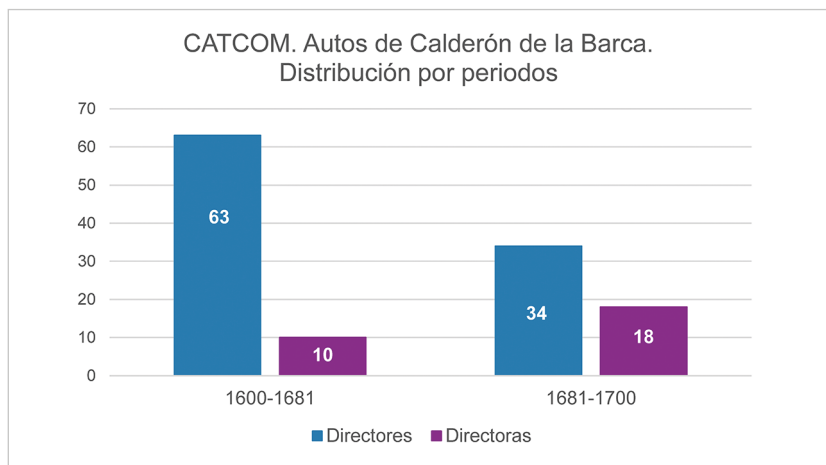


Imagen 4

La búsqueda que he enunciado daría como resultado un total de 10 autos representados por compañías dirigidas por mujeres, un 13,69%, y un total de 63 representados por compañías dirigidas por hombres, lo que supone un 86%. Si acotamos la búsqueda por años al periodo comprendido entre 1681 y 1700, ya fallecido Calderón, la distancia entre resultados es menor, pues se constatan 34 autos puestos en escena por compañías dirigidas por autores y 18 por compañías dirigidas por autoras, es decir, un 65%

frente a un 34,61%. A parte del descenso evidente de representaciones de autos de Calderón, ya una vez fallecido el dramaturgo, la cifra refleja el aumento relativo de mujeres que se incorporan a la dirección de compañías en el último cuarto del siglo XVII y una mayor presencia de las mujeres como autoras en el contexto de la fiesta del Corpus.

A la vez que trabajábamos todos estos años en estas bases de datos, otros grupos de investigación, y en particular me refiero ahora a CLEMIT, dirigido por Héctor Urzáiz (Universidad de Valladolid) y a MANOS, dirigido inicialmente por Margaret Greer (Duke University) y después codirigido con A. García-Reidy (Universidad de Salamanca), avanzaban en el trabajo de catalogación y estudio de censuras de textos dramáticos y de manuscritos teatrales, respectivamente. La consulta aislada de estas bases de datos ya ponía en evidencia la utilidad de la reunión de los resultados de las distintas informaciones que cada una podía aportar y que, en ocasiones, ayudaban a resolver incógnitas.

Pondré un ejemplo que muestra las posibilidades que abre la federación y el trabajo colaborativo entre los diferentes grupos que la integran a partir del caso de la obra titulada *El amor bando-lero*. Sobre ella sabemos, gracias a las noticias recopiladas en CATCOM, que fue representada en la Alhambra de Granada ante Felipe IV el 7 de abril de 1624 por la compañía de Antonio Granados; y que la volvió a representar unos años después la compañía de Alonso de Olmedo, como particular, ante el mismo monarca en Madrid, el 23 de diciembre de 1636.

Por otro lado, en MANOS (<https://manos.net>) se cataloga un manuscrito conservado en la Biblioteca Nacional de España. Además, este manuscrito incluye una licencia de representación fechada en Córdoba el 18 de septiembre de 1645, de la que da cuenta CLEMIT (<https://clemit.uv.es>).

A partir de esta licencia es posible deducir que se representó este año en el corral de esta ciudad y, por lo tanto, completar el elenco de representaciones en el calendario de CATCOM, añadiendo el año 1645.

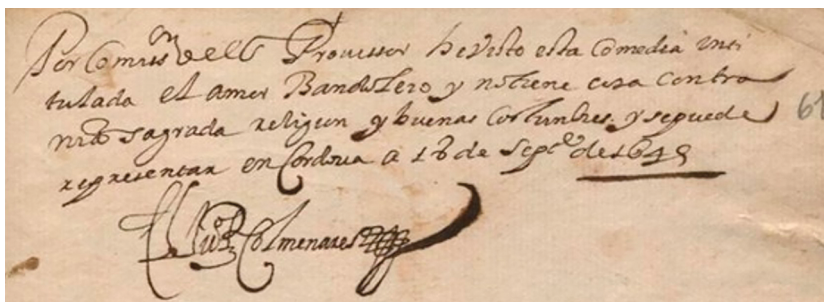


Imagen 5. El amor bandolero. BNE, mss. 14977, fol. 61r.
Licencia de representación

Por otro lado, si comprobamos a partir de DICAT qué compañía representó en Córdoba este año, haciendo la búsqueda correspondiente y una vez descartados los resultados no pertinentes, podemos aventurar con bastante certeza la hipótesis de que fue la compañía de Antonio de Rueda la que representó la obra, pues estuvo trabajando en la ciudad entre septiembre y octubre de este año.

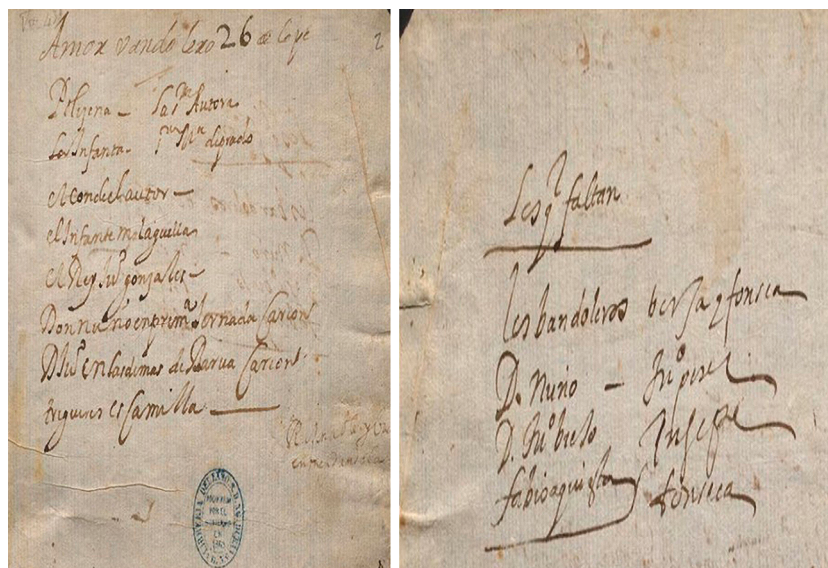


Imagen 6. BNE, mss. 14977, fol. 2r. y fol. 2v.

Pero, además, si acudimos de nuevo al manuscrito catalogado en MANOS y a la digitalización de éste conservada en la Biblioteca Nacional de España, podremos ver que en los folios 2r. y 2v. (imagen 6) se mencionan algunos nombres de actores en dos repartos parciales que proceden de manos distintas.

Aunque no tengamos la composición de la compañía de Antonio de Rueda en el año 1645, alguno de los mencionados en el reparto del fol. 2v. formó parte de ella en torno a esos años, como se comprueba en DICAT. Es el caso de Jusepe [Carrión], a quien encontramos en la compañía de Rueda en 1639, 1640, 1642, 1644 y 1646, y de [Pantaleón de] Borja, miembro de su compañía en 1639, 1640, 1642 y 1644. El mencionado como Fonseca, podría ser Nicolás Fonseca que había formado parte de su compañía en 1639. Todos estos datos, relacionados entre sí, corroboran la pertenencia a la compañía de Antonio de Rueda de este manuscrito y su utilización como material de trabajo para la puesta en escena⁶. Por otro lado, en MANOS, basándose en DICAT, y a partir de algunos de los nombres que figuran en el reparto del fol. 2r, se apunta que podría tratarse en este caso de miembros de la compañía de Diego Osorio de Velasco, y que la obra pudo representarse de nuevo en torno a 1658. Aunque la fecha de representación es difícil de determinar en este caso, resulta verosímil que este manuscrito pasase de manos de Antonio de Rueda a manos de Diego Osorio, pues este último fue actor de la compañía de Rueda durante varios años (está documentado en 1638, 1640, 1644 y 1646). En 1647 y 1648 Diego Osorio actuaba como apoderado de Rueda en los contratos de representación de su compañía, compañía que en la práctica debía haber pasado a gestionar, pues en un contrato de agosto de 1648 se explícita al referirse a Osorio: “a cuyo cargo está la compañía de Antonio de Rueda”. Este cambio en la gestión de la agrupación viene a coincidir con el momento en que Rueda parece empezar a

6 Cotarelo (1917, p. 362), al publicar la obra tuvo en cuenta el impreso y el manuscrito, señalando diferencias entre el texto conservado en el manuscrito y el de la edición de 1633: el manuscrito corrige errores del impreso y añade algunos versos.

ocuparse de otras actividades relacionadas con el mundo del teatro y a abandonar la dirección, llegando a convertirse en arrendador de los corrales de Madrid (así figura en 1653). Por su lado Osorio en 1649 ya figura como director de su propia compañía. En estas circunstancias cobra sentido la circulación del manuscrito de una compañía a otra. La relación de este segundo reparto con la compañía de Osorio permite que en MANOS se pueda apuntar la hipótesis de que uno de los varios copistas que intervino en el manuscrito pudo ser Gregorio González, quien en 1659 era apuntador de la compañía de Osorio.

Siguiendo con el ejemplo del manuscrito de *El amor bandolero*, conservado en la Biblioteca Nacional, en él se atribuye la obra a Lope de Vega. En las noticias de representación recopiladas en CATCOM a partir de la documentación existente, no se indica la autoría y tan solo se menciona el título, pero la obra fue publicada a nombre de Lope de Vega en la *Parte XXIV, extravagante* (Zaragoza, Diego Dormer, 1633), dato que junto con otros se recoge en ARTELOPE. Morley y Bruerton (1968, pp. 419-420), en su análisis métrico de las obras de Lope, rechazaron categóricamente que la comedia fuera suya. Desde 2022, dentro de la federación ASODAT, se van incluyendo enlaces a los informes de estilometría elaborados en ETSO (<https://etso.es>), proyecto que sostienen Germán Vega García-Luengos y Álvaro Cuéllar. Aunque el informe estilométrico no se encuentra publicado todavía en ETSO, Germán Vega ha tenido la amabilidad de indicarme que efectivamente el análisis no apunta a Lope y podría tratarse de una obra cuyo autor sería Jerónimo Bermúdez.

Esta es una muestra del modo en que la correlación de los datos puede ayudar a plantear sendas de investigación y a resolver en ocasiones algunas incógnitas. Cuantos más datos seamos capaces de relacionar y poner en común, más productividad encontraremos de cara a su explotación como investigadores. La red de posibles conexiones que se puede establecer entre las informaciones contenidas en las distintas bases de datos nos hizo plantearnos en 2016 la conveniencia de asociarnos con otros

grupos y crear la federación de bases de datos ASODAT (<https://asodat.uv.es>), plataforma que en su primera fase vio la luz en 2019⁷. Se partió de las mencionadas bases de datos CATCOM y CLEMIT, que se vieron transformadas para su integración, y paulatinamente se fueron creando e incorporando hasta el día de hoy las bases de datos DMP (Digital Música Poética), que cataloga y estudia canciones insertas en textos teatrales y dirige Lola Josa (Universitat de Barcelona); ISTAE, que dirige Alejandra Ulla (Universidad de Santiago de Compostela), base de datos que recoge y describe impresos sueltos publicados entre 1600 y 1834 que contengan comedias del Siglo de Oro; y la renovada DICAT, que ahora codirige conmigo Josefa Badía. Al mismo tiempo se está enlazando desde ASODAT con los mencionados informes de estilometría de ETSO y con bases de datos veteranas como son la mencionada MANOS y también ARTELOPE y EMOTHE, que dirigió Joan Oleza y ahora Jesús Tronch (Universitat de València), y nos proponemos en la siguiente fase la asociación con CALDERÓN DIGITAL, una base de datos cuyo trabajo dirige Fausta Antonucci (Università di Toma Tre) y que se ideó como tal a partir de una estructura de datos marcados inspirada en la de ARTELOPE.

Me referiré brevemente a las posibilidades que abre ISTAE, la más reciente base de datos incorporada a la federación. El grupo que dirige Alejandra Ulla se encarga de la catalogación de comedias sueltas, es decir impresos del XVII y del XVIII, que contienen una sola obra dramática de algún autor del siglo XVII. El interés del equipo investigador durante la primera fase del proyecto se dirige a una colección albergada en la BNE, no catalogada por distintas razones, que incluye 3.000 sueltas que carecen de datos

7 Las etapas del desarrollo desde la perspectiva técnica, que aquí no trato, pueden verse en Ferrer Valls, *et al.* (2021) y Ferrer Valls, Fuertes Seder (2023). Coordiné las dos primeras fases de este proyecto. La actual, que comenzó en septiembre de 2023, la coordina Alejandro García-Reidy. Agradezco a Ariadna Fuertes, Josefa Badía, Alejandro García-Reidy, Alejandra Ulla, Jesús Tronch, Héctor Urzáiz y Lola Josa su colaboración en este trabajo.

bibliográficos (año de publicación, lugar o imprenta), pero cuyas características tipográficas permiten atribuir las a imprentas del siglo XVII. En algunos casos ISTAE está aportando el descubrimiento de testimonios únicos de sueltas que ayudan a completar con textos dramáticos noticias de representación. Es el caso de *El rayo de Cataluña, primera parte* y *El rayo de Cataluña, segunda parte*, obras de Francisco Serrano Carimo de las que tan solo conocíamos que las representó la compañía de Agustín Coronel, pero carecíamos de textos. Ahora las dos sueltas descubiertas son identificadas en ISTAE como salidas de la imprenta de Francisco de Lyra y, por tanto, impresas antes de 1650. Si añadimos que la actividad conocida de Agustín Coronel, cuya compañía la representó no sabemos en qué año, solo se puede documentar en DICAT hasta 1643, esto nos permite afirmar que su fecha de composición es anterior a esta y situar cronológicamente al escritor, de cuya biografía nada se conoce. La inclusión de las sueltas de ISTAE en el marco de ASODAT contribuye a alimentar la red de informaciones que sobre el teatro áureo y desde diferentes perspectivas se había desplegado hasta la fecha. En un futuro cercano será posible, pues, trazar redes de relación o sociabilidad entre los diferentes agentes que participaron en el fenómeno teatral del siglo XVII: impresores, dramaturgos, directores de compañía, copistas, censores y músicos.

Una de las posibilidades que facilita ASODAT es la de agilizar la recuperación de los datos disponibles sobre la vida escénica de las obras de un dramaturgo, pues permite obtener a partir de un único metabuscador información sobre textos, manuscritos, licencias de representación y compañías teatrales. Un caso de muestra de estas posibilidades lo ofrece Guillén de Castro, que presenta ahora mismo ya numerosas entradas en las bases de datos federadas. El análisis sistemático de toda esta información puede permitirnos entender mejor la circulación y recepción de la producción de este dramaturgo y ahondar en las razones del mayor o menor éxito de unas obras respecto a otras. La trayectoria de Guillén de Castro como poeta dramático abarca más de un cuarto de siglo, desde los años finales del siglo XVI hasta alrededor de 1625, pero

su pervivencia en los tablados se extendió mucho más allá de su muerte, ocurrida en 1631.

Nos fijaremos en una de sus obras *Dido y Eneas*, cuya composición Bruerton situaba por razones métricas entre 1613 y 1616, aunque inclinándose hacia la segunda de estas fechas. Lope de Vega, al dedicar al autor valenciano *Las almenas de Toro* en la *Parte XIV* (Madrid, Juan de la Cuesta, 1620) afirmaba haber visto representar *Dido y Eneas* durante una estancia en Valencia. Bruerton argumenta en favor de que la estancia a la que alude sería la que realizó en 1616, aproximadamente entre el 2 de julio y el 9 de agosto, en que abandonó la ciudad, como se deduce de su correspondencia con el duque de Sessa (Bruerton, 1944, pp. 119-121). Si fue así, y se estrenó en Valencia en 1616 y la pudo ver Lope durante su estancia, se debió de estrenar en la casa de la Olivera y por parte de una de las compañías que actuaron por esas fechas en la ciudad: la de Francisco Mudarra o la de Diego López de Alcaraz, según se recoge en DICAT. Por otro lado, el día 8, el previo a la partida de Lope de Valencia, empezó a actuar la compañía de Hernán Sánchez de Vargas⁸. Si analizamos la reunión de noticias de diferentes bases de datos que nos facilita hoy por hoy ASODAT, podemos deducir que la obra se mantuvo en escena hasta años después. CATCOM documenta la representación que en junio de 1625 la compañía de

8 Como observa Bruerton (1944, p. 120), en una carta enviada por Lope desde Valencia a Sessa, el dramaturgo afirma haber estado durante veintidós días en cama, lo que nos permitiría estrechar más el cerco a la fecha en que pudo asistir a la representación. Pero prefiero no hacerlo, pues las afirmaciones de Lope no siempre se deben tomar al pie de la letra. Máxime en este caso, porque tenía que justificar ante su mecenas, siempre celoso de otras fidelidades, la dilación de su regreso a Madrid, para que no sospechase que el retraso tenía que ver con la inminente llegada del conde de Lemos y de la actriz Lucía de Salcedo. Efectivamente, justo antes de marchar Lope, llegó a puerto el conde don Pedro Fernández de Castro desde Barcelona, en una embarcación en la que viajaba con la compañía de Hernán Sánchez de Vargas, de la cual formaba parte la actriz Lucía de Salcedo, quien fue amante de Lope. El conde, gran mecenas y amante del teatro, regresaba de Nápoles, donde había ocupado el puesto de virrey.

Pedro de Valdés hizo para palacio como particular ante Felipe IV. El texto había salido a la venta a comienzos de este mismo año en la *Parte segunda de las comedias de don Guillén de Castro* (Valencia, Miguel Sorolla, 1625). A pesar de su impresión, la comedia circuló entre las compañías en forma de manuscritos, como testimonia el conservado en la Biblioteca Nacional de España, bajo la signatura mss 15020, que, según se indica en MANOS, muestra ser claramente una copia empleada por una compañía de actores. CLEMIT por su parte analiza la aprobación, otorgada por el doctor Pedro Garri-do en Valencia el 24 de octubre de 1653, que se incluye al final de este mismo manuscrito, que, sin embargo, no da cuenta de la compañía que pudo representarla en esa fecha. Pero DICAT nos permite completar la información y apuntar que fue la de Juan Vivas, que estuvo representando en la casa de la Olivera de Valencia durante el mes de octubre de 1653.

Este tipo de análisis que permite perseguir la circulación de una obra entre compañías y a lo largo del tiempo, resulta más productivo, como se muestra en este caso, si se pone en relación la información que hallamos en el conjunto de la federación. Si además hacemos extensible el análisis al resto de obras de un dramaturgo sobre las que podemos documentar noticias en el marco de ASODAT, podremos obtener un mapa de aquellas obras más representadas o de aquellas que pervivieron más en el tiempo, asociando estas noticias con el número de veces que se imprimieron y el número de manuscritos conservados, y podremos dar un paso más en nuestra investigación a partir del estudio de los textos mismos para tratar de apurar las razones que hicieron permanecer en el gusto del público algunas, mientras otras resultaron más efímeras.

Al comienzo me refería a la importancia que los músicos fueron adquiriendo como integrantes de las compañías de actores. DMP, que es una base de datos todavía en proceso de recopilación y tratamiento de la información, permitirá en su estadio más avanzado hacer estudios de conjunto sobre dramaturgos, analizar las preferencias musicales de cada uno reflejadas en su obra dramática y sus funcionalidades, establecer una relación con los distintos

géneros dramáticos cultivados por un autor o comparar sus usos con los de otros autores. El cruce de información, por ejemplo con CATCOM, referida a los años de representación, permitirá estudiar el mayor éxito de unos tonos musicales sobre otros a lo largo del tiempo; o el cruce de información con CLEMIT puede permitir indagar sobre la mayor o menor incidencia censora en los textos cantados, y si esta afectó más a unas modalidades que a otras.

Uno de los beneficios de la coordinación de grupos tiene que ver con la adopción de una metodología de trabajo común y el seguimiento de unos mismos protocolos para garantizar al máximo que la vinculación de la información, introducida desde cada una de las bases de datos, sea correcta. Explicaré muy brevemente para acabar algunos aspectos del funcionamiento interno que ASODAT ofrece al investigador que trabaja en el ámbito de la federación y, por tanto, desde la perspectiva de un trabajo colaborativo que va más allá de la propia base de datos de su grupo.

Cuando un investigador accede a la interfaz de gestión interna de su propia base de datos y desea dar de alta un título o un dramaturgo el sistema le conduce automáticamente a la interfaz de gestión interna de ASODAT (imagen 7). El formulario que se le presenta le facilita la búsqueda del título que pretende dar de alta en el resto de las bases de datos. Si no obtiene ningún resultado,

asodAT
BIENVENIDO/A, TERESA FERRER VALLS
CAMBIAR CONTRASEÑA
CERRAR SESIÓN

Edición

BIOGRAFÍAS
PUBLICACIONES
Autorización del IP
DRAMATURGOS
TÍTULOS CATCOM
Importar/Crear Títulos
CATCOM
Importar/Crear Dramaturgos

ASODAT - Gestión de Títulos Principales en CATCOM

Reinar después de morir FILTRAR

☐ Procedentes de otras bases de datos ☒ En todas las bases de datos ☐ Buscar sólo en títulos alternativos

Listado de 1 títulos principales, mostrando página 1 de 1

NUEVO TÍTULO PRINCIPAL

Título Principal	CATCOM	CLEMIT	DIGITALMP	MANOS	ISTAE	ARTELOPE	EMOTHE	ETSO	VISUALIZACIÓN WEB
Reinar después de morir	248	0	25	0	288	0	290	0	

© 2019 - Asociar los datos: Bases de datos integradas del teatro clásico español (ASODAT)
[CC BY-NC-SA]

Imagen 7

todavía puede emplear el mismo formulario para hacer una búsqueda solo entre los títulos alternativos. Esta manera de proceder tiene que ver con una casuística que en el teatro de la época resulta bastante habitual: una obra puede tener un título y variantes de títulos, algo que se complica cuando del ámbito de los impresos y de los manuscritos que nos han transmitido los textos pasamos al de la documentación, pues muchas veces los escribanos transmiten los títulos de manera parcial o con transcripciones de oído que resultan erróneas o confusas. Eso hace que sea muy importante que desde cada base de datos el investigador se asegure, antes de dar de alta un nuevo título, de que la obra no existe en el marco de la federación o de que no existe por un título alternativo distinto del que está buscando. La búsqueda por títulos alternativos le garantiza esta última posibilidad.

Otra de las casuísticas a la que nos enfrentamos tiene que ver con la existencia de títulos homónimos. Por ejemplo, pueden existir autos y comedias que comparten un mismo título, aunque no sean del mismo género dramático; o una obra puede contar con una versión seria y con una versión burlesca; o ser una refundición de otra anterior de un autor diferente. La interfaz de trabajo de ASODAT permite la consulta de toda información contenida en las distintas bases de datos, no solo de aquella que se encuentra publicada en abierto, sino también de aquella que todavía está en estado de elaboración. Tras la consulta de toda la información, el investigador podrá decidir si asocia la obra que posee con otra homónima o de título similar que ya exista en la federación o si crea un nuevo título. La creación de nuevos títulos debe ser verificada y autorizada por el investigador responsable de cada grupo para cada una de las bases de datos. Para la creación de dramaturgos se procede de igual modo: desde la interfaz de gestión se pueden consultar las biografías de dramaturgos y su producción dramática, información que puede servir de ayuda para establecer la asociación de un título.

A la hora de revisar la coherencia dentro del conjunto de nuestras decisiones independientes a partir de cada base de datos,

contamos con herramientas internas que sirven de ayuda para hacer la evaluación del trabajo de asociación. Por ejemplo, es posible cotejar automáticamente incoherencias entre bases de datos referidas a género, autores o títulos. Como se muestra en la imagen 8, un sistema de señalización por colores llama la atención sobre desajustes existentes entre distintas bases de datos: el rojo significa que no hay acuerdo entre ellas; el verde que si lo hay; y el amarillo que hay un acuerdo parcial entre algunas de ellas sobre el punto analizado.

asodat

BIENVENIDO/A, TERESA FERRER VALLS

🔑 CAMBIAR CONTRASEÑA

🚪 CERRAR SESIÓN

Edición

BIOGRAFÍAS

PUBLICACIONES

Autorización del IP

DRAMATURGOS

TÍTULOS CATCOM

Importar/Crear Títulos

ASODAT - Detalle Título Principal

TÍTULO PRINCIPAL: ■ Obras son amores, y no buenas razones

TÍTULOS RELACIONADOS EN ASODAT: ● Obras son amores

[VOLVER](#)

TÍTULOS ALTERNATIVOS:

GÉNERO: ● Auto ● Comedia ●

DRAMATURGOS RELACIONADOS ● Félix Lope de Vega Carpio ●

Imagen 8

En el ejemplo que puede verse en la imagen apreciamos una incoherencia en el título *Obras son amores* entre varias bases de datos. En este caso, hay acuerdo respecto al autor, pero no respecto al género. La consulta desde el registro interno de cada base de datos nos permite detectar el error de asociación que tiene que ver con que Lope de Vega posee dos obras de este título, un auto y una comedia, y en una de las bases de datos, al tratarse de obras homónimas, se ha asociado el auto a la comedia erróneamente. Esta detección automática de incongruencias facilita el trabajo de revisión de errores.

Lo prioritario en las dos primeras fases del proyecto coordinado ASODAT, finalizadas en 2023, era la propia creación de la federación y la suma de esfuerzos y bases de datos, recuperando las más veteranas y abriendo otras nuevas. Somos conscientes de que la mayor productividad de una base de datos, en nuestro caso de varias federadas en ASODAT, es tanto más alta y más representativa cuanto más información seamos capaces de reunir y procesar. Sin contenido relevante, por muy buena que sea la herramienta creada, no supera el cierre de la financiación de un proyecto, y muere. El avance de las nuevas tecnologías nos debe acompañar, pero no podemos empezar de cero cada vez el trabajo, desechando el ya realizado por los investigadores a lo largo de años, simplemente porque la herramienta ya no está en la vanguardia tecnológica o a la moda. Sería como desechar los manuscritos porque aparecieron los impresos. Después de estos años de trabajo, y ya con bastante información reunida, el desafío que tenemos delante ya no tiene que ver solo con ofrecer a los investigadores un buscador común para todas las bases de datos que permita visualizar el resultado de una búsqueda de conjunto, como se ofrece ahora, sino también con la creación de herramientas que favorezcan una explotación dinámica de los datos contenidos en las diferentes bases de datos de una manera más compleja para su utilización en nuevas investigaciones. En esta línea de trabajo esperamos poder continuar.

REFERENCIAS BIBLIOGRÁFICAS

- ARTELOPE = OLEZA, Joan y Jesús TRONCH (dirs.) et al., *ARTELOPE. Base de datos y argumentos del teatro de Lope de Vega*, <<https://artelope.uv.es/>> [consulta: 15/06/2023].
- ASODAT = FERRER, Teresa (coord.) et al., *ASODAT. Bases de datos integradas del teatro clásico español*, <<https://asodat.uv.es>> [consulta: 25/09/2023]
- BRUERTON, Courtney, “The chronology of the comedias of Guillén de Castro”, *Hispanic Review*, XII, 2, (1944), pp. 89-151.
- CATCOM = FERRER, Teresa (dir.) et al. *Base de datos de comedias mencionadas en la documentación teatral. 1540-1700*. CATCOM, <<https://catcom.uv.es>> [consulta: 3/08/2023].
- CLEMIT = URZÁIZ, Héctor (dir.) et al., *CLEMIT. Censuras y licencias en manuscritos e impresos teatrales*, <<https://clemit.uv.es>> [consulta: 11/08/2023].
- COTARELO, Emilio, *Obras de Lope de Vega publicadas por la Real Academia Española (Nueva Edición). Obras dramáticas, III*, Madrid, Real Academia Española, 1917.
- DAVIS, Charles y John E. VAREY, *Actividad teatral en la región de Madrid según los protocolos de Juan García de Albertos, 1634-1660. Estudio y documentos*, London: Tamesis Books, Fuentes para la Historia del Teatro en España XXXV-XXXVI, 2003.
- DICAT = FERRER VALLS, Teresa (dir.) et al., *Diccionario biográfico de actores del teatro clásico español*. DICAT, <<https://dicat.uv.es>>, [consulta: 07/07/2023]. Primera edición en DVD, Kassel, Edition Reichenberger, 2008.
- EMOTHE = TRONCH, Jesús (dir.) *Emothe. Teatro europeo de los siglos XVI y XVII: patrimonio y bases de datos* <<https://emothe.uv.es>>.
- ETSO = CUÉLLAR, Álvaro y Germán VEGA GARCÍA-LUENGOS. *ETSO: Estilometría aplicada al Teatro del Siglo de Oro*, <<https://etso.es>> [consulta: 20/08/2023].
- FERRER VALLS, Teresa, Ariadna FUERTES SEDER, Raúl PEÑA ORTIZ, Alejandro GARCÍA-REIDY, Dolores JOSA MARTÍNEZ y Héctor URZÁIZ TORTAJADA, “ASODAT: una plataforma de información sobre el teatro clásico español a partir de bases de datos federadas”, *Talía: Revista de estudios teatrales*, 3, (2021). pp. 45-58.

- FERRER VALLS, Teresa y Ariadna FUERTES SEDER, “Fundamentos y ampliación de la federación de bases de datos sobre teatro clásico español ASODAT (segunda fase)”, *Hipogrifo: Revista de Literatura y Cultura del Siglo de Oro*, 11. 1, (2023), pp. 15–37.
- FLÓREZ ASENSIO, María Asunción, *Músicos de compañía y empresa teatral en Madrid en el siglo XVII*, Kassel, Edition Reichenberger, 2015, 2 vols.
- ISTAE = ULLA LORENZO, Alejandra (dir.) et al., *Impresos sueltos del teatro antiguo español: base de datos integrada del teatro clásico español*, <<https://istae.uv.es>> [consulta: 12/09/2023]
- MORLEY, S. Griswold y Courtney BRUERTON, *Cronología de las comedias de Lope de Vega. Con un examen de las atribuciones dudosas, basado todo ello en un estudio de su versificación estrófica*, Madrid, Gredos, 1968.
- MANOS = GREER, Margaret R. y Alejandro GARCÍA-REIDY (dirs.), *Manos. Base de datos de manuscritos teatrales áureos*, <<http://www.manos.net>> [consulta: 15/08/2023].

EL USO DE LAS BASES DE DATOS Y LOS SIG EN ARQUEOLOGÍA.

BALANCE A TRAVÉS DEL ANÁLISIS DE ENCUESTAS, REVISIÓN HISTORIOGRÁFICA Y BIBLIOMETRÍA

Leticia Tobalina-Pulido, César González-Pérez

Resumen

Las bases de datos y los sistemas de información geográfica (SIG) son herramientas comunes en el ámbito arqueológico, pero ¿cómo de comunes?, ¿para qué se usan?, ¿cómo se aprende a usarlas adecuadamente? Para responder a preguntas como estas, durante el año 2021 se llevó a cabo una encuesta abierta, difundida a través de redes sociales y listas de difusión especializadas, dirigida a miembros de la comunidad arqueológica española.

En total se obtuvieron 134 respuestas de investigadores vinculados a centros públicos o privados de arqueología en el territorio nacional. El perfil demográfico de los encuestados incluyó representantes de prácticamente todos los estadios de la carrera académica hasta el nivel R3. La información recopilada fue complementada mediante una bibliometría.

Entre las conclusiones principales se destacan las tendencias y carencias en la formación de los arqueólogos en el uso de SIG y bases de da-

Leticia Tobalina-Pulido, Universidad de Cantabria
leticia.tobalina@unican.es
ORCID: 0000-0002-3315-5506

César González-Pérez, Incipit CSIC
cesar.gonzalez-perez@incipit.csic.es
ORCID: 0000-0002-3976-7589

tos, así como la diversidad en su aplicación práctica dentro del ámbito profesional. Este estudio proporciona una visión general sobre la integración de estas herramientas tecnológicas en la disciplina arqueológica.

Palabras clave: SIG, bases de datos, encuesta, análisis bibliométrico.

Abstract

Geographic Information Systems (GIS) and databases are common tools in the field of archaeology, but how common are they? What are they used for? How does one learn to use them properly? To answer questions like these, an open survey was conducted during 2021, disseminated through social media and specialized mailing lists, targeting members of the Spanish archaeological community.

A total of 134 responses were collected from researchers affiliated with public or private archaeology centers across the country. The demographic profile of the respondents included representatives from nearly all stages of the academic career up to the R3 level. The information collected was complemented by a bibliometric analysis.

Among the main conclusions, the study highlights trends and gaps in the training of archaeologists in the use of GIS and databases, as well as the diversity of their practical application within the professional field. This study provides an overview of the integration of these technological tools into the archaeological discipline.

Keywords: GIS, databases, survey, bibliometric analysis.

INTRODUCCIÓN

En los últimos 20 años, gracias a la popularización de los ordenadores con prestaciones gráficas y de procesamiento comparables a las estaciones de trabajo, se ha producido una auténtica revolución en el uso de herramientas digitales en la disciplina arqueológica. Esto ha permitido, además, una amplia difusión de estas herramientas a todos los niveles, estando actualmente presentes en prácticamente todos los proyectos de investigación (Menéndez-Marsh, 2023). Desde el inicio de la arqueología como disciplina, los investigadores han intentado compilar grandes volúmenes

de datos y encontrar maneras de procesar dicha información (Barceló, 2008). En este sentido, el gran volumen de datos y la diversidad de sus características suponen un verdadero problema para el arqueólogo (Tobalina-Pulido y González-Pérez, 2020), siendo las tecnologías de la información las que nos ofrecen múltiples opciones para hacer frente a la gestión, análisis y tratamiento de los datos. Entre estas soluciones, el empleo de los SIG (Sistemas de Información Geográfica) y de los Sistemas gestores de bases de datos se ha vuelto imprescindible en el trabajo arqueológico (Mínguez García y Capdevila Montes, 2016; Carrero-Pazos, 2023), ya que la enorme cantidad de datos recuperados hace inviable su procesamiento de manera manual. Por tanto, podríamos decir que su uso está en plena expansión en nuestra disciplina. Esto hace que resulte complicado e, incluso, como señala X. Rodier (2011: 20), presuntuoso, realizar un inventario de todas las investigaciones llevadas a cabo en la última década que utilicen este tipo de herramientas. Por ello, quisimos ir un paso más allá y comprobar cuál es la realidad de esta cuestión preguntando al colectivo arqueológico.

En este sentido, durante el año 2021 llevamos a cabo una encuesta abierta (difundida a través de las redes sociales y de listas de difusión especializadas) a la comunidad arqueológica. El objetivo era analizar el grado de formación y uso que se hace de las bases de datos y de los sistemas de información geográfica (en adelante SIG) en la disciplina. En total se consiguieron 134 respuestas de investigadores de nacionalidad española vinculados a algún centro público o privado de arqueología del territorio nacional, de perfiles demográficos muy variados, teniendo representados prácticamente todos los estadios de la carrera académica hasta el nivel R3. Consideramos, por tanto, para este trabajo, la edad límite de 45 años, dado que la edad media de consolidación investigadora en España está en 43 años (Precarios, 2022). A partir de las respuestas dadas y de la realización de un análisis bibliométrico de las publicaciones recogidas por las principales bases de datos bibliográficas, presentamos aquí algunas de las conclusiones extraídas de dicho estudio.

1. METODOLOGÍA

La metodología se ha dividido en dos fases: en primer lugar, la elaboración y difusión de la encuesta; en segundo lugar, la realización de una bibliometría.

1.1. LA ENCUESTA

Creada mediante *Google Forms*, se planteó una encuesta completamente anónima, con solo algunas preguntas demográficas generales para poder analizar los datos considerando valores como edad y posición académica/profesional, dado que la realidad de perfiles en arqueología es muy variada (Vaquerizo, 2015: 101). A partir de ahí, optamos por dividirla en dos bloques: uno dedicado a los SIG y otro a las bases de datos, con preguntas muy similares como las que siguen y que unificamos aquí para ambos casos: 1) ¿Has usado alguna vez SIG/bases de datos), 2) ¿Consideras importante el uso de SIG/bases de datos en arqueología?, 3) ¿Para el estudio de qué periodos has empleado este SIG/bases de datos? o 4) ¿Qué uso has hecho de SIG/bases de datos? En todos los casos se ha optado por preguntas con respuestas de opción múltiple.

En cuanto a la muestra, es de tipo no probabilístico por conveniencia, es decir, los participantes no fueron seleccionados de manera aleatoria, sino que la encuesta se difundió a través de redes sociales (Facebook, Twitter) así como por correo electrónico a diversas universidades y centros de investigación durante 5 meses a lo largo del año 2021.

1.2. EL ANÁLISIS BIBLIOMÉTRICO

Para el análisis bibliométrico hemos optado por centrar la búsqueda en *Web of Science Core Collection* (WOS) y *Scopus*. La selección final de los artículos se efectuó el 20 de julio de 2023. En primer lugar, se eligieron términos o palabras clave para la búsqueda como “SIG”, “GIS”, “Arqueología”, “Archaeology”, “database”. Esta búsqueda se realizó en tres campos específicos: “título”, “abstract” y “palabras clave”. Se utilizó el método booleano para delimitar una búsqueda avanzada en las bases de datos utilizando los

operadores “AND” y “OR”. A continuación, se hizo una selección específica de términos en inglés que incluyera los periodos históricos, los principales *softwares* de SIG, así como otros complementarios y algunas palabras más específicas relacionadas con los SIG y las bases de datos en arqueología. Así, en WOS se realizó bajo el esquema de búsqueda siguiente: TS=(GIS OR QGIS OR ArcGIS OR “Geographic* Information System*” OR SAGAGIS OR “SAGA-GIS” OR GisArchaeo OR GISci OR GRASSGIS OR 3DGIS OR “3D-GIS” OR WEBGIS OR “WEB-GIS” OR “WEB_GIS” OR “MOBILE-GIS” OR MOBILEGIS OR “MICRO-GIS” OR MICROGIS OR “GIS-functions” OR “Global Information System*” OR *database*) AND TS=(site OR pottery OR settlement OR landscape OR territory OR Prehistory OR Paleolithic OR Palaeolithic OR Mesolithic OR “Middle Stone Age” OR Epipaleolithic OR Epipalaeolithic OR Neolithic OR “New Stone Age” OR “Copper Age” OR Chalcolithic OR Eneolithic OR Aeneolithic OR “Bronze Age” OR “Iron Age” OR “Ancient history” OR “Classical antiquity” OR “classical era” OR “classical period” OR “classical age” OR “post-classical history” OR “Late antiquity” OR “Middle Ages” OR “medieval period” OR “Renaissance” OR “Modern era” OR “modern period” OR “Industrial Age” OR “Contemporary history”) AND TS=(*Archeolog* OR *Archaeolog*). Esto devolvió 2668 resultados. A esta búsqueda se aplicó un filtro por país, lo que redujo el número de resultados a 257.

A continuación se hizo una búsqueda en Scopus con la siguiente estrategia: (TITLE-ABS-KEY(GIS OR QGIS OR ArcGIS OR {Geographic* Information System*} OR SAGAGIS OR {SAGA-GIS} OR GisArchaeo OR GISci OR GRASSGIS OR 3DGIS OR {3D-GIS} OR WEBGIS OR {WEB-GIS} OR {WEB_GIS} OR {MOBILE-GIS} OR MOBILEGIS OR {MICRO-GIS} OR MICROGIS OR {GIS-functions} OR {Global Information System*} OR *database*) AND TITLE-ABS-KEY(site OR pottery OR settlement OR landscape OR territory OR Prehistory OR Paleolithic OR Palaeolithic OR Mesolithic OR {Middle Stone Age} OR Epipaleolithic OR Epipalaeolithic OR Neolithic OR {New Stone Age} OR {Copper Age} OR Chalcolithic OR Eneolithic OR Aeneolithic OR {Bronze Age} OR {Iron Age} OR {Ancient

history} OR {Classical antiquity} OR {classical era} OR {classical period} OR {classical age} OR {post-classical history} OR {Late antiquity} OR {Middle Ages} OR {medieval period} OR {Renaissance} OR {Modern era} OR {modern period} OR {Industrial Age} OR {Contemporary history}) AND TITLE-ABS-KEY(*Archeolog* OR *Archaeolog*)). Esta búsqueda devolvió 3237 resultados, reduciéndose a 266 tras la aplicación del filtro de país de España.

Tras dichas búsquedas se unificaron los resultados de ambas bases de datos y se eliminaron los duplicados, primando los registros de la WOS. Así, nos quedamos con un total de 339 registros. Después, aplicamos un filtro propio en base a nuestra experiencia como investigadores en la materia sobre el total de registros, eliminando aquellos que no estaban relacionados con la temática pese a haber aparecido en la búsqueda. Tras este filtrado nos quedamos con un total de 327 registros. Finalmente, se detectó y eliminó uno de los resultados de Scopus (2-s2.0-85023199500) cuyos autores son de una institución argentina, a pesar de que en el país ponía “Spain”. Por tanto, el total de registros para el análisis es de 326. Para algunos análisis hemos optado por excluir las publicaciones del año 2023, dado que la búsqueda la realizamos a mediados del mismo y no contábamos con el volumen final de textos publicados.

2. RESULTADOS Y DISCUSIÓN

2.1. ENCUESTAS

De las 134 respuestas obtenidas, hemos conseguido una muestra que tiene representación en todas las franjas etarias entre los 18 y los 45 años desde la edad en la que se inicia la etapa universitaria hasta la edad media en la obtención del R3, certificado de investigador establecido del Ministerio de Ciencia, Innovación y Universidades. La franja de edad que más respuestas ha aportado es la de mayor edad, entre 41 y 45 años, y la que menos la de 18 a 25 años, con 32 y 10 contestaciones, respectivamente.

En cuanto a la profesión de las personas encuestadas, el grupo mayoritario, con 32 respuestas, se dedica a la arqueología

profesional o de empresa, seguido de los investigadores independientes, con 20. Los grupos más minoritarios son los de Profesor Ayudante Doctor, Contratado Doctor y Titular de Universidad, con apenas 6 respuestas en conjunto (Fig. 1). Si bien no tenemos cifras del número total de investigadores y/o profesionales a nivel estatal para cada grupo profesional, no es de extrañar ese considerable número de arqueólogos de empresa, dado que el volumen de profesionales de dicho sector es bastante importante, pudiendo reflejar la muestra, en cierta manera, la realidad arqueológica nacional (Parga y Varela, 2014).

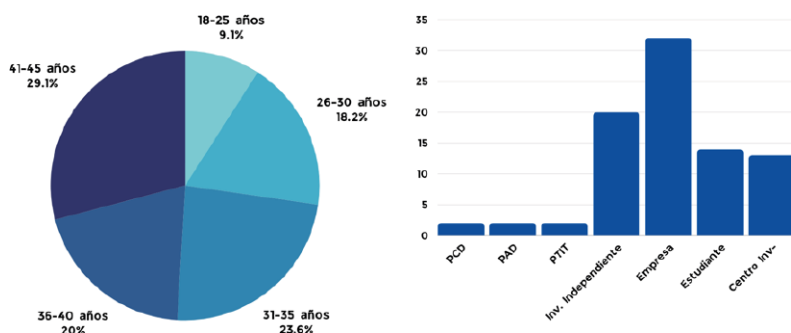


Figura 1. Porcentaje de la muestra por grupos etarios (izquierda) y número de respuestas por categoría profesional (derecha)

Centrándonos en la representación a nivel de comunidades autónomas y provincias, el mayor número de respuestas procede de la Comunidad Autónoma de Madrid, si bien hemos conseguido obtener respuestas de prácticamente todo el territorio nacional. Así, aunque no contamos con respuestas procedentes de todas las provincias, sí registramos respuestas de todas las comunidades autónomas, salvo las ciudades autónomas de Ceuta y Melilla y la comunidad de La Rioja, lo que permite partir de un abanico bastante representativo (Fig. 2). No es de extrañar que la Comunidad de Madrid sea la más representada, dado que, según los datos del informe realizado en 2014 por el CSIC, “la mayor parte de las empresas [...] se localizan en la Comunidad Autónoma de Madrid,

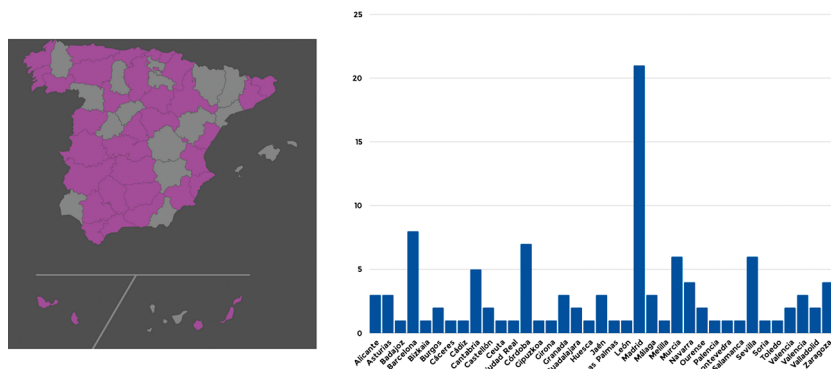


Figura 2. Distribución de la muestra por provincias

en donde se concentra el 19,8% de la muestra; le sigue Andalucía con un 17,9% y Cataluña con un 10,4%” (Parga y Varela, 2014: 12).

Centrándonos ya en las preguntas de la encuesta sobre los SIG y las bases de datos, a la cuestión: “¿Has usado SIG en algún proyecto de Arqueología (incluido TFG, TFM, tesis)?”, un 15,2% de los encuestados indica que no ha empleado los Sistemas de Investigación Geográfica. Esto se acentúa para la consulta: “¿Has usado bases de datos en algún proyecto de Arqueología (incluido TFG, TFM, tesis)?”, ya que un 9% no ha utilizado las bases de datos en proyectos arqueológicos (Fig. 3). Esto resulta bas-

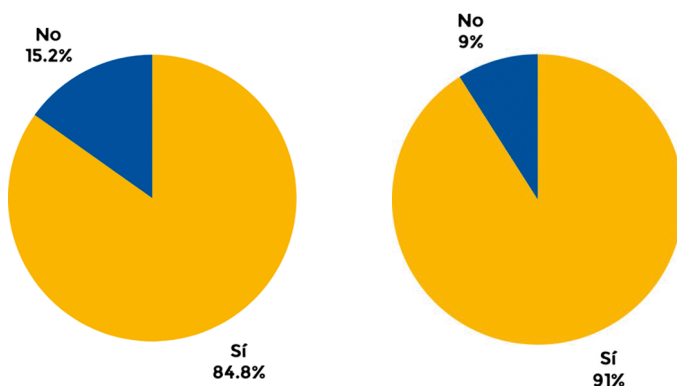


Figura 3. Porcentaje de encuestados que ha empleado las herramientas (SIG a la izquierda y bases de datos a la derecha)

tante significativo si consideramos que “la digitalización y virtualización en arqueología se viene desarrollando desde hace décadas gracias al gran avance de la tecnología” (Acevedo y Staropoli, 2023: 379).

Por otro lado, resulta relevador que, ante la pregunta de por qué consideran que es importante el uso de base de datos en arqueología, una mayoría de las respuestas indica que el motivo es la gestión de los datos (30 respuestas), seguido de la ordenación de datos, con 25 respuestas. Es interesante, sin embargo, que, de las contestaciones dadas, solo 22 de ellas indican que su empleo es importante para agilizar la entrada de datos (Fig. 4). También es llamativo que, aunque sea un número insignificante, ya que solo se registra una respuesta, se indique que el uso de las bases de datos no es importante en arqueología, cuando la rápida asimilación de los sistemas informáticos “[...] ha hecho que nuestra disciplina se viera, de manera pionera, fortalecida en sus posibilidades de interpretación” (Baena Preysler y Ríos Mendoza, 2006: 15).

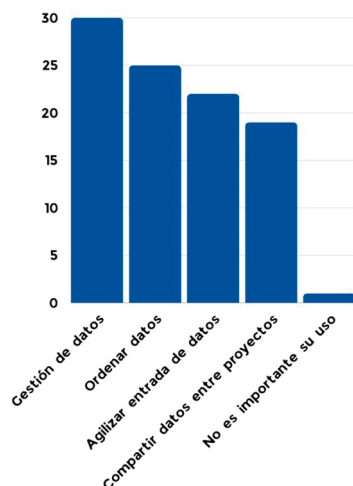


Figura 4. Respuestas dadas a la pregunta de por qué consideran importante el uso de bases de datos en arqueología

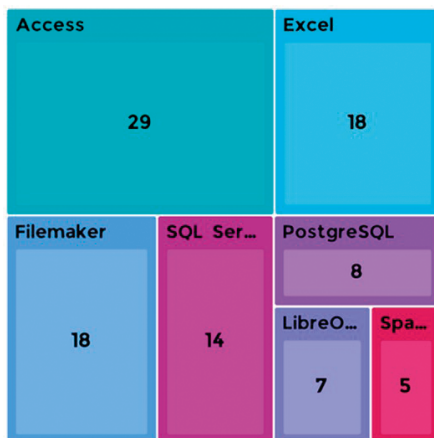


Figura 5. Productos *software* que se consideran como Sistema Gestor de Bases de Datos (SGDB)

Otra cuestión relevante es qué productos *software* se consideran un Sistema Gestor de Bases de Datos (SGDB) y cuáles se emplean como tal realmente. En este sentido, Access (29 casos) es el que más personas consideran como SGDB seguido de Excel y FileMaker (18 respuestas, respectivamente), otros programas como SQL Server, PostgreSQL o SpatiaLite tienen menos consideración como tal, siendo realmente SGDB (Fig. 5).

Cuando preguntamos sobre el programa que emplean para implementar las bases de datos, Access vuelve a ser el más mencionado, si bien LibreOffice Base (que en la pregunta anterior solo aparecía en 7 respuestas) es otro de los más utilizados (22 casos). Le siguen Excel y FileMaker, con 18 y 14 respuestas, respectivamente. Finalmente, Postgre, SpatiaLite y SQL Server solo se mencionan en 7 casos en conjunto (Fig. 6).

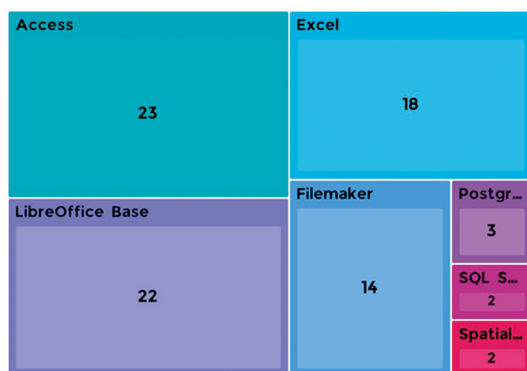


Figura 6. *Software* empleado para hacer bases de datos según las respuestas de los encuestados

Además, se realizaron pruebas de independencia chi-cuadrado de Pearson para estudiar la dependencia de algunos pares de variables. Se utilizó un valor de alfa de 0,05. Los resultados pueden verse en la tabla (Fig. 7).

Así pues, se puede observar que el uso de bases de datos o SIG no parece depender de la edad de las personas que respondieron a las encuestas. En este sentido, considerando un nivel de significancia de alfa = 0,05, se observa que la edad de las personas no influye significativamente en el uso de estas tecnologías, ya

VARIABLES	TEST CHI-CUADRADO	INTERPRETACIÓN
Edad / Uso de BBDD	0,3863	Independientes
Edad / Uso de SIG	0,8536	Independientes
Formación / Uso de BBDD	0,0388	Dependientes
Formación / Uso de SIG	0,0000	Dependientes
Uso de BDD / Uso de SIG	0,0630	Independientes

Figura 7. Resultados de las pruebas de independencia realizadas entre pares de variables

que los valores p obtenidos tanto para la relación entre edad y uso de bases de datos (0,3863) como para edad y uso de SIG (0,8536) son superiores a alfa. Es decir, que el uso de estas herramientas está distribuido de manera homogénea entre los distintos grupos etarios.

Sin embargo, la formación recibida sí se correlaciona con el uso de ambas tecnologías, con más intensidad en el caso de los SIG. Así, los valores p obtenidos para la relación entre formación y uso de bases de datos (0,0388) así como entre formación y uso de SIG (0,0000) son menores a alfa, lo que muestra dependencia entre las variables. Esto indica que las personas con mayor formación tienen más probabilidades de utilizar este tipo de herramientas. En este sentido, es algo lógico que el uso de los Sistemas de Información Geográfica (SIG) en arqueología esté estrechamente relacionado con la formación del usuario, pues para poder utilizar correctamente los SIG es necesario contar con un cierto nivel de formación (del tipo que sea) en estas tecnologías.

Finalmente, el análisis de la relación entre el empleo de bases de datos y el uso de SIG muestra un valor p de 0,0630, es decir, ligeramente superior a alfa. Esto sugiere que, técnicamente, estas dos variables son independientes, es decir, que la utilización de una de estas tecnologías no está necesariamente asociada al empleo de la otra. Sin embargo, dado que el valor p está muy próximo del nivel de significancia, es necesario tomar esta conclusión con precaución.

2.2. ANÁLISIS BIBLIOMÉTRICOS

Para poder orientar el análisis optamos por plantear algunas preguntas guía: ¿cómo evoluciona el volumen de publicaciones por años?, ¿y desglosado por universidades?, ¿en qué revistas se publican más trabajos de este tipo y en qué cuartil están? o ¿qué tipo de publicaciones son las más representadas?

En cuanto a la evolución del volumen de publicaciones por años (Fig. 8) podemos observar varias cuestiones interesantes.

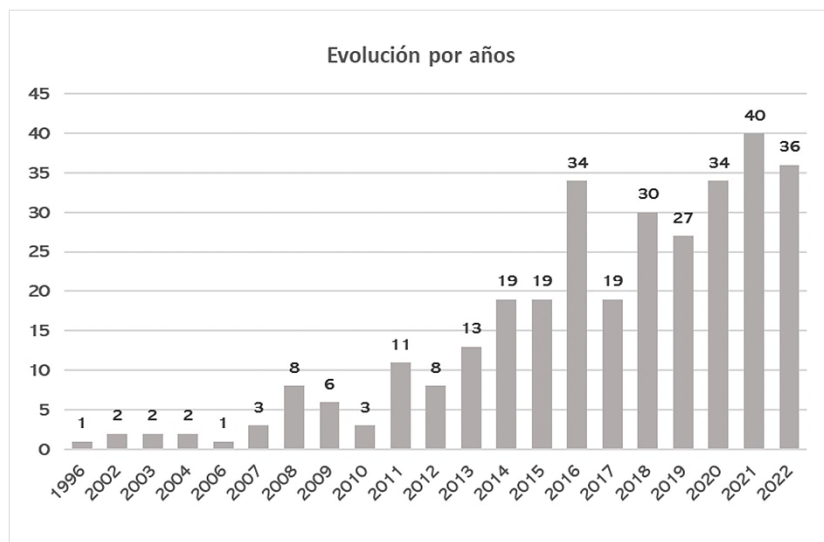


Figura 8. Evolución de las publicaciones por años

En primer lugar, entre los años 1996 y 2007 se observa un número muy bajo de publicaciones. Es importante tener en cuenta, sin embargo, que los datos analizados provienen de plataformas como Web of Science, creada con ese nombre en 1997, y Scopus, establecida en 2004. Esto implica que el bajo número de publicaciones observado antes de estas fechas, especialmente entre 1996 y principios de los años 2000, podría estar influido por la limitada existencia de este tipo de plataformas para indexar publicaciones científicas en ese momento, más allá de considerar que esto pueda

deberse a una falta general de adopción de estas tecnologías en la arqueología.

El primer cambio significativo se produce en 2008, donde las publicaciones relativas a dicho año aumentan a 6, marcando un pequeño pico que podría coincidir tanto con una mayor accesibilidad al *software* SIG como con el fortalecimiento de plataformas bibliográficas *online* como Scopus, que empezaron a recopilar de manera más sistemática las publicaciones producidas desde las diferentes disciplinas. Además, a partir de ese momento el incremento de las publicaciones relacionadas con las temáticas que nos conciernen aquí es bastante progresivo. Hay que tener en cuenta que en España los Sistemas de Información Geográfica estaban ya empezando a afianzarse en esos años. Ejemplo de ello es que en el año 2006 se inició la celebración de los cursos de posgrado en “Tecnologías de Información Geográfica aplicadas a la Arqueología” organizados anualmente desde entonces conjuntamente por el Instituto de Arqueología de Mérida y el Instituto de Ciencias del Patrimonio, ambos centros del CSIC, y que han permitido formar a una gran cantidad de investigadores en el uso de los SIG. Además, durante esos primeros años del siglo XXI, se publican algunas de las obras de referencia sobre esta cuestión (Grau Mira, 2006; Connolly y Lake, 2006).

Por otra parte, el periodo comprendido entre 2015 y 2017 muestra un crecimiento aún más evidente, alcanzando un máximo de 34 publicaciones en 2016 y manteniendo cifras similares en los años siguientes. En este sentido, en la última década, el desarrollo de los SIG y su aplicación en arqueología es indiscutible y los autores que emplean esta tecnología en sus investigaciones (con mayor o menor acierto) se han multiplicado. Basta hacer una búsqueda en Dialnet para observar el creciente número de publicaciones que los emplean en arqueología entre los años 2010 y 2018, principalmente manuales y tesis doctorales (Rodier, 2011; Mínguez García y Capdevila Montes, 2016; Mayoral Herrera *et al.*, 2017, son algunos ejemplos). Finalmente, entre 2018 y 2022 se observa un crecimiento sostenido en el número de publicaciones, alcanzando el máximo

de 40 en 2021. Este auge refleja no sólo un empleo continuado de las herramientas, sino que también puede coincidir con el “parón” que supuso el 2020 con la pandemia de la Covid-19 y que conllevó una eclosión de publicaciones durante el año siguiente. Finalmente, aunque en 2022 se registra una ligera disminución a 36 publicaciones, este número sigue siendo elevado, lo que evidencia que el uso de estas herramientas se han consolidado en la investigación arqueológica.

A nivel de universidades, se han identificado 48 centros entre las publicaciones analizadas. Destaca entre ellas la Universidad de Barcelona, que aparece en 34 de los documentos, seguida de la Universidad Complutense de Madrid con 23 y las universidades Autónoma de Barcelona y de Sevilla con 22 cada una. En los tres casos se trata de algunas de las universidades españolas con departamentos de arqueología más potentes de la península ibérica. Además, ofertan el grado de arqueología y másteres de especialización en la disciplina, lo que aumenta el número de egresados con filiación en dichos centros. En la tabla (Fig. 9) se muestran las 10 universidades que más aparecen en la filiación de las publicaciones.

POSICIÓN	UNIVERSIDAD	Nº DOCS.
1	Barcelona	34
2	Complutense de Madrid	23
3	Autónoma de Barcelona	22
4	Sevilla	22
5	Granada	21
6	País Vasco	19
7	Cantabria	18
8	Rovira i Virgili	18
9	Alicante	14
10	Jaén	13

Figura 9. Universidades con mayor número de publicaciones

POSICIÓN	UNIVERSIDAD	Nº DOCS. 2013-2022
1	Barcelona	28
2	Complutense de Madrid	20
3	Granada	20
4	Sevilla	18
5	País Vasco	17
6	Cantabria	16
7	Autónoma de Barcelona	15
8	Rovira i Virgili	15
9	Jaén	13
10	Salamanca	13

Figura 10. Ranking de universidades en función del número de publicaciones en los últimos 10 años

Por otro lado, en 51 de las 326 publicaciones no aparece ninguna universidad española en la firma de los autores (15,64%), estando los investigadores vinculados a otro tipo de instituciones.

Si atendemos a los resultados de los últimos 10 años (Fig. 10) se puede observar, además, que la Universidad de Barcelona y la Universidad Complutense de Madrid se sitúan en los primeros puestos de la tabla, mientras que, a nivel general, el resto de las universidades siguen apareciendo, aunque con algunas variaciones en cuanto al orden. Así, la Universidad de Alicante ya no está en la lista de las 10 primeras posiciones, pero sí aparece Salamanca con 13 documentos publicados entre 2013–2022.

Por otro lado, si atendemos al QS World University Rankings 2024 para la disciplina “Arqueología” (QS, 2024), las universidades que están en los primeros puestos en España son: la Universidad Complutense de Madrid, que ocupa el puesto 50, seguida de la Universidad de Sevilla y la Universidad Autónoma de Barcelona. También están presentes en el ranking la Universidad de Barcelona, la Autónoma de Madrid, Universidad de Santiago de Compostela, la Universidad Pompeu y la Universidad de Granada.

Centrándonos en el idioma de las publicaciones, el predominante es el inglés, dado que existe un sesgo de lenguaje al haber utilizado las bases de datos WoS y Scopus para extraer los datos, por lo que realmente este indicador no nos aporta información relevante.

En cuanto a la tipología documental de los textos, la mayor parte de los documentos son artículos, representando más del 86% de la producción científica analizada. Esto pone de relieve la publicación de múltiples artículos buscando la rentabilidad científica y esa cultura de la rapidez en la Academia que se está impulsando en los últimos años (Berg *et al.*, 2022). En este sentido, conviene precisar que en 9 de los registros se indican dos tipologías diferentes. Se han duplicado para poder tener todas las tipologías en cuenta, por lo que la suma es mayor. Por ejemplo, la publicación que se indica como *data paper* también tiene la tipología “artículo” (Fig. 11).

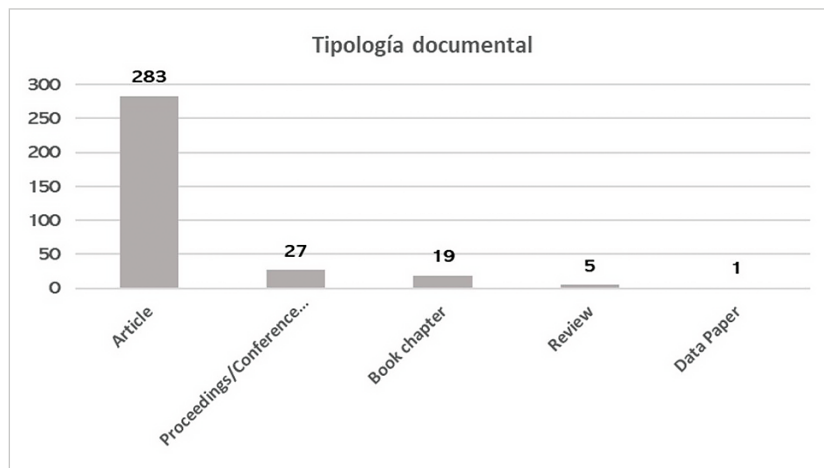


Figura 11. Número de resultados por tipología de documentos

Si observamos la evolución del tipo de publicación, se aprecia que el número de artículos ha ido incrementándose progresivamente hasta sufrir un aumento más acentuado a partir del año 2012. Así, los datos sobre la tipología documental de las publicaciones en arqueología, extraídos de Web of Science y Scopus, muestran un patrón claro en cuanto a las formas en las que se difunde la investigación científica en el campo de la arqueología. En primer lugar, los artículos destacan como la forma de publicación predominante, con un total de 283 documentos, lo que refleja la tendencia general en la investigación académica de priorizar las publicaciones en revistas científicas indexadas. Esto no es de extrañar, ya que la ANECA, agencia de acreditación española, valora más positivamente los artículos revisados por pares y publicados en revistas de impacto. Dado que tanto WOS como Scopus son bases de datos líderes en el ámbito académico, las publicaciones en estas plataformas son consideradas de alto impacto, lo que explica la alta concentración de artículos en estos repositorios. Sí se pueden detectar algunos picos en los que el número de artículos se reduce drásticamente, como por ejemplo en 2017, así como otros momentos de incremento notable, caso de los años 2016 o 2021

(Fig. 12). Las actas de conferencias ocupan el segundo lugar, con 27 publicaciones, seguidas de los capítulos de libro (19).

En cuanto a la evolución que sigue el tipo de publicación, desde 2009 los artículos científicos (línea azul) experimentan (Fig.12) un aumento constante, alcanzando su punto más alto en 2021 con, aproximadamente, 35 publicaciones. En este sentido, aunque se observa un ligero descenso en el año 2022, los artículos siguen siendo la categoría dominante. Aquí entrarían en juego dos cuestiones. Por un lado, el mayor número de publicaciones que se produjo tras la Covid-19 y, por otro, las demandas de la propia Academia en los méritos académicos específicos.

Los capítulos de libros (línea naranja) muestran una evolución intermitente, con picos en 2016 y 2018 y un ligero repunte en 2020 que se mantiene hasta el final del periodo estudiado. Esto podría estar relacionado con proyectos editoriales específicos o compilaciones temáticas, si bien es difícil de interpretar dado el bajo número de publicaciones y las escasas modificaciones que se producen en la tendencia. Las actas de conferencias (línea amari-

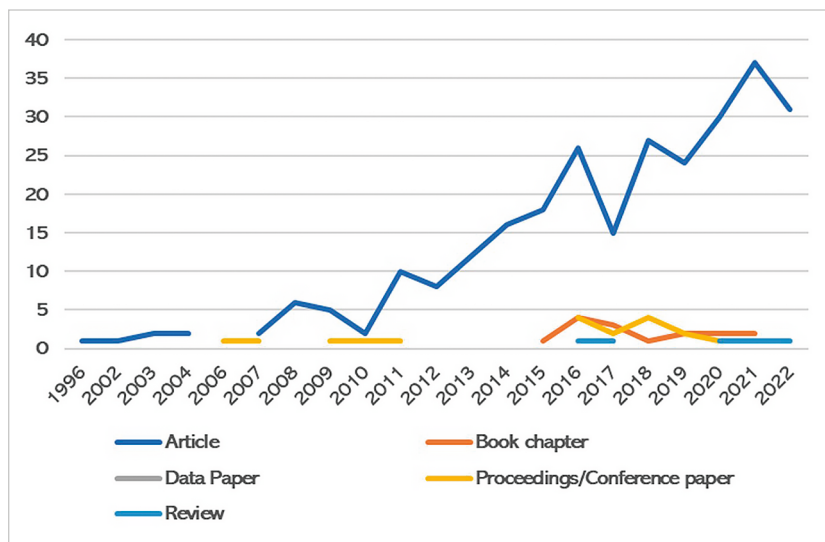


Figura 12. Evolución de la tipología de los documentos por años

lla) tienen una representación limitada, si bien parece que son más visibles desde 2015. Finalmente, los *data papers* (línea gris) prácticamente no tienen representación, lo que indica que la publicación de datos estructurados sigue siendo una práctica poco desarrollada en arqueología pese a los mandatos actuales de *open Access* que invitan a la publicación de las bases de datos en los artículos científicos.

Pasando a continuación a profundizar algo más en el análisis, nos hemos centrado en las publicaciones en revistas, incluyendo las categorías documentales “artículo” y “reseña”, lo que supone un total de 288 registros. Se han obviado, así, las categorías de “capítulo de libro” y “actas de congreso”. El resultado obtenido es de 133 revistas, de las cuales 92, es decir, un 69,17%, solo cuentan con una publicación relacionada con la temática objeto de estudio (Fig. 13).

Entre las revistas con más artículos publicados sobre la temática que nos concierne, destacan *Journal of Archaeological Science-Reports* y *Journal of Archaeological Science*, con 19 y 15 documentos, respectivamente. En cuanto a las revistas españolas en la lista de las 12 con más publicaciones estarían: *Virtual Archaeology*

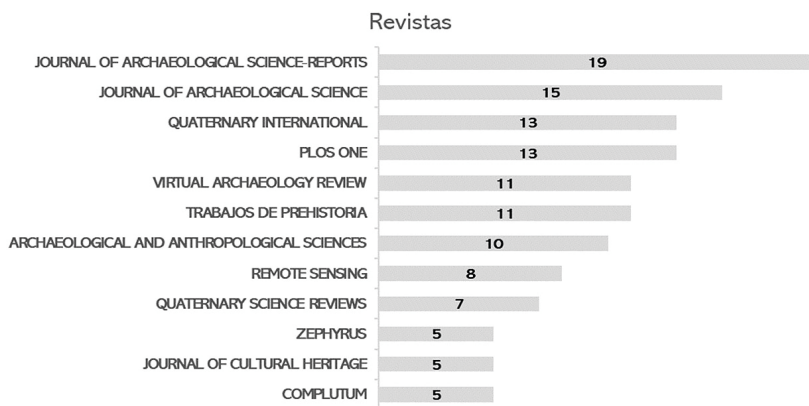


Figura 13. Revistas con mayor número de documentos publicados sobre la temática

Review, *Trabajos de Prehistoria*, *Zephyrus* y *Complutum*, con 11, 11, 5 y 5 documentos, respectivamente (Fig. 13). En todos los casos se trata de algunas de las revistas de mayor impacto en la disciplina a nivel estatal, siendo la primera de ellas una de las revistas de referencia en la publicación de trabajos arqueológicos relacionados con la modelización 3D y la fotogrametría.

En cuanto al impacto de las publicaciones en las que aparecen los documentos, hemos utilizado el *Journal Citation Reports* (JCR), base de datos creada por el *Institute for Scientific Information* (ISI), que recopila datos estadísticos para evaluar y comparar la relevancia de revistas científicas en diferentes áreas de estudio. Para este análisis hemos considerado solo las publicaciones cuya tipología documental es “artículo” y “reseña”.

En este sentido, 233 artículos tienen un índice de impacto, de los cuales 93 proceden de los índices AHCI, ESCI o BKCI-SSH. Además, hay que tener presente que para el JCR 2022, las revistas del AHCI y del ESCI tienen factor de impacto, pero no cuartil. Por tanto, hemos extraído los cuartiles de las 140 publicaciones restantes (60,1%). Así, 105 documentos están en Q1 o Q2, lo que supone un 74,9% (Fig. 14).

CUARTIL / INDEX	Nº DOCS.	% Nº DOCS.
Q1	62	44,1%
Q2	43	30,8%
Q3	30	21,5%
Q4	5	3,6%
TOTAL	140	100,0%

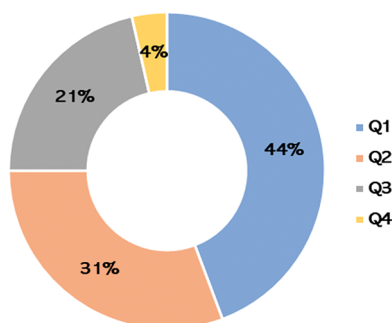


Figura 14. Índice de impacto de las revistas

En conclusión, el uso de herramientas tecnológicas como SIG y bases de datos en arqueología está cada vez más consolidado, aunque su adopción depende de la formación de los usuarios. Además, el aumento de publicaciones científicas en los últimos años

refleja el interés creciente en estas tecnologías, con un predominio claro de artículos publicados en revistas de alto impacto. Esto pone de relieve su importancia en la investigación arqueológica y el reconocimiento que tienen en el ámbito académico, especialmente en revistas internacionales y en proyectos vinculados a universidades de prestigio.

3. DISCUSIÓN Y CONCLUSIONES

En los últimos 20 años, la arqueología ha cambiado mucho gracias a los avances tecnológicos, especialmente con el uso de ordenadores más potentes y herramientas digitales. Los sistemas de información geográfica (SIG) y las bases de datos han revolucionado la forma en que las arqueólogas y los arqueólogos gestionan y analizan los grandes volúmenes de datos que se obtienen en las intervenciones arqueológicas (excavaciones, prospecciones, etc.). Estas tecnologías han hecho más fácil organizar la información, interpretarla de manera más precisa y presentarla de forma clara. Sin embargo, el crecimiento en el uso de estas herramientas aún no ha sido tan rápido como se pudiera esperar, y no todo el colectivo arqueológico parece que más las utiliza de forma habitual. Los Sistemas de Información Geográfica han evolucionado en arqueología desde que se empezaran a utilizar los primeros sistemas en la década de 1980. En la actualidad se han convertido en herramientas fundamentales para cualquier estudio arqueológico que trabaje con datos espaciales. En este sentido, su valor radica en sus capacidades de gestión de información y su potencial para el análisis espacial. Así, el progresivo incremento del uso de estas tecnologías ha sido impulsado por la disponibilidad de datos, el desarrollo de aplicaciones *software* con interfaces más amigables e intuitivas así como la posibilidad de utilizar recursos de código abierto. Desde 2010, los estudios que emplean los análisis espaciales han crecido progresivamente, utilizándose principalmente con fines de visualización y para ayudar en la recopilación y almacenamiento de datos arqueológicos, (Menéndez *et al.* 2023).

El estudio que aquí hemos presentado, con una muestra de 134 arqueólogas y arqueólogos de diferentes edades y trayectorias profesionales muestra que, aunque estas herramientas son cada vez más comunes en la disciplina, aún hay una parte del colectivo que no las utiliza o que tiene cierta reticencia a emplearlas. Muestra de ello es que un 9% nunca ha utilizado bases de datos o que un 15,2% no haya trabajado con los SIG. Las razones pueden ser diversas: la falta de formación, las limitaciones económicas o incluso la falta de acceso a las herramientas adecuadas.

Por otro lado, se ha observado que la formación recibida por los usuarios es uno de los factores más importantes a considerar cuando se trata de utilizar estas tecnologías. En este sentido, el estudio muestra que los arqueólogos con más formación son los que más suelen utilizar los SIG y las bases de datos en sus investigaciones. Esto sugiere que, aunque las herramientas tecnológicas estén disponibles, el conocimiento previo sobre cómo utilizarlas influye mucho en su empleo.

En cuanto a las herramientas específicas, se constata una tendencia a usar *software* accesible y conocido, como Access, Excel o FileMaker, en lugar de opciones más especializadas y avanzadas como Postgre-SQL o SQL Server. Aunque estas herramientas más comunes pueden ser útiles en algunos casos, no son las más adecuadas cuando se manejan grandes volúmenes de datos o se necesitan capacidades avanzadas. Esto sugiere que las arqueólogas y los arqueólogos a menudo optan por lo que ya conocen o lo que es más fácil de obtener, por ejemplo, en función de las licencias o *software* disponible en sus laboratorios y centros.

Finalmente, la bibliometría pone en evidencia un progresivo aumento de las publicaciones de temática arqueológica que emplean los SIG y las bases de datos. Este incremento ha sido mayor en la última década, cuando se ha producido una eclosión de las publicaciones que emplean estas herramientas.

Si bien este pequeño estudio que hemos presentado aquí es solo un modesto aporte, permite obtener una visión bastante completa de cuál es el uso que se está haciendo de las tecnologías en la disciplina arqueológica.

Agradecimientos

Nos gustaría agradecer a Marta Benítez Berrocal, bibliotecaria de la Universidad de Navarra (Servicios de apoyo a la investigación, Unidad de Bibliometría), su ayuda con la elaboración de la bibliometría que aquí presentamos. Por otro lado, la primera autora fue beneficiaria durante la realización de parte de este artículo de un contrato parte de la ayuda FJC 2020-043923-I, financiado por MCIN/AEI/10.13039/501100011033, y por la Unión Europea “NextGenerationEU/PRTR”. Instituto de Ciencias del Patrimonio. Incipit, CSIC. CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS. Edificio Fontán, Bloque 4, Despacho 220. Cidade da Cultura Monte Gaiás s/n, 15707 Santiago de Compostela (Spain).

REFERENCIAS BIBLIOGRÁFICAS

- ACEVEDO, Verónica y Laura STAROPOLI, “La digitalización de la arqueología en un mundo de patrimonio virtual y Ciencia Abierta (Versátil, Flexible y Expeditiva)”, *Revista del Museo de Antropología*, 16(3), (2023), pp. 377-394. <https://doi.org/10.31048/1852.4826.v16.n2.42498>.
- BARCELÓ, Miquel, *Una historia de la informática*, Barcelona, Editorial UOC, 2008.
- BERG, Maggie y Barbara K. SEEBER, *The Slow Professor: Desafiando la cultura de la rapidez en la academia*, Granada, Editorial Universidad de Granada, 2022.
- CARRERO-PAZOS, Miguel, *Arqueología computacional del territorio*, Archaeopress, 2023.
- CONOLLY, James y Mark LAKE, *Sistemas de información geográfica en arqueología*, Cambridge, Cambridge University Press, 2006.
- GRAU MIRA, Ignacio (ed.), *La aplicación de los SIG en la arqueología del paisaje*, Alicante, Universidad de Alicante, 2006.
- MAYORAL HERRERA, Victorino, César PARCERO OUBIÑA y Pastor FÁBREGA ÁLVAREZ, *Archaeology & Geomatics, Harvesting the benefits of 10 years of training in the Iberian Peninsula (2006-2015)*, Leiden, Sidestone Press, 2017.
- MENÉNDEZ-MARSH, Fernando, Mohammed AL-RAWI, João FONTE, Rita DIAS, Luís Jorge GONÇALVES, Luís GONÇALVES SECO, João HIPÓLITO, José Pedro MACHADO, Jorge MEDINA, José MOREIRA, Tiago DO PEREIRO,

- Marta VÁZQUEZ y António NEVES, “Geographic Information Systems in Archaeology: A Systematic Review”, *Journal of Computer Applications in Archaeology*, 6(1), (2023), pp. 40–50. <https://doi.org/10.5334/jcaa.104>.
- MÍNGUEZ GARCÍA, María del Carmen y Enrique CAPDEVILA MONTES (Coords.), *Manual de tecnologías de la información geográfica aplicadas a la arqueología*, Comunidad de Madrid, Museo Arqueológico Regional – Ilustre Colegio Oficial de Doctores y Licenciados en Filosofía y Letras y en Ciencias de la Comunidad de Madrid, 2016.
- PARGA DANS, Eva y Rocío VARELA POUSA, *Caracterización de la arqueología comercial en España*, Santiago de Compostela, Instituto de Ciencias del Patrimonio (Incipit) CSIC, 2014.
- PRECARIOS, “¿Cuánto cobra un/a postdoc en España?”, Precarios.org, <https://precarios.org/article376-Cuanto-cobra-un-a-postdoc-en-Espana#footnote1> [consulta: 19/12/2024].
- QS, QS World University Rankings by Subject, QS.com, <https://www.qs.com/tag/qs-world-university-rankings-by-subject/> [consulta: 19/12/2024].
- RODIER, Xavier, «Les SIG appliqués à l’archéologie», en *Information spatiale et archéologie*, X. RODIER (Dir.), O. BARGE, L. SALIGNY, L. NUNINGER y F. BERTONCELLO, Lugar de edición, Errance, 2011, pp. 13–32. <https://shs.hal.science/halshs-00956200>.
- ROMANÍ, Franco, Charles HUAMANÍ y Gregorio GONZÁLEZ-ALCAIDE, “Estudios bibliométricos como línea de investigación en las ciencias biomédicas: Una aproximación para el pregrado”, *CIMEL: Ciencia e Investigación Médica Estudiantil Latinoamericana*, 16(1), (2011), pp. 52–62.
- TOBALINA-PULIDO, Leticia y César GONZÁLEZ-PÉREZ, “Valoración de la calidad de los datos arqueológicos a través de la gestión de su variedad. Aplicación al estudio del poblamiento tardorromano”, *Complutum*, 31(2), (2020), pp. 343–360. <https://doi.org/10.5209/cmpl.72488>.
- VAQUERIZO GIL, Desiderio, “Arqueología para un futuro incierto: La profesión de arqueólogo tras la crisis devastadora del «pelotazo»”, *Pyrenae: Journal of Western Mediterranean Prehistory and Antiquity*, 46(2), (2015), pp. 89–120. <https://doi.org/10.1344/Pyrenae2015.vol-46num2.5>.

DISEÑO DEL ACCESO A UN CATÁLOGO SOBRE LA ANTIGUA GRECIA

Ana García Serrano, Fernando Chuquimarca,
Francisco Paños Merino, César Fernández Sánchez

Resumen

El capítulo presenta un proyecto de Humanidades Digitales¹ que combina métodos informáticos y humanísticos para digitalizar un catálogo sobre la Antigua Grecia desarrollado por el filoheleno D. César Fernández (autor). La investigación abarca la organización de una colección multimedia compuesta por textos, imágenes y planos, y detalla las decisiones tecnológicas involucradas en el desarrollo del catálogo: ¿Cuál es la estructura de información y la organización de ficheros del corpus? ¿Qué tipo de visualización es útil para los potenciales usuarios de la información? ¿Cómo se quiere navegar por el corpus multimedia o catálogo? ¿Qué tipo de búsquedas de información pretenden los expertos de dominio o estudiosos que les sean respondidas? Entre las herramientas utilizadas, se destacan Solr para indexación, Python para procesamiento de datos, y WordPress para la visua-

1 <https://arqueogriegos.iatext.ulpgc.es/>.

Ana García Serrano, ETSI Informática
UNED, agarcia@lsi.uned.es
ORCID: 0000-0003-0975-7205

Fernando Chuquimarca, ETSI Informática
UNED, fchuquimarca@lsi.uned.es

Francisco Paños Merino, ETSI Informática
UNED, fpanos3@alumno.uned.es

César Fernández Sánchez
cesarcesarfer@telefonica.net

lización de contenidos. El visor del catálogo digital incluye información estructurada sobre museos y yacimientos arqueológicos de 8 de las regiones de la Antigua Grecia. Se subraya el potencial de estas tecnologías para optimizar el acceso y análisis de datos, y se identifican desafíos como la escalabilidad y la integración de nuevas funcionalidades. Este proyecto también considera las implicaciones de la preservación digital y la accesibilidad abierta para la comunidad académica y el público general. El trabajo representa un esfuerzo interdisciplinario que combina informática, arqueología e historia, sirviendo como modelo para proyectos similares en Humanidades Digitales.

Palabras clave: corpus multimedia, acceso a la información, gestión y organización de información, Antigua Grecia.

Abstract

The chapter introduces a Digital Humanities project combining computational and humanistic methods to digitalize a catalogue on Ancient Greece developed by the philhellene Mr. César Fernández (author). The research covers the organization of a multimedia collection comprising texts, images, and plans, detailing the technological decisions involved in the catalog's development: What is the information structure and file organization of the corpus? What type of visualization is useful for potential users of the information? How do you want to navigate the multimedia corpus or catalogue? What type of information search do domain experts or scholars expect to be answered? Tools such as Solr for indexing, Python for data processing, and WordPress for content visualization are highlighted. The catalog includes structured information on museums and archaeological sites in 8 regions of Ancient Greece. The potential of these technologies to optimize data access and analysis is emphasized, along with challenges such as scalability and the integration of new features. The project also addresses the implications of digital preservation and open access for academic and general audiences. This work represents an interdisciplinary effort combining computing, archaeology, and history, serving as a model for similar initiatives in Digital Humanities.

Keywords: multimedia corpus, information access, information management and organization, Ancient Greece.

1. INTRODUCCIÓN Y ANTECEDENTES

Las Humanidades Digitales (HD) se encuentran en pleno desarrollo y continuamente surgen nuevas tecnologías que mejoran a las anteriores o que abren nuevas rutas para la consecución de los objetivos planteados en diversos proyectos de carácter multidisciplinar.

En general, las técnicas más utilizadas pueden agruparse en técnicas cualitativas, para el modelado y la organización de los datos, su etiquetado y anotación, la interpretación ontológica u otras y en técnicas cuantitativas, por ejemplo, para la recuperación de información, la clasificación de textos, la minería de textos para la identificación de temas o aspectos abstractos, identificación de entidades nombradas, análisis de sentimientos y otras tareas relacionadas con textos. Para ello se utilizan métodos del área informática de Inteligencia Artificial (IA) como son las redes de neuronas artificiales, los métodos de análisis estadísticos o probabilísticos en general o técnicas de aprendizaje automático profundo, así como los modelos del lenguaje que en los últimos dos años han revolucionado el panorama de las aplicaciones (García-Serrano y Menta-Garuz, 2022). En 2018 apareció una nueva arquitectura desarrollada por Google, que marcó el inicio de una nueva era en el PLN (Devlin *et al.*, 2019). Se trata de una aplicación a partir de un modelo informático que está pre-entrenado con una cantidad inmensa de datos: 800 millones de palabras del *BooksCorpus*² y 2.500 millones de palabras de *Wikipedia* en inglés (y mucha otra información de datos tomados de la actividad de usuarios en la web). Bien conocidas son aplicaciones como ChatGPT³ (aún con funcionalidades gratis), Midjourney⁴ (aplicación solo de pago) o Copilot⁵ (disponible para Microsoft 365), y están revolucionando la

2 <https://paperswithcode.com/dataset/bookcorpus>.

3 <https://openai.com/chatgpt/>.

4 <https://www.midjourney.com/home?callbackUrl=%2Fexplore>.

5 <https://learn.microsoft.com/es-es/microsoft-365-copilot/microsoft-365-copilot-overview>.

forma de trabajar y el ocio, y su rápida implantación en actividades profesionales y de investigación está siendo impactante, aunque son gestionados por grandes corporaciones e instituciones. Estas aplicaciones han sido diseñadas (entrenadas) para lenguas (principalmente el inglés) conocidas y no para su uso en textos muy técnicos o históricos (García-Serrano y Menta-Garuz, 2021), aunque su evolución es muy rápida (Menta, Sánchez-Salido y García-Serrano *et al.*, 2022; Sánchez-Salido, Menta y García-Serrano *et al.*, 2023).

En este trabajo se han utilizado tecnologías más o menos convencionales en informática para que conformen un proyecto como el que se presenta, sobre la colección multimedia de documentos o “Catálogo universal de museos y yacimientos arqueológicos de la Antigua Grecia” que contiene una descripción de 154 museos y 277 yacimientos arqueológicos con más de 20.000 textos, casi 10.000 imágenes y más de mil planos en AutoCAD (García-Serrano, Paños y Fernández, 2023).

Afrontar una colección de estas características exige una serie de decisiones sobre las tecnologías disponibles que faciliten la realización de ciertas tareas más o menos complejas (o funcionalidades en terminología relacionada con herramientas software). Hay que conocer las posibilidades, estudiar su adecuación a los problemas de proyecto planteado y diseñar la solución que resuelva el acceso de los usuarios con interés en la información contenida.

Una máxima en este tipo de proyectos multidisciplinarios es que siempre hay que tener en cuenta la estructura aportada por el estudioso, y en este caso el catálogo se ha estructurado en 20 regiones distintas o zonas parciales de estudio. Estas regiones coinciden, en el caso de Grecia actual, con las de su propia configuración regional. En el caso de Turquía, Italia y Chipre, la región coincide con el propio país.

De todas las regiones del catálogo, en este momento, solo tenemos información disponible de **8 regiones**, que constituyen una muestra suficiente para desarrollar el trabajo informático para su almacenamiento, procesado y acceso digital más adecuado a las características intrínsecas tanto de la estructura como

de la información contenida en el catálogo. Estas son: (1) ATE-NAS: 9 Yacimientos y 4 Museos; (2) ÁTICA: 13 Yacimientos y 1 Museo; (3) EL PELOPONESO: 45 Yacimientos y 5 Museos; (4) STERÉA ELLADA: 27 Yacimientos y 4 Museos; (5) TESALIA: 10 Yacimientos y 3 Museos; (6) ÉPIRO: 8 Yacimientos y 2 Museos; (7) MACEDONIA: 12 Yacimientos y 2 Museos; (8) TRACIA: 3 Yacimientos y 1 Museo.

A su vez, cada región tiene tres tipos de información clasificada en Geografía e Historia, Varios y Yacimientos. Cada tipo de información contiene en subcarpetas los documentos aportados por el autor en formato .doc (sub-carpeta entradas), las fotos que van relacionadas a las entradas (en formato .jpg) (sub-carpeta fotos) y los planos que van relacionados a las entradas (en formato CAD) (sub-carpeta planos).

Según se trate de la sub-colección de yacimientos o la de museos, el contenido de los tres tipos de fichero (texto, foto o plano) es diferente. Cada **museo arqueológico** consta de dos partes bien diferenciadas: la información literaria y la información gráfica. La información literaria de un museo consta, en general, de un único apartado en donde aparece la situación concreta del museo dentro de la localidad en que se encuentra, y una descripción del material expuesto más importante, estructurada sobre la base de las diferentes SALAS que constituyen el museo. La información gráfica consta de (a) el Plano de la planta del museo con su distribución en SALAS, elaborado (AutoCAD) generalmente a partir del escaneado de los planos que proporciona el propio museo, y que se han reelaborado con las leyendas en español; (b) Fotos propiedad del autor. Constituyen el 90% de la información gráfica. Se han realizado durante la visita al museo y están ordenadas congruentemente con la descripción de este. Cada una de ellas tiene un “pie de foto” en el que se da cuenta del elemento concreto fotografiado; (c) En muy escasas ocasiones, fotos extraídas de internet o escaneadas de alguna publicación. Cada **yacimiento arqueológico** consta también de información literaria y gráfica, aunque diferente a la de los museos. Por ejemplo, hay información literaria sobre

la forma de acceso, historia, mitología, descripción del yacimiento y del museo (si lo hay) del mismo (en formato texto). La información gráfica contiene fotos aéreas del yacimiento (extraídas generalmente de internet, planos generales del yacimiento (elaborados con AutoCAD) o fotos propiedad del autor.

Respecto a la tecnología disponible, y sin ánimo de ser exhaustivos se citan las siguientes herramientas básicas para este tipo de trabajos relacionados con las colecciones creadas y gestionadas por humanistas.

■ Voyant Tools⁶ se trata de una aplicación web (código libre en abierto bajo licencia pública general GNU) para el análisis de textos digitales, por ejemplo, análisis de frecuencias, de concordancias y otros.

■ spaCy⁷ es una librería de software libre con licencia, para procesamiento del lenguaje natural en Python y con muy variadas funcionalidades.

■ Omeka⁸ es una plataforma en abierto para publicar y compartir colecciones digitales enriquecidas con materiales multimedia.

■ NetLine⁹ es una herramienta de libre acceso para representación espacio temporal para la creación de mapas anotados a partir de colecciones OMEKA.

■ Zotero¹⁰ es un gestor de referencias bibliográficas que permite crear, almacenar, organizar, compartir e insertar esas referencias. Es software de código abierto, multiplataforma y con complementos para navegadores de Internet.

6 <https://voyant-tools.org/>.

7 <https://spacy.io/>.

8 <https://omeka.org/>.

9 <https://www.netline.com/>.

10 <https://www.zotero.org/>.

■ Solr¹¹ es un paquete software para indexar grandes colecciones de documentos, de libre acceso, que exige conocimiento informático para su uso. Muy útil para el desarrollo de sistemas de búsqueda y visualización de contenidos.

■ WordPress¹² es una plataforma de gestión de contenido utilizada para la presentación visual. Permite crear páginas web funcionales que con ayuda de sus *plugin* facilitan el desarrollo. De hecho, el 43% de las webs están creadas con WordPress (wordpress.com).

■ Python¹³ es un lenguaje de programación de propósito general muy utilizado en los estudios humanistas y sociales al disponer de numerosas y variadas librerías (programas predefinidos utilizables dentro de otro programa). Necesita un cierto nivel de conocimiento del lenguaje para su uso.

Como antecedentes de digitalización de otros catálogos, como es el caso del trabajo que se presenta, se describen brevemente dos proyectos de la UNED: El proyecto DIMH y el Musivaria.

El proyecto DIMH (HAR2012-31117; 2023 - 15) financiado por el Ministerio de Economía y Competitividad en una convocatoria competitiva. De carácter multidisciplinar por sus participantes y objetivos, incorpora un corpus digital compuesto por una colección de mapas, planos y dibujos (en formato de imagen) y 7792 fichas con texto semiestructurado. Para el análisis de la información se utilizó la técnica de “análisis formal de conceptos” (FCA, de sus siglas en inglés) que detecta relaciones entre contenidos y posteriormente los organiza en base a esas relaciones. En la aplicación de esta técnica al corpus DIMH, se representaron las fichas mediante su texto y los términos en la taxonomía, y a continua-

11 <https://solr.apache.org/>.

12 <https://es.wordpress.org/download/>.

13 <https://www.python.org/>.

ción, para hacer mas eficiente la ejecución, se realizó un refinamiento de la información, y se consiguió un sistema de información mas cercano a los intereses de los usuarios, aunque su desarrollo se finalizó mas tarde que el proyecto de los humanistas, por lo que no pudieron realizarse pruebas de satisfacción de los usuarios (García-Serrano, Castellanos y Cigarrán, 2015; Cámara *et al.*, 2016; García-Serrano y Castellanos, 2016; Castellanos y García-Serrano, 2017). En este caso, hay una publicación con contribuciones de los participantes¹⁴ y no está disponible el software desarrollado para el servicio de la comunidad de historiadores, debido a la obsolescencia del mismo y la ausencia de mantenimiento, una vez que el proyecto competitivo acabó. Este es un problema bien conocido en el entorno de las Humanidades Digitales, aunque esperamos que iniciativas como CLARIAH-ES¹⁵ (desde septiembre de 2023) aporten una solución en España.

El segundo proyecto muy relacionado con este trabajo es “Iconografía musivaria en la Península Ibérica en época romana: investigación y difusión desde el campo de las Humanidades Digitales” (Garrido Ramos, 2019). La palabra musivaria designa a la técnica y arte de elaborar mosaicos. El proyecto es el diseño de una herramienta informática que permite la gestión de la información recopilada sobre los mosaicos romanos encontrados en la península ibérica. El corpus de base usado para este proyecto es el Corpus de Mosaicos de España¹⁶ que aúna almacenamiento de datos gráficos y textuales, modelos 3D, vídeos, presentación de la información mediante geolocalización, utilización de estándares geográficos y vocabularios controlados. A partir de ese corpus se crea inicialmente una base de datos con una muestra representativa de 141 municipios y distritos, con la intención de ir ampliándola progresivamente, con 276 mosaicos, conteniendo informa-

14 <https://biblioteca.juaneloturriano.com/Record/Xebook1-10583>.

15 <https://www.clariah.es/>.

16 <http://www.proyectos.cchs.csic.es/mosaicosromanos/>.

ción textual y geoespacial, así como imágenes, vídeos y modelos 3D para conformar un completo catálogo. En su web de referencia¹⁷ se puede encontrar más información y los contenidos completos del proyecto.

Uno de los objetivos de este trabajo es determinar la idoneidad del uso de tecnologías de almacenamiento, indexación y búsqueda para proyectos en Humanidades Digitales y ver como estas tecnologías son más adecuadas por diversos factores como pueden ser la eficiencia o escalabilidad y en general, mostrar que tecnologías como SOLR o directamente WordPress son mejores que las bases de datos en los proyectos en Humanidades Digitales, en los que se necesite realizar búsquedas complejas para acceso a la información por parte de los usuarios especialistas.

2. PROPUESTA DE DESARROLLO EN DIFERENTES FASES Y SUS DISEÑOS

El objetivo global de este trabajo iba mas allá de encontrar una solución de almacenamiento mas sofisticada que la disponible originalmente, la realizada por el autor D. Cesar Fernández, que ya era de mucha calidad desde el punto de vista de la organización de contenidos. El objetivo era disponer de un sistema de navegación y búsqueda de información, teniendo en cuenta que la colección es de gran tamaño.

El tipo de base de datos más usado hasta la primera década del siglo XXI son las bases de datos relacionales, el proyecto MUSIVARIA HD usa este modelo de base de datos, en concreto usa el gestor de bases de datos MySQL (de libre acceso). Pero cuando los datos no tienen una estructura definida, los conceptos de tablas, campos, registros, de las bases de datos relacionales pierden su razón de ser y se necesitan otros conceptos de almacenamiento basados en la indexación de diversos objetos.

17 <https://musivariahd.com/>.

SOLR (*Searching On Lucene Replication*)¹⁸ es un motor de búsqueda basado en Apache y en la librería LUCENE, escrito en JAVA y de código abierto. Entre sus características principales se destaca la posibilidad de restringir las búsquedas mediante filtrado, la búsqueda por facetas que ofrece sugerencias de filtrado, la clasificación de los resultados de las búsquedas, las búsquedas por sinónimos, o la integración con bases de datos (Granger y Poter, 2014). Solr funciona recorriendo los documentos seleccionados e incorporándonos a un índice, indexado, añadiendo las palabras clave de los documentos. Este índice acepta datos de muchas fuentes, tales como archivos XML, CSV, archivos Word o PDF. SOLR en lugar de buscar en el texto mismo, realiza la búsqueda de la palabra clave del índice, y a continuación indica en qué documentos se encuentra dicha palabra clave. Es una tecnología muy utilizada cuando el corpus puede organizarse sobre la base de una ontología, aunque el campo de la ingeniería ontológica tiene sus propias técnicas y tecnologías, que están fuera de esta exposición.

Finalmente hay que indicar que, al ser imprescindible diseñar un acceso suficientemente útil para los estudiosos del tema, se ha desarrollado una aplicación web para la fase de pruebas de usuario que en este tipo de proyectos deben realizarse para su puesta en explotación asociando un buscador.

En lo que sigue, se describe el trabajo realizado para el proyecto ARQUEOGRIEGOS en sus etapas de gestión y almacenamiento de la documentación y desarrollo de un sistema de búsqueda para la visualización de los documentos. En la primera etapa se seleccionan dos regiones de las 8 disponibles para probar las diferentes aproximaciones.

El sub-corpus de trabajo seleccionado incluye todos los tipos de datos sobre las regiones de Ática y Tracia. Los objetos son documentos de texto (en formato .doc), documentos de hoja de cálculo (en formato .xls), archivos de imagen (en formato .jpg) y

archivos AutoCAD (en formato .dwg). Una vez probada la viabilidad de la propuesta, se incluirán las 6 regiones restantes y así desarrollar una segunda versión, que aquí se presenta.

2.1. PRIMERA APROXIMACIÓN: DISEÑO DE UNA BASE DE DATOS RELACIONAL

Se ha implementado una base de datos relacional en la que insertar la información de los documentos del estudioso, de acuerdo con las siguientes fases.

(1) Diseño de la estructura de la base de datos, atendiendo a la distribución de los documentos entregados por el humanista sobre la región Ática se puede observar una estructura bien definida en cada uno de los documentos de Word que hay por yacimiento (ver tabla 1).

Tabla 1. Estructura de los documentos de Ática

ID	COD	REGION	ARCHIVO	INTRO	ACCESO	HISTORIA	MITOLOGIA	YACIMIENTO	MUSEO
1	0	GEOGRAFIA DEL ATICA	X						
2	0	LOS 11 REYES MITICOS DEL ATICA	X						
3	1	ELEUSIS	X	X	X	X	X	X	X
4	2	MEGARA	X	X	X	X	X	X	X
5	3	HERAION DE PERACHORA	X	X	X		X		
6	4	AIGOSTHENA	X	X	X	X	X		
7	5	ELEUTERAS	X	X	X	X	X		
8	6	TEMPLO DE APOLO ZOSTER	X	X	X	X	X		
9	7	SOUNION	X	X	X	X	X		
10	8	MUSEO DEL LAVRIO	X	X				X	X
11	9	THORIKOS	X	X	X	X	X		
12	10	VRAUNION	X	X	X	X	X	X	X
13	11	ICARION	X	X	X	X	X		
14	12	MARATON	X	X	X	X	X	X	X
15	13	RAMNOUS	X	X	X	X	X		
16	14	ANFIAREIO	X	X	X	X	X		

(2) Formato de los datos a insertar en las relaciones. Los datos aportados por el estudioso necesitan ser adaptados para poder ser insertados en la base de datos relacional. Así, los documentos .doc se convierten a texto plano, o a .xml, con cuidado con el formato del tipo de codificación, para que al insertarlos en la base de datos los textos se mantengan de manera íntegra y no se produzcan sustituciones de unos caracteres por otros. En el caso de

(2) Formato de los datos a insertar en las relaciones. Los datos aportados por el estudioso necesitan ser adaptados para poder ser insertados en la base de datos relacional. Así, los documentos .doc se convierten a texto plano, o a .xml, con cuidado con el formato del tipo de codificación, para que al insertarlos en la base de datos los textos se mantengan de manera íntegra y no se produzcan sustituciones de unos caracteres por otros. En el caso de

las imágenes .jpg de las fotografías, planos y documentos hay que adaptar los nombres de los archivos para que sean accesibles desde enlaces de hipervínculo.

Para habilitar el acceso a la base de datos en primer lugar se necesita una plataforma que la almacene y otra que posteriormente permita acceder a los datos, esto es un servidor local realizado con la aplicación XAMPP (Apache, MySQL y PHP) con el motor de bases de datos MySQL para crear y consultar la base de datos (figura 1) y un servidor remoto, para el que se habilitó un subdominio para la realización de pruebas de rendimiento en un entorno de producción, como se describe en la memoria del trabajo fin de máster de D. Francisco Paños Merino (Paños, 2023).

2.2. DISEÑO DE UN BUSCADOR WEB BASADO EN LA INDEXACIÓN

La configuración del sistema implementado con la herramienta SOLR ha seguido los siguientes pasos.

Paso 1, creación de la colección: Si los documentos que necesita indexar están en formato binario, como Word, Excel, PDF, etc., la herramienta SOLR incluye un controlador (que utiliza Apache Tika¹⁹) para extraer texto para indexarlo en SOLR. Se crea la colección con un esquema personalizado, ARQUEGRIEGOS-ESQ, con un diseño interactivo a través de una interfaz web de SOLR.

Paso 2, selección del modo de esquema. La siguiente decisión que tomar es la granularidad de los documentos que se insertarán en la colección para su indexación y es clave pensar en lo que los usuarios del sistema querrán encontrar y ver en las respuestas a sus búsquedas (Granger y Potter, 2014). En el sub-corpus de trabajo inicial de las regiones de Atica y Tracia, los textos son grandes y se organizan en apartados como introducción, acceso, historia, mitología, museo...

Esta estructuración subyacente proporciona una base para plantear una granularidad media a la hora de definir los campos del esquema. No sería una estrategia buena la de introducir en un único campo todo el texto de un documento (granularidad alta), al igual que tampoco lo sería definir multitud de campos para clasificar por ejemplo los elementos de características de las construcciones de los yacimientos (tipos de columnas, tipos de edificios...), pues los propios textos dan descripciones muy generales más que detalles.

Será importante también en la definición del esquema diferenciar qué campos se indicarán como indexados y cuáles como almacenados, los primeros son aquellos que son determinantes desde la perspectiva del proceso de búsqueda y los segundos aquellos que no, pero que igualmente son útiles para mostrar los resultados de una búsqueda. Para la realización del esquema de esta colección se definen los campos:

- ID numerador general de la colección.
- CODIGO código de la región siguiendo el formato de los documentos entregados por el humanista.
- REGION nombre de la región de la Antigua Grecia.
- ARCHIVO nombre original del archivo entregado por el humanista, suele referirse al nombre del yacimiento.
- INTRO texto de introducción del documento.
- ACCESO información sobre como es el acceso al yacimiento.
- HISTORIA información recopilada sobre la historia del yacimiento.
- MITOLOGIA información sobre los episodios mitológicos relacionados con el yacimiento.
- YACIMIENTO información relacionada con el propio yacimiento como tal.
- MUSEO información sobre el museo, si existiera, en el propio yacimiento.

- FOTOS número de imágenes fotográficas disponibles para el yacimiento.
- PLANOS número de imágenes de planos del yacimiento.
- DOCS numero de otro tipo de imágenes del yacimiento.
- LANG idioma del documento.
- TIMESTAMP fecha de registro del documento.

SOLR creará el esquema con los campos anteriores y decide los parámetros de cada uno de los campos. A continuación, se muestra una tabla con estos parámetros para los campos generados en el esquema SQL.

La asignación automática de los tipos de campo de SOLR, es de cuatro tipos:

- Campo de identificación del documento, en este grupo se ha clasificado a campo ID.
- Campos de gestión del documento, en este grupo se encuadrado los campos CODIGO, REGION, ARCHIVO, TIMESTAMP y LANG que son campos que básicamente contienen información relacionada con el propio archivo y que no contiene un texto interesante para el análisis de este.
- Campos de información del documento, se encuadran los campos INTRO, ACCESO, HISTORIA, MITOLOGIA, YACIMIENTO y MUSEO que como ya sabemos, o intuimos, son los campos en los que hemos incluido la información textual de los documentos.
- Campos numéricos del documento, se encuadran los campos FOTOS, PLANOS y DOCS que contienen datos numéricos sobre el número del elemento de esos tipos (foto, planos y docs) que se entregaron para cada uno de los yacimientos.

En caso de que la asignación automática de campos no fuera considerada la más adecuada, existe la opción de crear los campos manualmente. Si bien en este proyecto la distribución de

la información aportada por el estudioso está muy definida y los formatos en los que se almacena esa información también, la clasificación automática es óptima y ha detectado automáticamente correctamente todos y cada uno de los campos.

A continuación, se ingresaron cada uno de los documentos de texto de la región Ática en el sistema (16) en el sub-corpus de prueba. Además, se insertan los 6 documentos de la región Tracia, que permitirán hacer pruebas sobre el sub-corpus de trabajo. En su momento se habilitó un acceso remoto en la dirección privada, pero solo de prueba, no se podía poner en explotación esta versión.

2.3. EJEMPLOS DE BÚSQUEDAS PARA ACCESO A LA INFORMACIÓN

Como el sub-corpus de trabajo no es muy extenso, se hace una prueba previa con la herramienta VOYANT (ver figura 2), y manualmente estudiar algunos aspectos de los documentos, aspectos relacionados con las palabras de los textos. Así se comprueba con la lectura

Contextos Líneas de búrbuja Colocaciones ?				
Documento	Izquierda	Términos	Derecha	
11. ICARION	siempre se levantaba cerca del	teatro	. Otra posibilidad es que se	
13. RAMNOUS	por su forma tan simple,	teatro	(36) de Ramnous. Como cávea	
13. RAMNOUS	que se ha conservado del	teatro	. El teatro y el espacio	
13. RAMNOUS	ha conservado del teatro. El	teatro	y el espacio de alrededor	
13. RAMNOUS	Alrededor del gimnasio y del	teatro	se encuentran las ruinas de	
14. ANFIAREIO	ocupan hoy los pedestales. El	teatro	(c) se encuentra detrás y	
14. ANFIAREIO	parte de su escena. Al	teatro	se entraba a través de	
14. ANFIAREIO	reconstruida columnata del proscenio del	teatro	con 8 semicolumnas dóricas entre	
14. ANFIAREIO	lo que queda del llamado	teatro	del altar. En su forma	
14. ANFIAREIO	V o IV a.C.) el	teatro	era más grande y su	
14. ANFIAREIO	santuario, se decidió demantelar el	teatro	y utilizar sus piedras como	
9. THORIKÓS	de la torre cuadrada, del	teatro	y del ágora dórica. La	
9. THORIKÓS	Escuela Americana de Estudios Clásicos (	teatro	) y en 1893 por la	
9. THORIKÓS	a niños. Al norte del	teatro	se encontraron tumbas de los	
9. THORIKÓS	encuentra inmediatamente al sur del	teatro	(13, 16) se encontraba en	
9. THORIKÓS	del siglo IV a.C. El	teatro	(7 y 9 a 13	
9. THORIKÓS	sur de la orquesta del	teatro	se construyese un cementerio contemporáneo	
9. THORIKÓS	talleres y galerías alrededor del	teatro	muestran la densidad de la	
9. THORIKÓS	lado del graderío oeste del	teatro	está la llamada Casa 2	
9. THORIKÓS	un patio. La separa del	teatro	un estrecho camino que sube	
9. THORIKÓS	la ladera. Al sur del	teatro	y al oeste del cementerio	
14. ANFIAREIO	para el estudio de otros	teatros	puesto que se conserva una	

Figura 2. Capturas de la web de VOYANT TOOLS

Diseño del acceso a un catálogo sobre la Antigua Grecia

de los documentos, que la palabra “teatro” aparece en los yacimientos de THORIKOS, ICARION, RAMNOUS, ANFIAREIO en Ática y ABDERA y MARONEIA de Tracia. Se comprueba que salvo ICARION (uso genérico del término), todos tienen restos de un teatro.

Si buscamos la palabra “teatro” en la base de datos implementada, se obtiene la respuesta de la figura 3.

Mostrando filas 0 - 5 (total de 6, La consulta tardó 0,0023 segundos)

SELECT * FROM "arqueologia_vsl"."yacimientos" WHERE (CONVERT('id' USING utf8) LIKE 'teatro' OR CONVERT('codigo' USING utf8) LIKE 'teatro' OR CONVERT('region' USING utf8) LIKE 'teatro' OR CONVERT('archivo' USING utf8) LIKE 'teatro' OR CONVERT('intro' USING utf8) LIKE 'teatro' OR CONVERT('acceso' USING utf8) LIKE 'teatro' OR CONVERT('historia' USING utf8) LIKE 'teatro' OR CONVERT('mitologia' USING utf8) LIKE 'teatro' OR CONVERT('yacimiento' USING utf8) LIKE 'teatro' OR CONVERT('museo' USING utf8) LIKE 'teatro')

Perifoneo [Editar en línea] [Editar] [Explicar SQL] [Copiar código PHP] [Actualizar]

☐ Mostrar todo | Número de filas: 25 | Filtrar filas: [Buscar en esta tabla] | Ordenar según la clave: Ninguna

Opciones extra:

	id	codigo	region	archivo	intro	acceso	historia	mitologia	yacimiento	museo	fotos	planos	docs
☐	11	9	ATICA	THORIKOS	Cefaleo, tipo de Delos, se casó en Thouria del Á.	Pasado la localidad de Laurio, seguimos la carretera.	Thourikos es uno de los más antiguos asentamientos.	1. Thourikos era la patria del rey Cefaleo, hijo de	Vigentes del siglo XVII y XIX según descripción.		29	4	2
☐	13	11	ATICA	ICARION	La penia Mera, auxiliando por el difunto ciano, mi.	El acceso al yacimiento se hace a través del entrec.	El lugar en donde está el yacimiento pertenece a:	Atique Eneo fue el primer mortal al que Dioniso di.	A finales del siglo XIX, en Rapazouza, un vaporio		13	2	1
☐	15	13	ATICA	RAMNOUS	pero se pasa hacia el interior hay un santuario.	El acceso al yacimiento puede hacerse a través del.	El nombre del lugar procede de la palabra "ramnos".	1. Algunos dicen que cuando Zeus se enamoró de Ním.	La primera excavación científica que tuvo lugar en.		54	15	1
☐	16	14	ATICA	ANFIAREIO	El templo de Anfiaros se trasladó aquí desde la I.	El acceso al yacimiento se hace directamente desde.	El Anfiteatro de Orpéus se fundó a finales del sig.	1. Tras haber sido tragado por la tierra durante l.	Gracias a la florecida arqueología empezaron se e.		41	9	1
☐	19	15	TRACIA	ABDERA	?????ABDERA - en su viaje de vuelta, según l.	Dece de Alexandropolis cogemos la autopista hacia Xa.	La ciudad de Abdera, con dos puertos naturales, se.	1. Eneidos ordenó a Heracles, para su octavo lab.	La ciudad fue identificada por el arqueólogo aca.	Los hallazgos de las excavaciones de la ciudad y l.	51	13	2
☐	20	16	TRACIA	MARONEIA	????????MARONEIA - "Llevaba vino negro que añeja.	El acceso se realiza desde Komotini. Una carretera.	El teatro es referido en la Odissea como la ciudad y l.		Las excavaciones empezaron a través de la década		12	1	0

☐ Seleccionar todo | Para los elementos que estén marcados: ☐ Editar ☐ Copiar ☐ Borrar ☐ Exportar

☐ Mostrar todo | Número de filas: 25 | Filtrar filas: [Buscar en esta tabla] | Ordenar según la clave: Ninguna

Operaciones sobre los resultados de la consulta

☐ Imprimir ☐ Copiar al portapapeles ☐ Exportar ☐ Mostrar gráficos ☐ Crear vista

☐ Guardar esta consulta en favoritos

Etiqueta: ☐ Permitir que todo usuario pueda acceder a este favorito

☐ Guardar esta consulta en favoritos

Figura 3. Consulta del término TEATRO en la base de datos

Se producen también 6 coincidencias, es decir el termino aparece en 6 yacimientos, a partir de la consulta (que exige conocimiento no básico del lenguaje SQL):

```
SELECT COUNT(*) AS num_coincidencias FROM yacimientos WHERE (CONVERT('id' USING utf8) LIKE '%teatro%' OR CONVERT('codigo' USING utf8) LIKE '%teatro%' OR CONVERT('region' USING utf8) LIKE '%teatro%' OR CONVERT('archivo' USING utf8) LIKE '%teatro%' OR CONVERT('intro' USING utf8) LIKE '%teatro%' OR CONVERT('acceso' USING utf8) LIKE '%teatro%' OR CONVERT('historia' USING utf8) LIKE '%teatro%' OR CONVERT('mitologia' USING utf8) LIKE '%teatro%' OR CONVERT('yacimiento' USING utf8) LIKE '%teatro%' OR CONVERT('museo' USING utf8) LIKE '%teatro%')
```

```
LIKE '%teatro%' OR CONVERT(`fotos` USING utf8)
LIKE '%teatro%' OR CONVERT(`planos` USING utf8)
LIKE '%teatro%' OR CONVERT(`docs` USING utf8) LIKE
'%teatro%');
```

El usuario puede leer el campo por completo, pero se accede a él de manera mas engorrosa que en el caso anterior, que lo mostraba en pantalla, por lo que no es una solución aceptable para usuarios no informáticos. Finalmente, para el sub-corpus disponible en un servidor Solr se configura la consulta QUERY con sus parámetros como se muestra en la Figura 4 (esta aplicación ya no está disponible en aquella primera versión, pero sí disponible en la versión actual según se describe en la siguiente sección de este documento).

Para mejorar la visualización anterior, se transformó la respuesta Solr a formato JSON y se implementó una página web para este proyecto con el administrador de contenidos WordPress, con una aplicación, WPSOLR, que lo comunicaba con SOLR (solo para Windows). La visualización de la pregunta se muestra en la figura 5 (Paños, 2023).

La mejora de visualización de los resultados con respecto a la de una base de datos es notoria, pues la inspección es directa

[illegible]

Figura 4. Búsqueda y resultados en SOLR

ARQUEOGRIEGOS

INICIO SOLR-SQL BUSCADOR YACIMIENTOS CONTACTO

BUSCADOR

teatro

Sort by
More relevant

Filters
☐ All results

By Displaydate d:
☐ 2023-06-31T10:35:09Z (1)
☐ 2023-06-31T10:37:19Z (1)
☐ 2023-06-31T10:46:17Z (1)
☐ 2023-06-31T10:46:00Z (1)
☐ 2023-06-31T10:52:02Z (1)
☐ 2023-06-31T10:53:00Z (1)

By Displaymodified d:
☐ 2023-06-31T10:13:00Z (1)
☐ 2023-06-31T10:37:19Z (1)
☐ 2023-06-31T10:46:17Z (1)
☐ 2023-06-31T10:48:03Z (1)
☐ 2023-06-31T10:52:02Z (1)
☐ 2023-06-31T10:53:09Z (1)

By Meta type s
☐ post_type (6)

Showing 1 to 6 results out of 6

THORIKOS

 Ilustración el hecho de que justamente al sur de la orquesta del teatro se construyese un cementerio contemporáneo con él. Casas, talleres y galerías alrededor del teatro muestran la densidad de la ocupación y la variedad de ocupaciones de los antiguos habitantes de Thorikos. Al lado del graderío oeste del teatro está la llamada Casa 2 (7) con una habitación de banquetes y un patio....
 By fpanos3, on 06/31/2023

MARONEIA

 Se accede a él por el mismo camino que nos lleva al teatro, encontrándose antes de llegar a éste. La gran casa helénica se encuentra a la izquierda de la carretera....
 By fpanos3, on 06/31/2023

ANF. AREIO

 Es todo lo que queda del llamado teatro del altar. En su forma primitiva (siglo V o V a.C.) el teatro era más grande y su función era la de ofrecer asiento a los que observaban los sacrificios que realizaban los sacerdotes sobre el altar....
 By fpanos3, on 06/31/2023

RAMNOUS

 Al norte de ella había un gimnasio (35) de unos 1.200 m2 que tenía unos propileos monumentales en su lado sur con un dintel de una sola pieza. Al norte del gimnasio se

Figura 5. Primer buscador y vista de los resultados para el término TEATRO

(con o sin página web). También se puede asociar a la base de datos una página web para mostrar los resultados de las búsquedas SQL, pero en Solr es posible además implementar que se muestren todas las apariciones del término en cada documento. Finalmente hay que indicar que cualquier comparación relacionada con los tiempos de acceso o ejecución de consultas en ambas tecnologías también beneficia la segunda aproximación.

En este momento ya se podía decidir la aproximación mas adecuada para trabajar con un gran número de archivos textuales, los correspondientes a las 8 regiones, e incluir las fotos y los planos en la actual aplicación de acceso a la información del catálogo aportado por el estudioso. En la siguiente etapa, se incluyen las imágenes de la colección y se realizan las mejoras que se indican en la siguiente sección.

3. APLICACIÓN PARA EL ACCESO A LA INFORMACIÓN DEL CATÁLOGO

La segunda versión de la aplicación web y del buscador se ha realizado para incluir las imágenes del corpus multimedia. En cuanto a las herramientas adicionales necesarias, se han empleado, por ejemplo, *Notepad++* para una edición avanzada y eficiente sobre archivos de texto y de otros formatos, *Visual Studio Code* para editar el código en Python, un lenguaje de programación flexible, versátil y con gran cantidad de bibliotecas y módulos de ayuda para crear los scripts (programas) necesarios para la implementación de las funcionalidades del proyecto o también, SOLR y WordPress ya introducidas en este trabajo.

La preparación del corpus también ha variado. Primeramente, se transforman los ficheros en formato .doc a texto plano (formato .txt), creando ficheros con un nombre que se corresponde exactamente con la ruta de su localización en la colección, esto es el nombre de cada una de las carpetas que le contienen. Por ejemplo, “Atenas” para el fichero que contiene información sobre “EL MUSEO ARQUEOLÓGICO DE EL PIREO”. A continuación, se genera un único fichero en formato XML (lenguaje de marcado) tanto con el contenido de cada uno de los ficheros en formato .doc como una etiqueta que indica su contenido.

Un poco mas en detalle. En el fichero XML, habrá información de cada fichero .doc entre dos etiquetas: <doc></doc>. A su vez en cada fichero habrá marcas o etiquetas para indicar el tipo de contenido (semántica). Por ejemplo, si observamos el fichero **ELEUSIS.doc** hay unas secciones, que, observando otros documentos de otros yacimientos, vemos que se repiten en todos ellos y que son exactamente: INTRO, contiene el nombre del yacimiento y una breve cita; ACCESO, información sobre cómo se accede al yacimiento; HISTORIA, información sobre la historia del yacimiento; MITOLOGIA, información sobre los episodios mitológicos relacionados con el yacimiento; YACIMIENTO, información técnica y descriptiva de los elementos arquitectónicos y otros del yacimiento; MUSEO, en algunos yacimientos además del propio yacimiento hay un museo que se describe en esta sección.

A continuación, se ha procedido a preparar el corpus de fotos (imágenes) y de planos de las 8 regiones. Para integrar 4746 imágenes que se corresponden a las distintas regiones del catálogo de arqueogriegos, hay que descargarlas y el primer paso, es identificar la estructura de ficheros del catálogo, como en el caso de los ficheros de texto. El segundo paso es decidir la estructura de las imágenes dentro del XML, y se ha optado por la siguiente estructura:

- ID: El identificador de cada imagen insertada en el XML.
- ID_ENTRADA: El identificador que se indica de la entrada a la que va asociada la imagen.
- TIPO: Si la imagen consiste en que es una Foto o un Plano.
- REGIÓN: Región a la que pertenece la imagen.
- ARCHIVO: Nombre del Yacimiento o Museo al que pertenece.
- NOMBRE: Nombre de la imagen.
- RUTA: Ruta completa de la imagen.

Una vez ejecutemos el script, en la figura 9 se muestra un ejemplo de lo que sería el resultado final del XML.

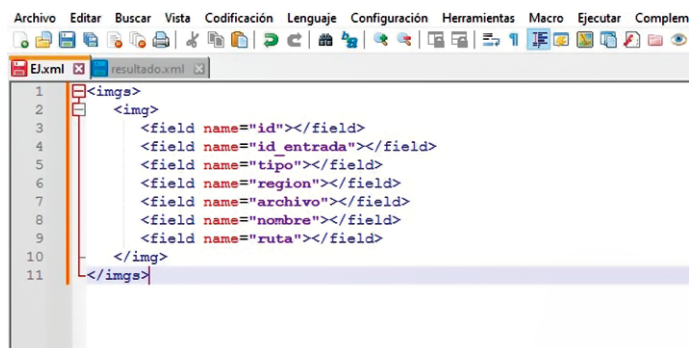


Figura 8. Estructura en etiquetas XML del contenido de los ficheros de imágenes

```

1 <?xml version='1.0' encoding='utf-8'?>
2 <doc>
3   <field name='id'>1</field>
4   <field name='id_entrada'>Diapo 1.1</field>
5   <field name='tipo'>PLANO</field>
6   <field name='region'>TRACIA</field>
7   <field name='archivo'>ABDERA</field>
8   <field name='nombre'>Planta general del yacimiento de Abdera</field>
9   <field name='ruta'>https://alball.lsi.uned.es/arqueogriegos/wp-content/uploads/2024/04/Diapo-1.1-Planta-general-del-yacimiento-de-Abdera.jpg</field>
10 </doc>
11 <doc>
12   <field name='id'>2</field>
13   <field name='id_entrada'>Diapo 1.10</field>
14   <field name='tipo'>FOTO</field>
15   <field name='region'>TRACIA</field>
16   <field name='archivo'>ABDERA</field>
17   <field name='nombre'>El Edificio Maxé8n</field>
18   <field name='ruta'>https://alball.lsi.uned.es/arqueogriegos/wp-content/uploads/2024/04/Diapo-1.10-El-Edificio-Maxé8n.jpg</field>
19 </doc>
20 <doc>
21   <field name='id'>3</field>
22   <field name='id_entrada'>Diapo 1.11</field>
23   <field name='tipo'>FOTO</field>
24   <field name='region'>TRACIA</field>
25   <field name='archivo'>ABDERA</field>
26   <field name='nombre'>Restos de la basílica cementerio</field>
27   <field name='ruta'>https://alball.lsi.uned.es/arqueogriegos/wp-content/uploads/2024/04/Diapo-1.11-Restos-de-la-basílica-cementerio.jpg</field>
28 </doc>
29 <doc>
30   <field name='id'>4</field>
31   <field name='id_entrada'>Diapo 1.12</field>
32   <field name='tipo'>FOTO</field>
33   <field name='region'>TRACIA</field>
34   <field name='archivo'>ABDERA</field>
35   <field name='nombre'>Termas</field>
36   <field name='ruta'>https://alball.lsi.uned.es/arqueogriegos/wp-content/uploads/2024/04/Diapo-1.12-Termas.jpg</field>
37 </doc>
38 <doc>
39   <field name='id'>5</field>
40   <field name='id_entrada'>Diapo 1.13</field>
41   <field name='tipo'>FOTO</field>
42   <field name='region'>TRACIA</field>
43   <field name='archivo'>ABDERA</field>
44   <field name='nombre'>Monumento funerario</field>
45   <field name='ruta'>https://alball.lsi.uned.es/arqueogriegos/wp-content/uploads/2024/04/Diapo-1.13-Monumento-funerario.jpg</field>
46 </doc>
47 <doc>
48   <field name='id'>6</field>
49   <field name='id_entrada'>Diapo 1.14</field>
50   <field name='tipo'>FOTO</field>
51   <field name='region'>TRACIA</field>
52   <field name='archivo'>ABDERA</field>
53   <field name='nombre'>Monumento funerario</field>
54   <field name='ruta'>https://alball.lsi.uned.es/arqueogriegos/wp-content/uploads/2024/04/Diapo-1.14-Monumento-funerario.jpg</field>
55 </doc>
56 </doc>

```

Figura 9. Vista parcial del fichero XML resultado del proceso de los ficheros con imágenes

En el caso de los planos AutoCAD, los ficheros (formato .dwg) se transforman a fichero imagen (en formato .jpg). Como los archivos de AutoCAD, tienen licencia de pago y su manipulación requiere de manejar esta aplicación, se utiliza una herramienta que permite convertir archivos .dwg a .dxf que es de uso libre, se llama *ODA File Converter*²⁰ y a continuación se utiliza un script Python propio que permite la conversión de múltiples archivos de formato .dxf (LibreCAD) a .jpeg. Ahora, teniendo todas las imágenes en formato adecuado se aplica un script que convierte esas imágenes a un único XML bien formateado. En este caso la gestión del script se va a realizar desde la nube ya que *Google Colab* no permite el recorrer carpetas en un entorno local debido a las restricciones de permisos. Por ello se usará *Google Drive* como herramienta adicional.

Una vez indexado el fichero único en Solr e ingresado en las páginas web realizadas con WordPress en un servidor de la

20 https://www.opendesign.com/oda_online_converter.

UNED (usando el plugin llamado *WP All Import*) puede accederse a la información²¹ como se muestra en la figura 10 y convertirlas en entradas de WordPress.



Figura 10. Captura de pantalla de la aplicación (2024)

En el menú de la aplicación web el usuario observa las distintas categorías de las distintas regiones en las que se dividen todas las 163 entradas de la colección y sus tipos primarios. Para realizar una búsqueda no navegando, en el apartado del ícono de la lupa en la parte superior a la derecha se encuentra el cuadro de búsqueda que se abrirá para buscar en el catálogo de Arqueogriegos.

CONCLUSIONES

En este artículo, se presenta un problema, explotar un corpus multimedia digital, y dos posibles soluciones tecnológicas que ayudan a plantear el desarrollo final de la solución tecnológica para su explotación.

21 <https://arqueogriegos.iatext.ulpgc.es/>.

Se muestra cómo las bases de datos se pueden utilizar, pero que no son la mejor tecnología de almacenamiento para alojar corpus con muchos documentos textuales y, en cambio, sí lo son las bases de datos NoSQL y herramientas tecnológicas como SOLR, mejorándose tanto la eficiencia en términos de velocidad de búsqueda y respuesta como la facilidad de acceso a la información por usuarios no informáticos.

Se ha trabajado no solo en la aplicación web (en sus varias versiones) para acceder a la información sino también en aspectos relacionados con su disponibilidad a través de un buscador web.

El trabajo que se va a realizar a continuación es la prueba con usuarios para evaluar la adecuación de la aplicación al catálogo y las necesidades de los estudiosos del tema. Además, ya se está gestionando la disposición como recurso en abierto tanto de la colección multimedia como de las diferentes aplicaciones.

Agradecimientos

Los autores desean agradecer la financiación parcial de este trabajo por los proyectos CLARA-HD (PID2020-116001RB-C32)²² y el proyecto GRE-SEL-UNED (PID2023-151280OB-C22).

REFERENCIAS BIBLIOGRÁFICAS

CÁMARA MUÑOZ, Alicia, Alfonso MUÑOZ COSME, José CALVO LÓPEZ, Javier ORTEGA VIDAL, Juan Miguel MUÑOZ CORBALÁN, Fernando COBOS GUERRA, Carlos José HERNANDO SÁNCHEZ, Pablo de la FUENTE DE PABLO, Consuelo GÓMEZ LÓPEZ, Antonio BRAVO NIETO, Sergio RAMÍREZ GONZÁLEZ, Maurizio VESCO, Annalisa DAMERI, Isabelle WARMOES, Emilie D. ORGEIX, Ana GARCÍA-SERRANO, Ángel CASTELLANOS, Jesús LÓPEZ DÍAZ, Margarita VÁZQUEZ MANASSERO, *El dibujante ingeniero al servicio de la monarquía hispánica: siglos XVI-XVIII*, ISBN: 9788494269561, Fundación Juanelo Turriano, <https://biblioteca.juaneloturriano.com/Record/Xebook1-10583>, 2016.

22 <https://clara-nlp.uned.es/home/dh/>.

- CASTELLANOS GONZÁLEZ, Ángel y Ana GARCÍA-SERRANO, *Representación y Organización de documentos digitales: Detalles y Práctica sobre la Ontología DIMH*, Revista de Humanidades Digitales, v. 1, ISSN 2531-1786, <https://doi.org/10.5944/rhd.vol.1.2017.17155>, (2017), pp. 314-344.
- DEVLIN, Jacob, Ming-Wei CHANG, Kenton LEE y Kristina TOUTANOVA, *Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding*, <http://arxiv.org/abs/1810.04805>, (2019).
- GARCÍA-SERRANO, Ana, ÁNGEL CASTELLANOS y Juan CIGARRÁN, “De la preservación digital al acceso semántico de documentos históricos” en *II Congreso de Humanidades Digitales Hispánicas – Innovación, globalización e impacto*. 05/10/2015 Madrid. <https://humanidadesdigitaleshispnicas.es/wp-content/uploads/2020/02/Humanidades-Digitales-Hisp%C3%AAnicas.-Innovaci%C3%B3n-Globalizaci%C3%B3n-e-Impacto.pdf>, (2015).
- GARCÍA-SERRANO, Ana y Ángel CASTELLANOS, “Conceptualización, acceso y visibilidad de la información en el proyecto DIMH”, en Muñoz, A. C. *El dibujante ingeniero al servicio de la monarquía hispánica: Siglos XVI-XVIII*. Fundación Juanelo Turriano. Alicia CÁMARA y Bernardo REVUELTA (Eds). <http://www.juaneloturriano.com/coleccion-digital/lecciones-juanelo-turriano-de-historia-de-la-ingenieria>, (2016), pp. 379-400.
- GARCÍA-SERRANO, Ana y Antonio MENTA GARUZ, “Orientaciones y evaluación de técnicas en Humanidades Digitales: de la estadística al *deep-learning*”, en *HDH 2021–Scire Vías. Humanidades Digitales y Conocimiento*. 4-8 octubre, España, (2021).
- GARCÍA-SERRANO, Ana y Antonio MENTA GARUZ, *La inteligencia artificial en las Humanidades Digitales: dos experiencias con corpus digitales*. Revista de Humanidades Digitales, V7, <https://doi.org/10.5944/rhd.vol.7.2022.30928>, (2022), pp. 19-39.
- GARCÍA-SERRANO, Ana, Francisco PAÑOS MERINO y César FERNÁNDEZ SÁNCHEZ, “Una mirada digital a los datos: el trabajo de un estudioso de la Antigua Grecia” en VI congreso de la Sociedad Internacional de Humanidades Digitales Hispánicas – HDH 2023 “Encuentros y transformaciones” Logroño, 18 y 19 octubre, (2023).
- GARRIDO RAMOS, Beatriz, *Iconografía musivaria en la Península Ibérica en época romana: investigación y difusión desde el campo de las Hu-*

- manidades Digitales*. Repositorio institucional de la Universidad Nacional de Educación a Distancia (España). <http://e-spacio.uned.es/fez/view/tesisuned:ED-Pg-HHAT-Bgarrido>, (2019).
- GRANGER, Ter y Timothy POTTER, *Solr in action*. Manning Publications; ISBN 16172 91021, (2014).
- MENTA, Antonio, Eva SÁNCHEZ-SALIDO y Ana GARCÍA-SERRANO, “Transcripción de periódicos históricos: aproximación CLARA-HD” en *Proc. of the Annual Conf. of the SEPLN*. A Coruña, Spain, septb 21-23. CEUR-WS V 3224, ISSN: 1613-0073 <https://ceur-ws.org/Vol-3224/> (2022), pp. 70-74.
- PAÑOS MERINO, Francisco (2023) *Arqueogriegos*. Memoria del Trabajo Fin de Máster dirigido por Dra. Ana García Serrano. Disponible repositorio institucional de la UNED <http://e-spacio.uned.es/fez/view/bibliuned:master-ETSInformatica-II-Fpanos>.
- SÁNCHEZ-SALIDO, Eva, Antonio MENTA y Ana GARCÍA-SERRANO, *Seeking Information in Spanish Historical Newspapers: The Case of Diario de Madrid (18th and 19th Centuries)* Digital Humanities Quarterly. V17 N 4. DHQ journal. ISSN: 1938-4122. <http://digitalhumanities.org/dhq/vol/17/4/000735/000735.html> (2023).

SISTEMA DE ETIQUETADO AUTOMÁTICO DE CATÁLOGOS DE EXPOSICIONES PARA LA EXTRACCIÓN DE CARACTERÍSTICAS Y EL RECONOCIMIENTO DE ENTIDADES NOMBRADAS: PRIMERA APROXIMACIÓN

M. Luisa Díez Platas, María Ortiz Tello,
Manuela García Lirio, Nuria Rodríguez Ortega

Resumen

Este artículo presenta las primeras fases del desarrollo de un sistema de etiquetado automático para catálogos de exposiciones artísticas, con el fin de extraer información estructurada e integrarla en la ontología OntoExhibit y su correspondiente grafo de conocimiento.

Para la implementación del sistema se utilizan técnicas de procesamiento de lenguaje natural y aprendizaje profundo, mediante el diseño de un modelo LSTM bidireccional, para el reconocimiento de entidades nombradas y elementos específicos propios de la estructura de la información presente en los catálogos objeto de estudio. Los resultados

M.^a Luisa Díez Platas (UNIR)
marialuisa.diez@unir.net
ORCID: 0000-0002-9416-3147

María Ortiz Tello (UMA)
mortiztello@uma.es
ORCID: 0000-0001-8082-7064

Manuela García Lirio (UGR)
magali@ugr.es
ORCID: 0000-0001-7734-8142

Nuria Rodríguez Ortega (UMA)
nro@uma.es
ORCID: 0000-0003-3460-397X

preliminares del modelo, aunque prometedores, evidencian la necesidad de ajustar métricas como la puntuación F1 y optimizar el reconocimiento de entidades anidadas. Este enfoque representa un avance significativo en la automatización del análisis de grandes corpus textuales en el dominio del arte.

Palabras clave: Etiquetado automático, catálogos de exposiciones artísticas, procesamiento de lenguaje natural (PLN), aprendizaje profundo, reconocimiento de entidades nombradas (NER), Ontología, grafo de conocimiento.

Abstract

This paper presents the initial phases of the development of an automatic tagging system for art exhibition catalogs, aimed at extracting structured information and integrating it into the OntoExhibit ontology and its corresponding knowledge graph.

For the system's implementation, natural language processing techniques and deep learning are applied through the design of a bidirectional LSTM model for named entity recognition and specific elements inherent to the information structure present in the studied catalogs. While the preliminary model results are promising, they underscore the need to refine metrics such as the F1 score and optimize the recognition of nested entities. This approach represents a significant advancement in automating the analysis of large textual corpora in the art domain.

Keywords: Automatic labeling, art exhibition catalogs, natural language processing (NLP), deep learning, named entity recognition (NER), ontology, knowledge graph.

1. INTRODUCCIÓN

Esta propuesta se realiza en el contexto de tres proyectos de investigación (UMA20-FEDERJA-126, PY20_00508 y PID2021-125037NB-I00) [1] que se están desarrollando de manera simultánea y que cuentan entre sus objetivos el desarrollo de mecanismos que permitan la extracción automática de información estructurada de catálogos de exposiciones artísticas. Este objetivo se incardi-

na, a su vez, en un objetivo de mayor alcance: nutrir la ontología OntoExhibit [3], desarrollada *ad hoc* para el modelado del dominio de las exposiciones artísticas y generar el correspondiente grafo de conocimiento.

Se debe tener en cuenta que lo habitual es la creación de un catálogo, resultado de la exposición que recopila la información relevante de esta y la lista de las obras expuestas. Tanto la cantidad de catálogos con los que se cuenta como la información relevante que contienen hace casi imposible plantearse la extracción de la información de cada uno de ellos, de forma estructurada, mediante un proceso meramente manual.

Por tanto, se ha tomado la decisión de que, para conseguir el objetivo de extraer la cantidad ingente de información de forma estructurada de manera que se facilite el proceso de integración de ésta en el grafo de conocimiento de la ontología mencionada, se debían aplicar técnicas de procesamiento de lenguaje natural (PLN). Estas cuentan con un gran potencial para llevar a cabo procesos de extracción de información de grandes corpus textuales. Concretamente, la extracción de información ha consistido en la obtención de información estructurada de textos escritos en lenguaje natural y en formatos no legibles para una máquina.

La información que se puede extraer de los catálogos de arte es de muy distinta índole y, por tanto, se podría decir que la implementación de los procedimientos de PLN ha dado lugar a la integración de distintas formas de reconocimiento. Así, se ha tenido en cuenta el reconocimiento de entidades nombradas (NER) [2], pero también mecanismos específicos para la extracción de información propia del género catalogar, como la asociada a la identificación y descripción de las obras artísticas que comprende cada catálogo.

Para la extracción automática de la información, se necesita el etiquetado de los textos, es decir, la delimitación de las distintas entidades o información que se identifican. Dado que, como ya se ha mencionado, un procedimiento manual ha quedado descartado, en este artículo se presenta un sistema de etiqueta-

do automático basado en técnicas de IA y, más concretamente, en modelos de aprendizaje automático.

Este artículo es una primera aproximación para la creación del sistema de etiquetado.

2. CONSIDERACIONES SOBRE PLN

Previo al comienzo de la preparación de los datos se realizó un análisis exhaustivo de los distintos catálogos con los que se contaba. Se analizó la estructura completa, la parte de texto ensayístico y la lista de obras que constituyen lo que se conoce propiamente como el catálogo y la información la estructura de cada uno de los elementos de información de las obras.

Se observó que las estructuras usadas en los catálogos no eran las mismas para todos ni tampoco la información que se incluía.

De acuerdo con esto, y dada la posibilidad de usar distintas técnicas de PLN, se hizo un análisis de cada una de ellas aplicando criterios de acuerdo con los textos que se iban a procesar y la conveniencia del uso o no de la técnica considerada.

1. Se realizaron pruebas para el etiquetado automático mediante técnicas de PLN tradicionales a fin de crear un sistema basado en expresiones regulares y procesamiento semántico contextual mediante el uso de expresiones regulares para encontrar las entidades con un complemento semántico basado en el cambio de estado para identificar contextos como se propone en (Díez Platas *et al.*, 2021). La eficacia de estos sistemas es alta, especialmente en textos en los que se identifiquen características muy específicas, se cuente con pocos textos y con pocas regularidades.

Los resultados fueron buenos dada la posibilidad de definir reglas específicas para cada uno de los tipos de catálogos. Sin embargo, la complejidad de las especificaciones y la diversidad de los contextos es grande.

2. Se descartó un sistema de aprendizaje automático (*machine learning*) debido a la necesidad de información procedente del contexto de las frases y la secuencia de etiqueta de estas, ya que las frases pueden ser muy largas en los textos y estos sistemas no tienen capacidad para capturarla.

3. Se analizó un modelo de aprendizaje profundo (*deep learning*). Concretamente, un modelo de memoria de tiempo largo y corto (LSTM) bidireccional. Una red LSTM bidireccional combina dos LSTM que funcionan cada una en un sentido (izquierda-derecha, derecha-izquierda), proporcionando una cantidad infinita de contexto para la red que proviene de ambos lados de una palabra y elimina el problema del contexto limitado que se encuentra en otros modelos (Chiu & Nichols, 2016), vease Figura 1. Esta técnica mejora el rendimiento del problema en cuanto a

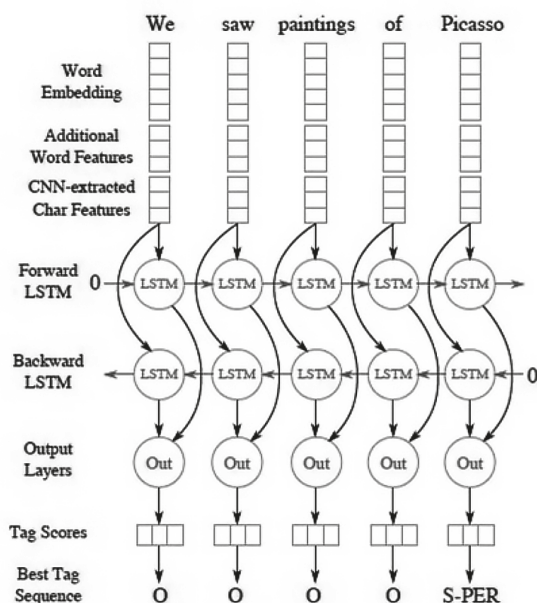


Figura 1. LSTM bidireccional. Fuente:(Chiu & Nichols, 2016)

la clasificación o etiquetado de las frases, ya que captura el contexto y el aprendizaje es suficientemente rápido.

Por tanto, se decide por la tercera opción se la que se muestra su estructura principal en la figura 1, considerando la información de entrada y las etiquetas de salida.

3. METODOLOGÍA Y DESARROLLO

La propuesta que se presenta en este artículo pretende llevar a cabo las fases presentadas en la Figura 2, algunas de las cuales se encuentran todavía en proceso de desarrollo.

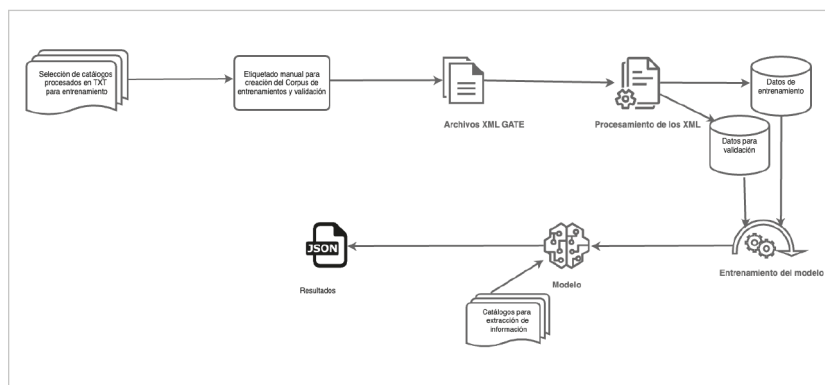


Figura 2. Procesos de entrenamiento y validación del modelo, ejecución del modelo y resultados obtenidos. Elaboración propia

Como se puede ver se realizan unas primeras fases de preparado de textos para el entrenamiento del modelo y una vez que se valida se procesa el conjunto completo de catálogos y se obtiene un archivo en formato JSON que servirá para la obtención futura de los triples que se integrarán en el grafo de conocimiento.

En este artículo se presenta en profundidad la preparación de los datos de entrada para el entrenamiento del modelo, el modelo elegido y los primeros resultados de validación.

3.1. PREPARACIÓN DE LOS TEXTOS Y CREACIÓN DEL CORPUS DE CATÁLOGOS PARA ENTRENAMIENTO Y VALIDACIÓN

En primer lugar, se recopiló un conjunto de catálogos procedentes de distintos organismos y fundaciones, casi todas ellos de carácter actual, obtenidos, especialmente, de galerías y centros de arte contemporáneo. El motivo de esta decisión es que, en el contexto de nuestra investigación, estos circunscriben un panorama actualizado del ecosistema expositivo español. De igual modo, la selección de los catálogos estuvo guiada por criterios de diversidad, a fin de configurar una muestra lo más representativa posible de las distintas características de edición y forma de maquetado de la información. Se tuvo en cuenta catálogos de exposiciones individuales y colectivas, de instituciones públicas y privadas, de índole heterogénea, con estructuras tradicionales e innovadoras. Como requisito previo, estos catálogos debían ser de acceso abierto y encontrarse digitalizados en línea. Además, se llevó a cabo una clasificación de acuerdo con el estilo de los distintos catálogos para futuras funcionalidades que están previstas en el sistema, como la clasificación de los catálogos a partir de los patrones de características estructurales diversas: organizados por secciones dedicadas a colecciones de piezas; por secciones dedicadas a artistas, en las que las piezas son ítems dentro de ellas, es decir, donde el artista creador es el verdadero protagonista del catálogo y no la pieza *per se*; por fichas descriptivas de las obras, donde el creador es únicamente un metadato más; etcétera. Hay que tener en cuenta, además, que los catálogos contemporáneos, que son los que constituyen el foco de atención de nuestros proyectos, cuentan con mucha más información y distintas formas de disposición que la que podemos encontrar en los catálogos históricos, tal y como se describe en (Gabay *et al.*, 2021), en los que se aprecia una disposición estable de los datos y de la información sobre el autor.

En todos los catálogos analizados aparece la información referente al artista o artistas participantes en la exposición, el lugar y momento de celebración, el comisario y comité organizador,

la entidad responsable de su realización y otros datos de interés muy relevantes para un análisis del ecosistema expositivo completo, como menciones a exposiciones anteriores, además de, como es de esperar, la información catalográfica, en general numerada (aunque no siempre fiel a la disposición habitual) y con datos clave como el título de la pieza, su creador, la fecha de creación, la técnica de elaboración, las dimensiones y, en algunos casos, el propietario actual y/o propietarios previos, además del lugar de creación, exposiciones anteriores, texto descriptivo e imagen. Asimismo, se debe indicar que, en general, aparece también texto unas veces con referencia al creador y su historia y otras veces textos ensayísticos acerca de la exposición y cuestiones relacionadas.

Las entidades que se van a identificar están formadas, en general, por varias palabras, y se encuentran situadas en distintas partes del texto. El resto de la información específica suele estar estructurada, generalmente en modo de listas anidadas y hacia el final de los catálogos. No obstante, como ya se ha indicado, la estabilidad no está garantizada ni tampoco un formato fijo, lo que hace que el proceso de etiquetado y la subsecuente extracción de información sea una tarea muy compleja. De este modo, el etiquetado se realiza con detenimiento y revisando punto por punto el catálogo en toda su extensión para identificar todas aquellas entidades relevantes para el objetivo establecido.

Originalmente, los textos se obtuvieron en formato pdf. Posteriormente se convirtieron en texto plano y se limpiaron para la eliminación de símbolos y términos no relevantes que podrían dar problemas en el procesamiento y etiquetado automático. A continuación, se etiquetan siguiendo una guía previamente elaborada con el protocolo de etiquetado.

Creación de una guía de anotación

Con el objeto de proceder al etiquetado de los textos, se construye una guía de anotación en la que se definen las etiquetas que se van a usar para delimitar los elementos de información que se pretende rescatar de los catálogos.

Entre estas, se seleccionaron tanto las etiquetas habituales para la identificación de entidades nombradas (personas, organizaciones, etc.), como etiquetas para marcado de la estructura básica del catálogo.

Dado que muchos de los elementos que se identifican cuentan con información adicional, para representar con semántica completa la información, se consideraron también algunos atributos específicos para algunas etiquetas.

En total se han considerado 17 etiquetas.

Las etiquetas que se van a usar para las entidades nombradas son:

■ Etiqueta para persona (etiqueta *person*) con el atributo *role* que puede tomar los siguientes valores: *creator*, *owner*, *funder*, *exhibitionCurator*, *museumCurator*, *exhibitionManager*, *assistant*.

Se identificación igualmente las personas que intervienen en la edición del catálogo como entre otros el autor (atributo *autor*) y otros como los maquettadores y editores.

■ Etiqueta para instituciones (etiqueta *institution*) que puede contar con dos atributos, para determinar el tipo de institución (atributo *type*) con posibles valores: *museum*, *foundation*, *company*, *gallery*, *artCenter* y el atributo *role* para indicar el papel que desempeña la institución en la exposición, con valores como: *organiser*, *host*, *funder*, *owner*, *depository*, *lender*.

■ Etiqueta para la exposición (*exhibition*) y su información como el título (*title*) y la fecha (*date*), entre otras.

En lo que respecta al etiquetado de las estructuras propias de los catálogos se cuenta con etiquetas como *catalogue* que delimita el conjunto de los elementos del catálogo, *catalogueItem* para marcar cada uno de los elementos del catálogo que representan a obras o conjunto de obras con su nombre, creador y sus características como la técnica de creación (etiqueta *technique*), dimensiones (etiqueta *dimensions*) y otras.

Corpus de entrenamiento

Para la generación del corpus correcto para el entrenamiento (*gold standard*) (Navarro-Colorado, 2007), se realiza el etiquetado manual de 120 catálogos en formato de texto plano que han sido sometidos a los procesos de limpieza y preparación necesarios.

El etiquetado ha sido realizado por dos etiquetadores con el objeto de generar el corpus buscado a partir del acuerdo entre estos.

Como ya se ha mencionado en el apartado anterior, la selección de los catálogos se ha realizado aplicando una serie de criterios que garanticen que la muestra es representativa y que contempla las distintas estructuras que se presentan en los catálogos analizados.

Para llevar a cabo el etiquetado manual se usó el conjunto de herramientas GATE¹, de código abierto y que, entre otras funcionalidades, cuenta con la opción de definir una serie de etiquetas para anotación de los textos.

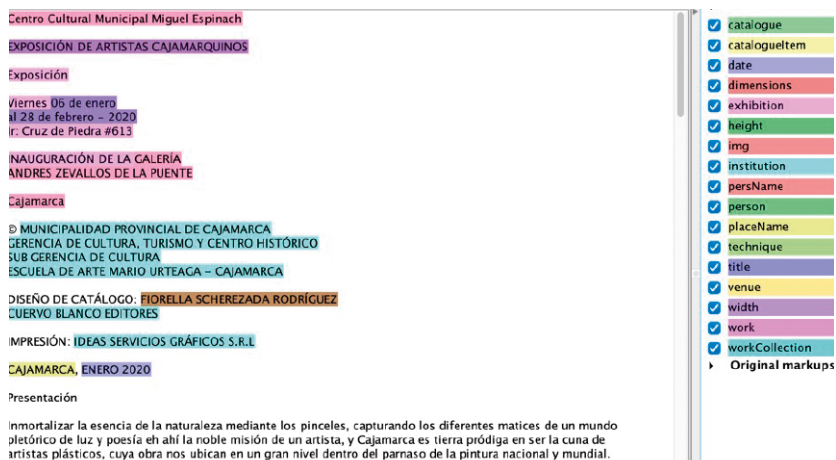


Figura 3. Etiquetado de un texto con GATE.
Elaboración propia

1 <https://gate.ac.uk>.

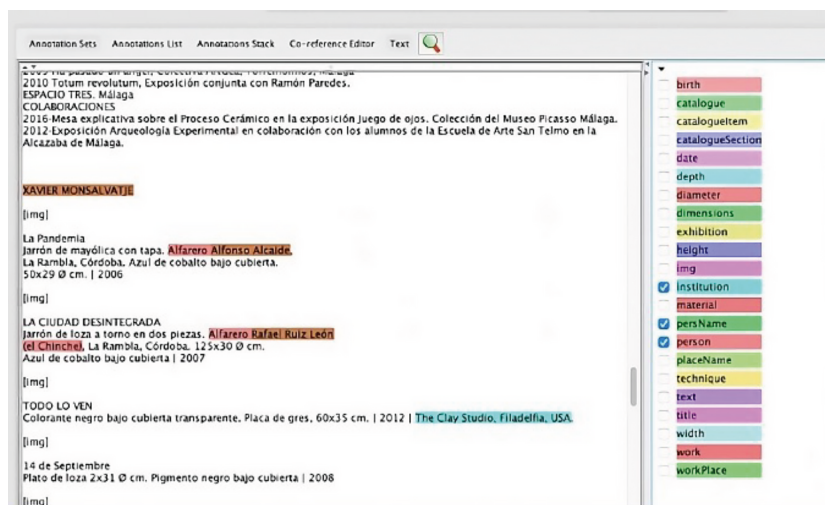


Figura 4. Etiquetado manual con GATE del listado de obras. Elaboración propia

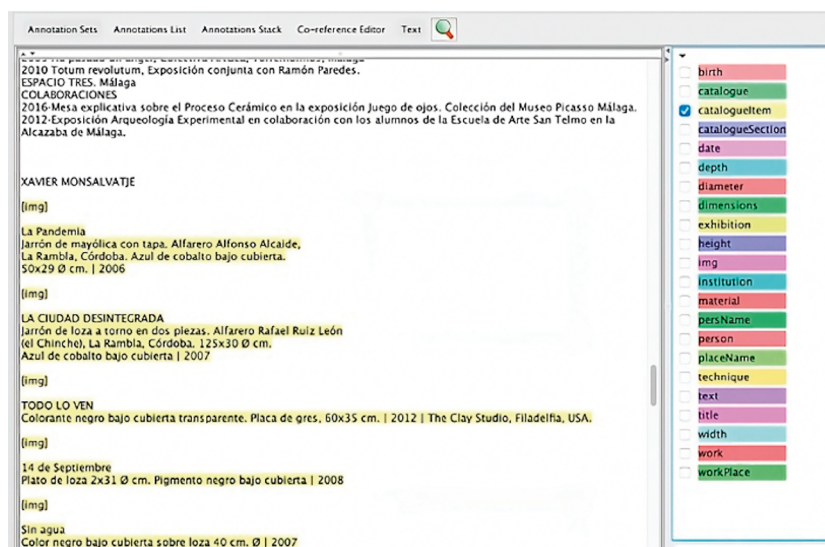


Figura 5. Etiquetado manual con GATE. Elementos del catálogo. Elaboración propia

En la Fig. 3, se muestra el etiquetado de la parte inicial del texto en el que aparecen datos relativos a la exposición y la edición del catálogo, entre otros.

En las figuras 4 y 5 se presenta el etiquetado de la parte del listado de obras que aparecen en el catálogo. En esta parte, hay anidamiento de etiquetas, en la figura 4 se muestran algunas anotaciones de personas o instituciones y en la Figura 5, la anotación de los elementos de catálogo completo. Estas anotaciones contienen otras como las indicadas en la Figura 4, título de las obras, características técnicas y tamaño, entre otras.

Tras la realización del etiquetado, se obtiene un archivo XML.

El corpus estará disponible en <https://github.com/Complexhibit-Project/Corpus>.

Construcción del conjunto de datos

La preparación del conjunto de datos para el entrenamiento y validación requiere procesar los XML obtenidos al realizar el etiquetado manual. El objetivo del proceso es extraer la información de las anotaciones considerando no sólo el tipo de elementos obtenidos, sino también su extensión y los atributos asociados a elemento. Esta información, de la que se muestra un extracto en la Figura 6, requiere ser integrada por frases para generar la entrada del modelo.

	Id	Type	StartNode	EndNode	Attributes
0	0	person	0	19	[(role, creator)]
1	1	person	88	107	[(role, creator)]
2	2	person	327	346	[(role, creator)]
3	3	person	4650	4669	[(role, creator)]
4	4	person	4706	4725	[(role, creator)]
5	5	person	4782	4801	[(role, creator)]
6	6	person	5530	5538	[(role, creator)]
7	7	person	6166	6174	[(role, creator)]
8	8	person	5965	5973	[(role, creator)]
9	9	person	6909	6928	[(role, creator)]

Figura 6. Extracto de las etiquetas y sus atributos de un texto. Elaboración propia

Para la generación del conjunto de datos por frases que se ha usado en el entrenamiento, se requirió la realización de las correspondencias entre las palabras de cada frase, el índice por el que van a estar representadas y las etiquetas y su índice de representación (Figura 7).

Sentencia	Word	Tag	Word_index	Tag_index
1	[Centro, Cultural, Municipal, Miguel, Espinach]	[B-title, I-title, I-title, I-title-title]	[20,35,43,24,50]	[3,4,4,4,5]

Figura 7. Conjunto de datos para una frase. Elaboración propia

3.2. IMPLEMENTACIÓN DEL MODELO LSTM

El modelo LSTM se ha implementado en Python y se han usado las principales bibliotecas para la implementación de un modelo de aprendizaje profundo de las características indicadas, como *keras*² y *tensorflow*³.

Construcción del modelo LSTM

En el modelo construido se han utilizado dos capas: una capa de entrada y una capa para incrustaciones (*embedding*) mediante el modelo ELMO (Tiwari, 2019) para la representación de la secuencia de palabras como la secuencia de sus vectores correspondientes y que se considera adecuado para esta propuesta.

En la compilación del modelo LSTM se utiliza el optimizador *Adam*.

En el desarrollo se ha realizado, inicialmente, una adaptación de la propuesta presentada en (Sakar, 2021) teniendo en cuenta las características específicas de los textos utilizados y las estructuras que contienen la información de las obras.

La entrada al modelo para su entrenamiento es el conjunto de sentencias con la estructura indica en el apartado anteriores. Para el entrenamiento y validación se han usado 49.680 sentencias

2 <https://keras.io>.

3 <https://www.tensorflow.org/?hl=es-419>.

con una media de 3 palabras por frase, lo que es un tamaño bueno para un entrenamiento de este tipo (Figura 7), pero en el futuro no se descarta incrementar el tamaño del conjunto si los datos del entrenamiento no mejoran.

Los archivos y la descripción completa del sistema estarán disponibles en <https://github.com/Complexhibit-Project/Automatic-Tagged-System>.

Evaluación

Tras el entrenamiento completado hasta el momento se ha obtenido una precisión del 0,65. No obstante, para la evaluación del modelo, la principal métrica que se considera es la puntuación F1, como equilibrio entre precisión y exhaustividad, para la que se obtiene un 63%, que no puede considerarse una muy buena puntuación.

Precisamente una de las cuestiones que se han analizado son los anidamientos en las entidades. Existen varias técnicas propuestas que puede dar buenos resultados como métodos de clasificación basados en span y el uso de grafos como se propone en (Wan, J. *et al.*, 2022) o bien o modelos de lenguaje extensos que abren una vía de aplicación para estos casos.

Por tanto, en las nuevas etapas de se pretende adaptar el modelo para tratar estos anidamientos y realizar la evaluación de los resultados con las técnicas usadas.

Conclusiones

Los valores de evaluación obtenidos requieren un ajuste del modelo para poder mejorar su rendimiento. Por tanto, las acciones se enfocarán principalmente en los tres aspectos que se enumeran a continuación:

1. Realización del entrenamiento para más épocas.
2. Tratamiento de las anotaciones con anidamiento mediante distintas técnicas.
3. Selección de nuevos textos para ampliar el corpus. En especial, se seleccionarán con características para las

que se han obtenido menos aciertos y, por tanto, menor precisión.

Unas veces que se consigan las métricas deseadas, realizando los diferentes ajustes se procederá a la aplicación de modelo y a la transformación de los resultados para almacenarlo en el archivo JSON que será la fuente de entrada para el sistema que generará los triples para su integración en el grafo de conocimiento de la ontología OntoExhibit.

Notas

[1] Esta comunicación es resultado de los proyectos de investigación: *Diseño de un modelo semántico guiado por ontologías del dominio de las exposiciones artísticas para la generación de conocimiento y valor en el ámbito de las Industrias Culturales y Creativas (ICC)* (UMA20-FEDERJA-126); *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* (PY20_00508); *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).

[2] Subtarea de la extracción de información que identifica distintos tipos de entidades, como pueden ser nombres de personas, nombres de instituciones, nombre de ubicaciones, entre otras.

[3] <https://w3id.org/OntoExhibit/>

REFERENCIAS BIBLIOGRÁFICAS

- CHIU, Jason P. C. y Eric NICHOLS, “Named Entity Recognition with Bidirectional LSTM-CNNs”, *Transactions of the Association for Computational Linguistics*, 3, (2016), pp. 357–370.
- DIEZ PLATAS, María Luisa, Salvador ROS MUÑOZ, Elena GONZÁLEZ-BLANCO, Pablo RUIZ FABO y Elena ÁLVAREZ MELLADO, “Medieval Spanish (12th–15th centuries) name entity recognition and attribute annotation system based on contextual information”, *Journal of the Association for Information Science and Technology*, 72(2), (2021), pp. 224–238.
- GABAY, Simon, Barbara TOPALOV, Caroline CORBIÈRES, Lucie RONDEAU DU NOYER, Béatrice JOYEUX-PRUNEL y Laurent ROMARY, “Automating Art1@s– extracting data from exhibition catalogues”, *EADH 2021–Second International Conference of the European Association for Digital Humanities*, (2021), pp. 1–9.
- NAVARRO-COLORADO, Borja, “Metodología, construcción y explotación de corpus anotados semántica y anafóricamente”, Universidad de Alicante, 2007.
- SARKAR, Subham, “Named Entity Recognition using Deep Learning (ELMo Embedding+ Bi-LSTM)”, *Analytics Vidhya*, <<https://medium.com/analytics-vidhya/named-entity-recognition-using-deep-learning-elmo-embedding-bi-lstm-48295bc66cab>> [consulta: 10/10/2023].
- TIWARI, Abhisht, “ELMo Embedding, Medium”, <<https://medium.com/@abhisht85/elmo-embedding-3c7bd0d-f20d2>> [consulta: 01/10/2023].
- WAN, Juncheng, Dongyu RU, Weinan ZHANG y Yong YU, “Nested Named Entity Recognition with Span-Level Graphs”, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 1, (2022), pp. 892–903.

COMPARATIVA DE LOS SENTIMIENTOS Y LAS EMOCIONES BÁSICAS MÁS FRECUENTES EN LAS NOVELAS DE JANE AUSTEN A TRAVÉS DEL ANÁLISIS DE SENTIMIENTOS

Eduarne Sáenz Moral

Resumen

A pesar de la breve aportación de Jane Austen a la literatura, autora de la que solo se publicaron seis novelas completas, esto no ha impedido que sea considerada uno de los referentes canónicos de la literatura universal. Sus obras se caracterizan por el retrato que hacen de la burguesía agraria georgiana, con mujeres jóvenes como protagonistas y una sutil crítica a la sociedad. Todo ello permite que sus novelas sean fácilmente reconocibles por el lector y nos lleva a plantearnos una cuestión que se convirtió en el objetivo de este trabajo: comparar las emociones básicas y sentimientos presentes en las seis novelas para determinar si estas siguen un patrón idéntico tal y como sucede con los demás elementos mencionados. Para ello, en primero lugar se ha utilizado el lenguaje de programación R y varios de sus paquetes, que nos han ayudado a comprobar y confirmar las similitudes presentes en cuanto a sentimientos y emociones. Posteriormente, teniendo en cuenta las limitaciones presentes en este primer análisis, se ha considerado el estado del arte del análisis de sentimientos,

Eduarne Sáenz Moral, UNIR
trad.tesm@gmail.com
ORCID: 0009-0002-9794-6718

actualmente basado en el aprendizaje profundo, para volver a realizar un análisis, esta vez utilizando en Python el modelo de lenguaje pre-entrenado para minería de opinión pysentimiento.

Palabras clave: Jane Austen, análisis de sentimientos, emociones básicas, Syuzhet, pysentimiento.

Abstract

Despite the brief contribution Jane Austen made to literature (she only published six complete novels), this author is considered as one of the canonical references in universal literature. Her works are characterized by the portrait she made of Georgian rural bourgeoisie, by having young women in the leading roles, and by containing a subtle critic to her society. All these facts make her novels easily recognizable by the reader, and they make us set out a question which became the objective of this essay: comparing the basic emotions and sentiments present in Austen's six novels. The aim is to establish whether they follow a pattern, as other characteristics from the novels do. In order to perform it, firstly we have employed the programming language R and several of its packages, which have helped us to verify and confirm the similarities regarding sentiments and emotions. Subsequently, considering the limitations of this first analysis, we have considered the state of art in sentiment analysis, currently based on deep learning. For this reason, we have developed a new analysis using pysentimiento, a Python pre-trained language model for opinion mining.

Keywords: Jane Austen, Sentiment analysis, Basic emotions, Syuzhet, Pysentimiento.

INTRODUCCIÓN

La influencia de Jane Austen en la literatura universal es, parafraseando a la propia autora, una verdad universalmente reconocida. A pesar de contar con detractores de su obra, es habitual verla incluida en los listados de autores clásicos de imprescindible lectura y en el canon literario occidental, como el elaborado por el crítico Harold Bloom (1994).

Labrarse dicha fama no fue algo sencillo para Austen, ya que durante los siglos XVII y XVIII, época en la que vivió, la novela estaba considerada un género literario menor que no podía competir con el prestigio del género ensayístico filosófico o los tratados sobre historia, escritos por hombres (Wyett, 2015). Así, la novela resultaba idónea para aquellas mujeres que querían desarrollar su creatividad, pero que se veían obligadas a firmar de forma anónima para evitar perder su reputación por verse expuestas públicamente. A pesar de estas trabas, tanto crítica como público comenzaron rápidamente a hablar de Austen, llegando a ser una autora estudiada en el ámbito académico universitario (Owen, 2018) y a tener un número de admiradores que no hace sino aumentar con el paso del tiempo.

Pero ¿qué hace que la obra de una autora que solamente escribió seis novelas completas sea tan especial y característica? De acuerdo con autores como Waldron (1999), en Austen destaca su capacidad para convertir en interesantes las situaciones más cotidianas y el poder de mantenerse fiel a la realidad sin necesidad de recurrir al sensacionalismo ni crear modelos de conducta, algo habitual en las novelas de su época. La autora retrató lo que mejor conocía: la clase social a la que ella misma pertenecía, la burguesía agraria georgiana, a través de historias protagonizadas por mujeres jóvenes, que bien podrían ser ella y sus hermanas. Y, a todo ello, le añadió una sutil crítica social llena de ironía y humor.

La presencia constante en todas sus novelas de las características anteriormente mencionadas hace que la obra de Austen sea fácilmente reconocible, pero, además, a esto debemos añadir el uso de un lenguaje con unos rasgos muy concretos: un vocabulario conservador y clásico combinado con una sintaxis experimental y un léxico innovador (Luu, 2018), pero a su vez natural y fluido, que permiten que la “voz” de Jane Austen sea reconocida de inmediato (Page, 2011). Además, la forma de estructurar sus novelas es similar en todas ellas: cuentan con una primera parte en la que nos presenta a la heroína de la historia y una segunda que comienza con la aparición del antagonista principal (Jordán Enamorado, 2017). La

trama fluye sin prisa, dedicando tiempo a conocer a los protagonistas, para finalmente acelerarse hasta llegar al momento álgido que nos lleva a la resolución de la novela: el final feliz en forma de matrimonio.

Por este motivo, sabiendo que las seis novelas narrativas creadas por la autora (*Sense and Sensibility*, *Pride and Prejudice*, *Mansfield Park*, *Emma*, *Northanger Abbey* y *Persuasion*) cuentan con un lenguaje, estructura y temática similares y característicos de su estilo, nos planteamos la hipótesis de si los sentimientos y las emociones básicas mostrados por los personajes de sus novelas también siguen unos patrones comunes y si están presentes en el mismo grado o si, por el contrario, las presencias diferentes de sentimientos y emociones son las que permiten hacer que cada novela sea distinta a pesar de los rasgos compartidos por todas.

Para realizar esta comparativa, se ha llevado a cabo un procesamiento del lenguaje natural, concretamente un análisis de sentimientos de las seis novelas. Se han analizado una a una con el objetivo de detectar los patrones presentes en cada una de ellas y, de este modo, establecer similitudes y diferencias entre las seis novelas. Considerando la popularidad de Jane Austen, no era de extrañar que ya existan análisis de sentimientos de sus obras. Sin embargo, nos hemos encontrado con carencias en dichos análisis: bien se centran en una única obra o analizan varias de forma individual sin llegar a hacer una comparativa, como en el análisis realizado por McKenzie (2018), o bien solamente determinan patrones entre sentimientos positivos, negativos o neutros, pero no prestan atención a las emociones básicas, como en el estudio de Horiuchi (2021).

METODOLOGÍA

El análisis de sentimientos es una de las aplicaciones que nos permite realizar el procesamiento del lenguaje natural y que consiste en el análisis de los sentimientos, las opiniones y las emociones de las personas hacia productos, organizaciones o temas, entre otros

(Liu, 2012), expresados en valores de positivo, negativo o neutro. Gracias al auge de los medios sociales, principalmente a partir del año 2000, se detectó el potencial de este tipo de estudio con fines comerciales, ya que permite a las empresas analizar las reseñas publicadas en la red sin necesidad de crear encuestas específicas. Con el paso del tiempo, esta técnica se ha extendido a otros ámbitos, como el político o el médico por mencionar algunos, especialmente gracias a las posibilidades que ofrece al permitir analizar la información creada por los propios usuarios de la World Wide Web. Asimismo, también se ha visto su utilidad en el ámbito académico para estudios literarios, puesto que posibilita un acercamiento al texto desde una perspectiva diferente a la humanística tradicional. Si bien es cierto que, debido a cuestiones como los derechos de autor y las dificultades que conllevan para recopilar el corpus a analizar, el número de trabajos realizados sobre análisis de sentimientos aplicados a la literatura es inferior a otros ámbitos. Sin embargo, este trabajo se ve facilitado cuando se opta por analizar obras que han pasado a ser de dominio público, como en nuestro caso, y que además nos ofrece la posibilidad de visitar dichas obras y extraer nueva información y conclusiones sobre ellas. De esta manera, se puede analizar fácilmente en su totalidad la obra de un autor, de una época creativa o un amplio volumen de obras de un género específico, sin necesidad de dedicar muchos meses, o incluso años, a esta tarea.

Gracias al aumento del uso comercial del análisis de sentimientos, han sido muchas las herramientas creadas para ello, tanto en el ámbito académico como por parte de empresas privadas. Para hacer un análisis de sentimientos de un corpus lingüístico contamos con softwares como, por ejemplo, SentiWordNet, basado en el diccionario WordNet, que reúne las palabras por categoría gramatical y por grupos de sinónimos o *synsets*. Cada uno de estos *synsets* tiene una valoración de positivo, negativo o neutro. También existen otros programas como Linguistic Inquiry and Word Count (LIWC), que analiza emociones y rasgos cognitivos basándose en un diccionario con palabras divididas por categorías;

SenticNet, que detecta la polaridad de los sentimientos en función de la frecuencia de la coocurrencia de las palabras; Opinion Parser; RapidMiner, o Semantria, por mencionar algunos softwares. Otra opción es utilizar lenguajes de programación a los que se instalan paquetes concretos que permitan realizar este tipo de análisis. Entre los más empleados están Python, que cuenta con paquetes como SailAil Sentiment Analyzer, o R, con paquetes como Rsentiment o Syuzhet. Por último, también existen aplicaciones web como Mallet, Google Cloud Natural Language o Meaning Cloud, que dan la opción de realizar diferentes procesamientos de lenguaje natural.

Teniendo en cuenta las características que ofrece cada una de las herramientas anteriormente mencionadas, la opción elegida para este análisis de las novelas de Jane Austen ha sido el lenguaje de programación R junto con el paquete Syuzhet. El principal de los motivos que nos ha llevado a decantarnos por él ha sido la posibilidad de analizar las novelas no solo por valores de sentimientos positivos o negativos, sino también profundizar en las emociones básicas presentes en ellas. Pero a esto también debemos añadir que R es un lenguaje de libre acceso, gratuito, multiplataforma, lo que fomenta la interoperabilidad, altamente flexible y ampliable gracias a los paquetes que permite añadirle, ya sean creados por los propios diseñadores del lenguaje o por programadores ajenos, y que ofrece numerosas opciones a la hora de mostrar los resultados, con salidas más dinámicas y atractivas (Equipo de edX, 2021). Por todo ello, R ha conseguido ser uno de los lenguajes más empleados en el mundo académico (Redondo Figuero, 2017).

Respecto al paquete Syuzhet, empleado para el análisis de sentimientos y emociones básicas, cabe destacar que fue desarrollado por Matthew L. Jockers en 2015. Consta de cuatro diccionarios de sentimientos: Bing, Afinn, Stanford y NRC, entre los que el usuario puede elegir (Isasi, 2021). En este caso, optamos por NRC, ya que lo forma un listado de palabras en inglés, idioma original de las novelas, que han sido asociadas a ocho emociones básicas: alegría, anticipación, confianza, disgusto, enfado, miedo, sorpresa y tristeza, y a dos sentimientos: positivo y negativo, de forma ma-

nual por parte de colaboradores (Mohammad y Turney, 2012). El paquete convierte el texto en vectores, que tokeniza en oraciones o palabras según nuestra elección. Estos tokens los asocia a sentimientos o emociones concretos en función del diccionario elegido y nos posibilita el mostrarlo en diferentes tipos de gráficos.

Aunque es cierto que estas características hacen que Syuzhet sea un buen recurso para conseguir el objetivo que nos hemos marcado, hay que tener en cuenta ciertas carencias encontradas. Por un lado, hay ocasiones en las que dependiendo de la puntuación utilizada no hace correctamente la división en oraciones. En este caso, dado que hemos optado por una tokenización por palabras no nos encontramos con este problema. Por otro lado, Syuzhet no termina de analizar correctamente oraciones negativas. En ejemplos como “No estoy feliz” el paquete detectará la palabra “feliz” y la categorizará como positiva, a pesar de que se encuentra negada (Swafford, 2015).

El proceso seguido para analizar las seis novelas de Jane Austen que forman nuestro corpus es el siguiente (figura 1):

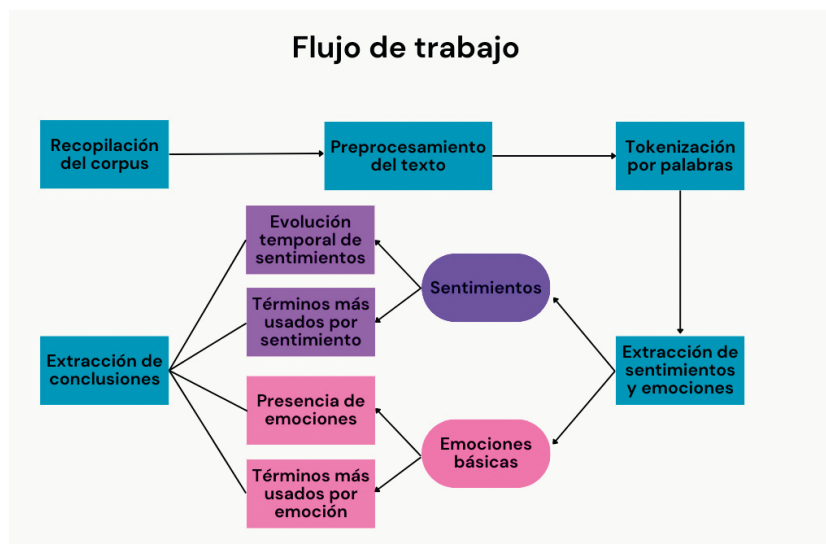


Fig. 1. Elaboración propia, Flujo de trabajo, 2023

Para comenzar, procedemos a la recopilación del corpus. Considerando que el análisis con R requiere que el corpus esté en formato de texto plano, utilizamos la biblioteca online Proyecto Gutenberg (s.f.), la cual ofrece la posibilidad de descargar las seis novelas en lengua inglesa en dicho formato. Posteriormente, llevamos a cabo el preprocesamiento de los textos con el fin de limpiar los seis archivos, correspondientes a cada una de las novelas, y eliminar así las secciones irrelevantes, de modo que se preservara únicamente el contenido literario principal. Una vez obtenidos los textos procesados, los cargamos en seis variables de tipo vector. Con el paquete Syuzhet ya instalado, realizamos la tokenización por palabras, lo cual permite determinar el sentimiento y las emociones asociadas a cada término, conforme a la anotación proporcionada por el diccionario NRC. Como resultado nos dará seis variables de tipo lista con cada uno de los unigramas que conforman las novelas. A partir de este punto, en la investigación se abren dos caminos: por un lado, el análisis de los sentimientos y, por otro, el de las emociones básicas.

En cuanto a los sentimientos, veremos cómo avanzan en la línea temporal de cada una de las novelas de Austen, lo que nos permitirá conocer si es frecuente que comiencen con un sentimiento concreto para evolucionar hacia otro diferente, y si las curvas de las trayectorias sentimentales son similares en todas las novelas, lo que permitiría determinar la existencia de una evolución de los sentimientos característica de la autora. Respecto a las emociones básicas, nos centraremos en ocho: alegría, anticipación, confianza, disgusto, enfado, miedo, sorpresa y tristeza. Se comprobará la presencia de cada una de ellas en las novelas para determinar cuáles son las más frecuentes tanto individualmente como de forma global en la obra de Austen. Además, en ambos casos se analizarán los términos más frecuentes utilizados para expresar cada sentimiento y emoción, y así establecer similitudes y diferencias entre novelas. Para todo ello, también se han utilizado paquetes adicionales que ofrecen visualizaciones más atractivas de los resultados, tales como RColorBrewer (Neuwirth, 2022), Wordcloud

(Fellows, 2018), Tm (Feinerer, Hornik y Artifex Software, Inc., 2023), Tidyverse (Wickham *et al.*, 2019) y Tidytext (Silge y Robinson, 2016).

RESULTADOS OBTENIDOS

Con los textos ya procesados y tokenizados, extraemos los sentimientos y emociones básicas de cada una de las novelas anotados con el diccionario NRC. Los resultados obtenidos se almacenan en una variable de tipo data frame, sin embargo, su visualización no nos da la posibilidad de extraer conclusiones, por lo que tenemos que procesarlo para obtener información más valiosa, es decir, los valores mínimos y máximos para los diez casos (dos sentimientos: positivo y negativo, y las ocho emociones básicas anteriormente mencionadas).

En primer lugar, como se aprecia en las siguientes tablas, los sentimientos positivos son más frecuentes en todas las novelas que los negativos, con una frecuencia incluso del doble en algunos casos (tabla 1). Así, atendiendo a los valores de la media aritmética, podemos ver que *Persuasion* es la novela con una frecuencia más positiva de las seis, no solo porque tenga la mayor valencia positiva sino porque su valencia negativa es la segunda más baja de todo el corpus. Por el contrario, encontramos que *Sense and Sensibility* es la novela con una mayor valencia negativa, pero a su vez es la segunda con una valencia positiva más alta.

Tabla 1. Medias de sentimientos positivo y negativo

NOVELA	NEGATIVO	POSITIVO
<i>Sense and Sensibility</i>	0.03278	0.06043
<i>Pride and Prejudice</i>	0.02902	0.0592
<i>Mansfield Park</i>	0.0288	0.05804
<i>Emma</i>	0.02681	0.05667
<i>Northanger Abbey</i>	0.03095	0.05865
<i>Persuasion</i>	0.02683	0.06383

Fuente: Elaboración propia, 2022

En cuanto a las ocho emociones básicas, podemos comprobar que las cifras de cada emoción son muy similares en las seis novelas, lo que nos permite pensar que existen patrones muy equilibrados en cuanto al uso de emociones. Así, podemos ver que aquellas emociones que podemos considerar positivas, como la alegría o la confianza, son las que cuentan con valores más elevados en los seis casos (tablas 2 y 3).

Tabla 2. Medias de emociones básicas (parte 1)

NOVELA	IRA	ANTICIPACIÓN	DISGUSTO	MIEDO
<i>Sense and Sensibility</i>	0.01083	0.03025	0.009642	0.01497
<i>Pride and Prejudice</i>	0.01023	0.02859	0.007704	0.01402
<i>Mansfield Park</i>	0.01067	0.0296	0.008039	0.01455
<i>Emma</i>	0.009252	0.02894	0.007918	0.014
<i>Northanger Abbey</i>	0.01074	0.02886	0.008773	0.01521
<i>Persuasion</i>	0.009684	0.02867	0.007699	0.01494

Fuente: Elaboración propia, 2022

Tabla 3. Medias de emociones básicas (parte 2)

NOVELA	ALEGRÍA	TRISTEZA	SORPRESA	CONFIANZA
<i>Sense and Sensibility</i>	0.02734	0.01701	0.01279	0.03479
<i>Pride and Prejudice</i>	0.0265	0.01406	0.01322	0.03359
<i>Mansfield Park</i>	0.02602	0.01499	0.01313	0.03452
<i>Emma</i>	0.02633	0.01354	0.01325	0.03385
<i>Northanger Abbey</i>	0.02623	0.01534	0.01416	0.03462
<i>Persuasion</i>	0.02569	0.01464	0.01399	0.03613

Fuente: Elaboración propia, 2022

Pero ¿cómo evolucionan los sentimientos a lo largo de cada una de las novelas? Para favorecer la interpretación de los datos obtenidos, aplicamos la función `simple_plot()` que construye dos gráficas en las que podemos visualizar en una línea temporal la evolución de los sentimientos a lo largo de cada novela: una con tres líneas, cada una de ellas perteneciente a un méto-

do de alisado diferente (concretamente la línea azul corresponde a la regresión local o alisado de Loess, la gris representa la media móvil o *rolling mean* y la roja está generada por Syuzhet), y otra con una única línea, resultado de la simplificación del primer gráfico. De esta forma, podemos apreciar diferencias en la distribución de los sentimientos en cada una de las novelas, aunque existen varias similitudes reseñables: todas las novelas comienzan en rangos de sentimientos positivos y tienen, a continuación, un descenso hacia valores negativos. Este descenso es más o menos pronunciado en cada novela: así, en *Pride and Prejudice* vemos una línea menos acusada que en *Mansfield Park*, *Emma* o *Persuasion*, donde el descenso es más abrupto. Encontramos también altibajos que no siempre se manifiestan en el mismo momento ni con la misma intensidad. De este modo, en *Sense and Sensibility* el punto más bajo se localiza en la parte central de la novela, mientras que en *Mansfield Park* y *Northanger Abbey* es precisamente el momento central el más positivo de la novela. Destaca también la evolución de *Pride and Prejudice*, donde los momentos con una valencia más positiva se encuentran en la primera parte de la novela. Lo que sí está presente en las seis novelas es la existencia de una curva ascendente al final, que ayuda a dejar al lector con la sensación de final feliz tan característico en las novelas de Austen (figuras 2 a 7).

En cuanto a la terminología utilizada con más frecuencia para expresar cada uno de los dos sentimientos, sorprende ver que el término anotado como negativo más empleado por Austen

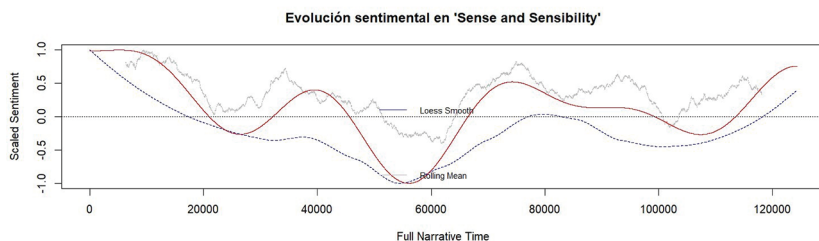
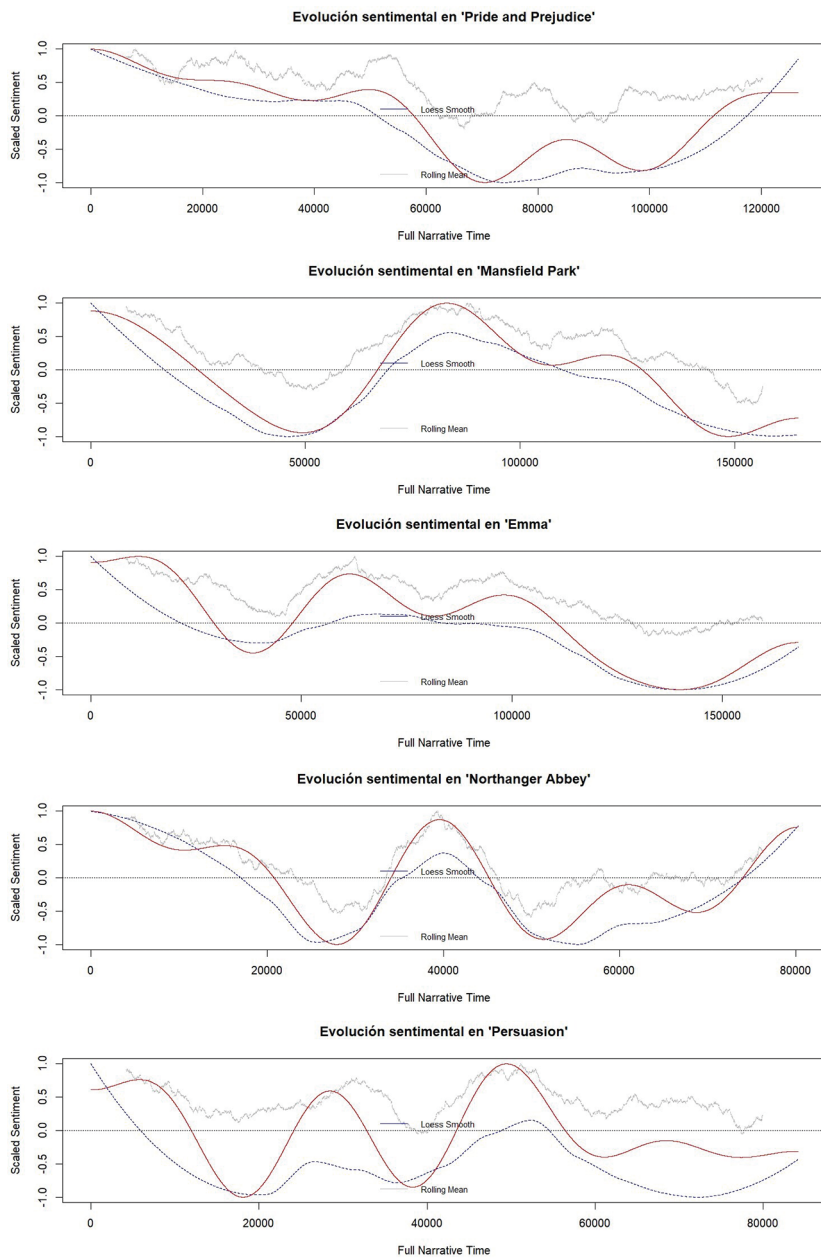


Fig. 2. Elaboración propia, Evoluciones sentimentales, 2022

Encuentros y transformaciones



Figs. 3-7. Elaboración propia, Evoluciones sentimentales, 2022

es *mother*, aunque podemos encontrarlo también en la nube de términos positivos. Esta clasificación negativa no la encontramos por el contrario en otros términos que expresan parentesco, como *aunt* o *brother*, incluidos entre los términos positivos. Este hecho nos lleva a pensar en otra de las limitaciones derivadas de un análisis basado en lexicones: la polisemia. En este caso, al analizar los sentimientos y emociones palabra a palabra, con los términos aislados de su contexto, no se verifica el significado con el que están empleados términos como, por ejemplo, *mother*, que puede ser a la vez sustantivo (con el significado de “madre”) o verbo (*to mother*, que puede significar “parir” o “preocuparse”, “hacerse cargo”), lo que provoca que aparezca presente como uno de los términos más empleados en ambas nubes de palabras.

Además, podemos presenciar el alto protagonismo del ejército en las novelas de Austen, donde encontramos varios términos para expresar cargos militares entre las palabras positivas más utilizadas. En cuanto a otros términos englobados en la valencia negativa, encontramos términos más habituales como *ill*, *evil*, *difficulty* o *dangerous*, entre otros (figura 8).

Palabras negativas y positivas más frecuentes en las seis novelas

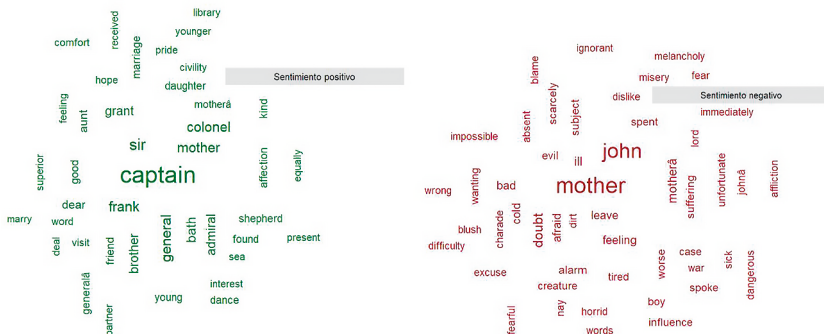


Fig. 8. Elaboración propia, Palabras negativas y positivas más frecuentes, 2022

En lo que concierne a las emociones básicas, hemos elaborado gráficos de barras por cada una de las novelas con el fin de comprobar la frecuencia de cada una de las de cada una de las ocho (figuras 9 a 14). Ya hemos podido ver similitudes entre las valencias positivas de todas las novelas y, gracias a estas visualizaciones, podemos observar otras coincidencias.

Podemos observar que la confianza es la emoción más presente en las seis novelas, seguida por la anticipación y la alegría, las tres con valores muy destacados y todas ellas emociones

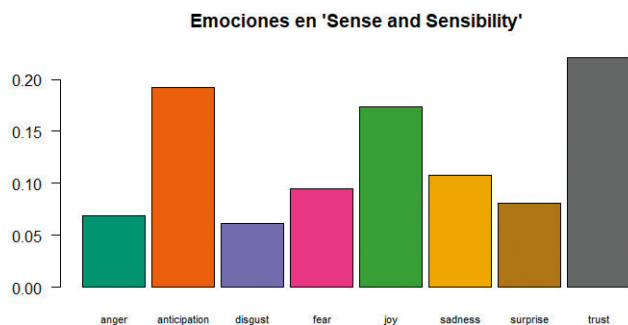


Fig. 9. Elaboración propia, Emociones en *Sense and Sensibility*, 2022

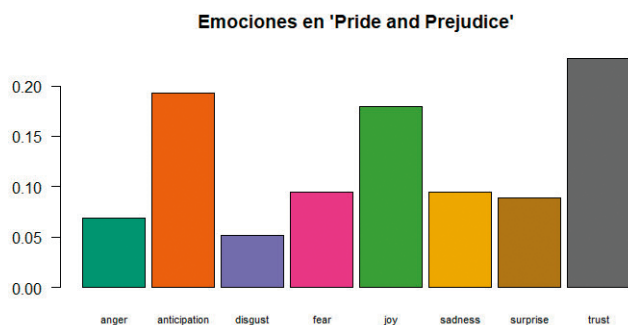


Fig. 10. Elaboración propia, Emociones en *Pride and Prejudice*, 2022

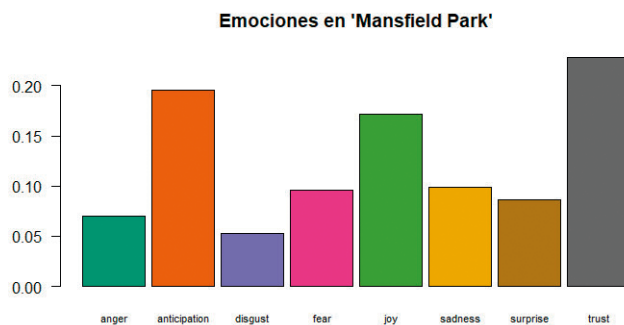


Fig. 11. Elaboración propia, Emociones en *Mansfield Park*, 2022

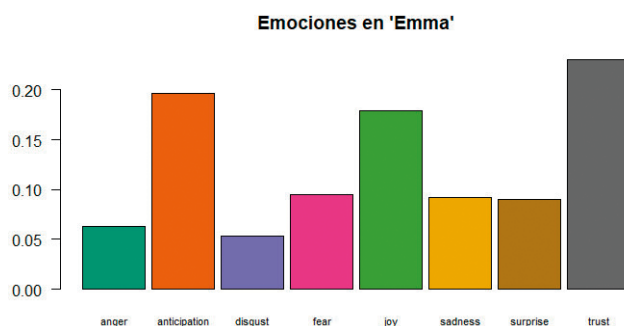


Fig. 12. Elaboración propia, Emociones en *Emma*, 2022

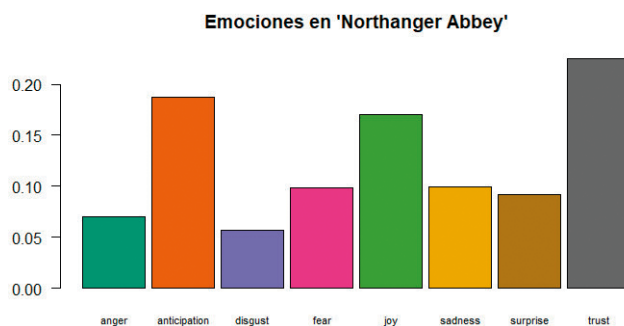


Fig. 13. Elaboración propia, Emociones en *Northanger Abbey*, 2022

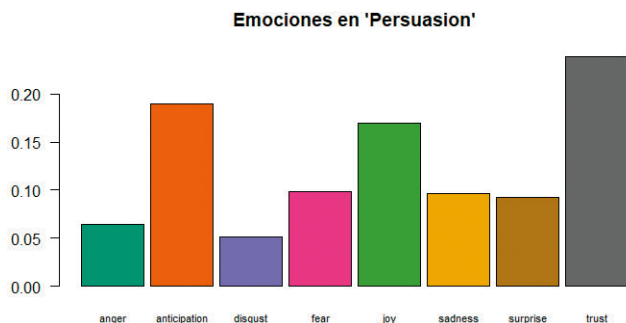


Fig. 14. Elaboración propia, Emociones en *Persuasion*, 2022

que podemos clasificar como positivas. Por otro lado, la primera emoción con carácter más negativo que aparece es el miedo, pero con una presencia bastante inferior a las anteriores. Con los anteriores gráficos podemos comprobar que la posición de todas las emociones es la misma en todas las novelas salvo dos excepciones, *Emma* y *Persuasion*, donde miedo y tristeza cambian sus posiciones, pero sin suponer una variación muy drástica.

Dado que este análisis se basa en un diccionario anotado con emociones y sentimientos, y comprobando la gran similitud entre la frecuencia de las emociones en las seis novelas, ¿debemos pensar que esto implica que el vocabulario empleado por la autora es similar en todas sus obras? Para dar respuesta a esta pregunta, hemos ejecutado una función que nos muestra los términos más utilizados en orden decreciente por cada emoción. Así, podemos crear una nube de palabras por cada emoción para comprobar los términos más empleados para manifestarlas en cada novela, como se puede ver en las figuras 15 a 22.

Mediante estas nubes de palabras, podemos confirmar algo ya comprobado anteriormente: la enorme presencia del ejército en las novelas de Austen, reflejo de la época en la que siempre localizó sus obras, y la consideración positiva de estos, con una gran presencia en emociones como la confianza. Por su parte, podemos comprobar que términos relacionados con el parentesco también se encuentran presentes en la confianza, al igual que el

Nube de palabras para la emoción de anticipación

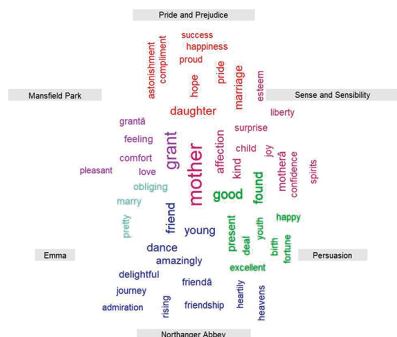


Fig. 15. Elaboración propia, Nube de palabras para alegría, 2022

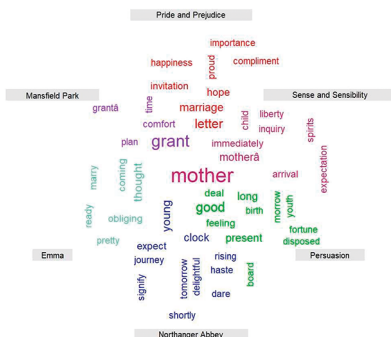


Fig. 16. Elaboración propia, Nube de palabras para anticipación, 2022

Nube de palabras para la emoción de confianza

Nube de palabras para la emoción de disgusto

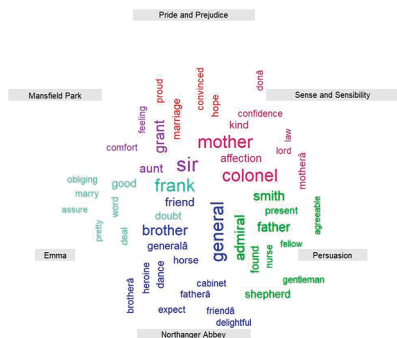


Fig. 17. Elaboración propia, Nube de palabras para confianza, 2022

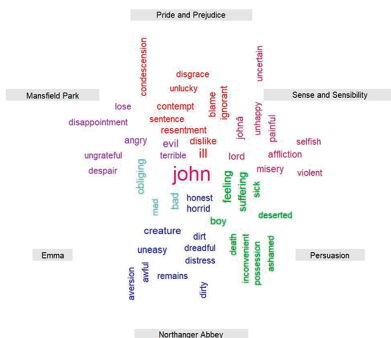


Fig. 18. Elaboración propia, Nube de palabras para disgusto, 2022

Nube de palabras para la emoción de ira

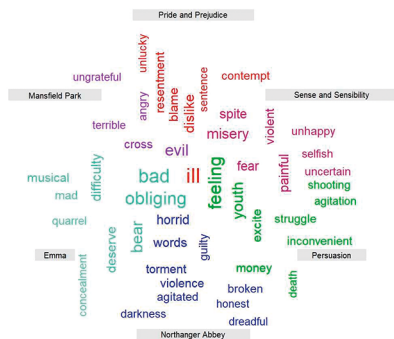


Fig. 19. Elaboración propia, Nube de palabras para ira, 2022

Nube de palabras para la emoción de miedo

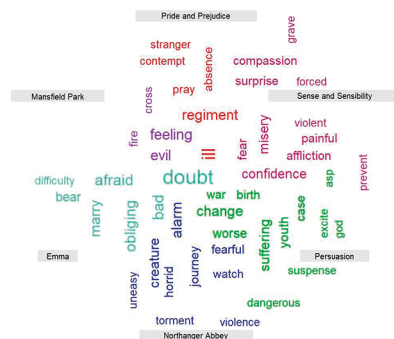


Fig. 20. Elaboración propia, Nube de palabras para miedo, 2022

Nube de palabras para la emoción de sorpresa

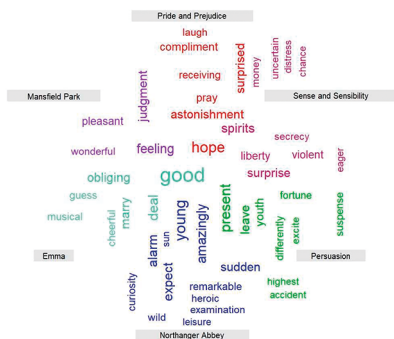


Fig. 21. Elaboración propia, Nube de palabras para sorpresa, 2022

Nube de palabras para la emoción de tristeza

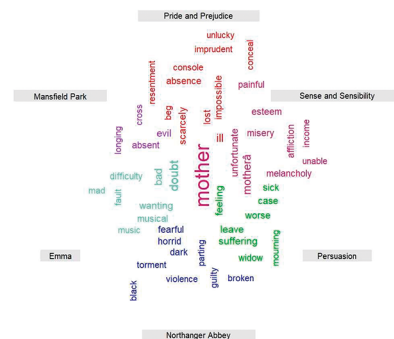


Fig. 22. Elaboración propia, Nube de palabras para tristeza, 2022

Teniendo en cuenta el vocabulario compartido por diferentes emociones, nos planteamos si la emoción menos presente, el disgusto, cuenta con términos muy específicos que hagan que sea la menos habitual en las novelas. Así, podemos comprobar que entre los términos más frecuentes para expresar dicha emoción se encuentra el nombre propio John, presente entre los personajes protagonistas de varias novelas. Además, encontramos otros

términos también presentes en otras emociones negativas, como el miedo o la tristeza, como *ill*, *painfull* o *suffering*.

Si algo podemos sacar en claro de este análisis del vocabulario es que las similitudes entre los términos empleados por Austen en cada una de las novelas nos hacen confirmar, por un lado, ese estilo narrativo característico de la autora y, por el otro, cómo se refleja el bagaje cultural y la subjetividad de los responsables de la anotación de los términos para elaborar el diccionario NRC, y que conlleva que ciertos términos estén incluidos en una emoción y sentimientos determinados.

LIMITACIONES DEL MÉTODO EMPLEADO Y NUEVO ANÁLISIS

Como hemos podido comprobar a lo largo de este proceso de análisis de las novelas de Jane Austen con el paquete Syuzhet, son varias las limitaciones que nos ofrece y que pueden condicionar los resultados obtenidos. La principal limitación viene dada por el hecho de analizar los textos utilizando lexicones, que no tienen en cuenta el contexto de cada palabra, y por tanto repercute en la interpretación de términos polisémicos, oraciones negativas o el cambio semántico. Esto es algo que no está presente en los métodos que utilizan modelos de aprendizaje profundo, como puede ser BERT, y más acordes al estado del arte del análisis de sentimientos.

Por este motivo, hemos buscado nuevos métodos de análisis que nos ofrezcan resultados más precisos y en los que se empleen algoritmos entrenados para identificar patrones en los textos. El uso de estos métodos no es sencillo, ya que su entrenamiento exige una gran cantidad de datos etiquetados, pero ofrecen una precisión más alta (Borrelli y Challiol, 2023), sin embargo, existen modelos ya preentrenados que permiten introducir nuestra información para su análisis. La dificultad reside ahora en encontrar el modelo más acorde a nuestra situación. Por un lado, aunque son muchas las herramientas privadas de pago que podemos encontrar, también tenemos a nuestra disposición numerosos modelos

gratuitos de código abierto. La elección entonces puede basarse en el dominio específico para el que estén entrenados y el idioma empleado. Una de las ventajas de analizar las novelas de Austen en su lengua original es que la mayoría de los modelos están entrenados para utilizarse en inglés, por lo que no contamos con este obstáculo. Sin embargo, el gran uso del análisis de sentimientos para analizar contenido generado por los usuarios en Internet, por ejemplo en reseñas o *tweets*, hace que muchas de estas herramientas se centren en las características de este tipo de textos: mensajes breves de un dominio específico. Además, no son tan abundantes los modelos que introduzcan también el análisis de emociones, lo que nos lleva a descartar varios de ellos.

Partiendo del análisis realizado por Borrelli y Challiol (2023) de varios de estos modelos preentrenados, nos encontramos con Hugging Face, la plataforma que se presenta indicando estar “on a mission to democratize *good* machine learning” (Hugging Face, 2024). Dentro de las publicaciones de su blog podemos leer la entrada *Getting Started with Sentiment Analysis using Python* (Pascual, 2022), donde se nos indican unos sencillos pasos para llevar a cabo análisis de sentimientos. Sin embargo, rápidamente nos encontramos con el primer obstáculo: el modelo está entrenado para analizar *tweets* y la extensión de una novela sobrepasa la longitud permitida para las secuencias a analizar. A pesar de ello, encontramos rápidamente una alternativa dentro de la propia plataforma Hugging Face: el modelo pysentimiento, que nos ofrece no solo la posibilidad de analizar los sentimientos y las emociones básicas, sino que también cuenta con funciones para analizar discursos de odio o detectar la ironía, y está disponible en idiomas como el inglés, español, portugués e italiano (Pérez *et al.*, 2023).

Respecto al análisis que hemos llevado a cabo previamente con Syuzhet, encontramos varias diferencias: por un lado, pysentimiento no nos muestra dos valencias a la hora de presentar los sentimientos, sino tres: positivo, neutro y negativo. Además, en cuanto a las emociones, siguen la teoría de Ekman (1999), quien indica que las emociones básicas son seis: ira, disgusto, miedo,

alegría, tristeza y sorpresa, y no las ocho incluidas en el diccionario NRC que utilizamos con Syuzhet (desaparecen la anticipación y la confianza). Pero los desarrolladores de pysentimiento incluyen una séptima categoría, *other* o *neutral*, por considerar que el etiquetado de emociones es una tarea más subjetiva que el etiquetado de sentimientos, y puede ser el resultado de la combinación de las seis mencionadas emociones básicas o de un conjunto de emociones más detalladas (Pérez *et al.*, 2023) en las que incluyen: admiración, diversión, ira, irritación, aprobación, preocupación, confusión, curiosidad, deseo, decepción, disconformidad, disgusto, vergüenza, entusiasmo, miedo, gratitud, pena, alegría, amor, nerviosismo, optimismo, orgullo, comprensión, alivio, arrepentimiento, tristeza y sorpresa.

Debido a estas diferencias, en cuanto analizamos las seis novelas con pysentimiento encontramos las primeras discrepancias respecto al análisis realizado inicialmente con Syuzhet. Tal y como se puede ver en la tabla 4, con la aparición de la valencia neutra, las novelas ya no se clasifican como positivas, sino como neutras (salvo *Emma*, en la que sigue dominando el sentimiento positivo), e incluso encontramos casos como *Northanger Abbey* donde el sentimiento positivo es el que tiene una menor presencia de los tres.

Respecto a las emociones, nos encontramos con que ya no se analiza de manera independiente la emoción más presente de acuerdo con los resultados ofrecidos por Syuzhet: la confian-

Tabla 4. Valores obtenidos en el análisis de sentimientos con pysentimiento

NOVELA	RESULTADO	POSITIVO	NEUTRO	NEGATIVO
<i>Sense and Sensibility</i>	Neutro	0.156	0.838	0.005
<i>Pride and Prejudice</i>	Neutro	0.117	0.880	0.004
<i>Mansfield Park</i>	Neutro	0.345	0.646	0.009
<i>Emma</i>	Positivo	0.935	0.063	0.002
<i>Northanger Abbey</i>	Neutro	0.014	0.904	0.082
<i>Persuasion</i>	Neutro	0.118	0.859	0.023

Fuente: Elaboración propia, 2024

za, y que en este caso forma parte del grupo de emociones etiquetado como “otros” (tabla 5). Precisamente es este grupo el que más valor tiene en cinco de las seis novelas, con la excepción de *Northanger Abbey*, donde predomina el disgusto. La segunda de las emociones más presentes según nuestros resultados iniciales, la anticipación, tampoco es una de las que se ha incluido en este segundo análisis ni se ha considerado dentro del grupo “otros”. Además, la alegría, tercera emoción según Syuzhet, se muestra con valores muy variados: por ejemplo, en el caso de *Persuasion* es la emoción con menor presencia (a pesar de mostrarse como una de las más frecuentes según los resultados de Syuzhet).

Tabla 5. Valores obtenidos en el análisis de emociones con pysentimiento

NOVELA	OTROS	IRA	DISGUSTO	MIEDO	ALEGRÍA	TRISTEZA	SORPRESA
<i>Sense and Sensibility</i>	0.494	0.004	0.100	0.010	0.010	0.378	0.004
<i>Pride and Prejudice</i>	0.965	0.001	0.017	0.003	0.005	0.006	0.002
<i>Mansfield Park</i>	0.910	0.002	0.072	0.003	0.007	0.004	0.003
<i>Emma</i>	0.946	0.001	0.014	0.003	0.015	0.019	0.002
<i>Northanger Abbey</i>	0.689	0.003	0.289	0.005	0.005	0.006	0.004
<i>Persuasion</i>	0.018	0.005	0.965	0.002	0.001	0.006	0.002

Fuente: Elaboración propia, 2024

A la hora de valorar el significado real de estos resultados, sería interesante conocer qué términos son los que se han incluido en cada una de las emociones y determinar así si obstáculos previos, como la polisemia o la presencia de oraciones negativas, se han superado de la manera correcta. Sin embargo, el modelo utilizado no nos ha permitido la extracción de los términos más frecuentes tanto para los sentimientos como para las emociones.

CONCLUSIONES

El cambio de paradigma hacia un mundo más digital nos abre un nuevo camino a los humanistas, dándonos la oportunidad de dar respuesta a preguntas más ambiciosas como el poder analizar en un breve periodo de tiempo la obra completa de un autor. Gracias al análisis informático de estas novelas, hemos podido detectar patrones no tan perceptibles a simple vista escondidos bajo el vocabulario empleado por la autora.

Sin embargo, es de vital importancia contar con herramientas que analicen de la manera más fiel posible el uso del lenguaje y que se centren en el tipo de contenido que nos interesa. Los resultados obtenidos en los dos análisis llevados a cabo, especialmente las diferencias encontradas, nos llevan a pensar en la importancia del contexto a la hora de analizar el lenguaje natural, ya que las mismas novelas, analizadas con dos metodologías diferentes, nos dan resultados dispares. Esto nos da pie a seguir analizando las novelas utilizando modelos preentrenados con otro tipo de corpus, no tan centrados en el uso de redes sociales sino enfocados en el análisis de dominios del conocimiento más amplios o incluso a trabajar de forma conjunta con los desarrolladores para conseguir modelos de análisis pensados para ser empleados con textos literarios.

REFERENCIAS BIBLIOGRÁFICAS

- AUSTEN, Jane (1994). *Sense and Sensibility*. Project Gutenberg. <https://www.gutenberg.org/files/161/161-0.txt>.
- AUSTEN, Jane (1998). *Pride and Prejudice*. Project Gutenberg. <https://www.gutenberg.org/files/1342/1342-0.txt>.
- AUSTEN, Jane (1994). *Mansfield Park*. Project Gutenberg. <https://www.gutenberg.org/files/141/141-0.txt>.
- AUSTEN, Jane (1994). *Emma*. Gutenberg. <https://www.gutenberg.org/cache/epub/158/pg158.txt>.
- AUSTEN, Jane (1994). *Northanger Abbey*. Project Gutenberg. <https://www.gutenberg.org/cache/epub/121/pg121.txt>.
- AUSTEN, Jane (2008). *Persuasion*. Project Gutenberg. <https://www.gutenberg.org/cache/epub/105/pg105.txt>.
- BLOOM, Harold (1994). *The Western Canon: The Books and School of the Ages*. Harcourt Brace.
- edX (2021). *R vs. Python para la ciencia de datos: Explicación y consejos de aprendizaje*. Blog de edX. <https://blog.edx.org/es/r-vs-python-para-la-ciencia-de-datos-explicacion-y-consejos-de-aprendizaje>.
- BORRELLI, Franco Martín y Cecilia CHALLIOL (2023). Comparing and evaluating tools for Sentiment Analysis. En *XI Jornadas de Cloud Computing, Big Data & Emerging Topics* (pp.36-40). <http://sedici.unlp.edu.ar/handle/10915/155432>.
- EKMAN, Paul (1999). Basic emotions. En Dalglish, Tim y Power, Mick J. (Ed.), *Handbook of cognition and emotion* (98, 45-60, 16). John Wiley & Sons.
- FEINERER, Ingo, Kurt HORNIK y Artifex SOFTWARE, Inc. (2023). Package «tm». *The Comprehensive R Archive Network*. <https://cran.r-project.org/web/packages/tm/tm.pdf>.
- FELLOWS, Ian (2018). Package «wordcloud». *The Comprehensive R Archive Network*. <https://cran.r-project.org/web/packages/wordcloud/wordcloud.pdf>.
- HORIUCHI, Kira (2021). Jane Austen Novels: Word Frequency and Sentiment Analysis. *RPubs*. <https://rpubs.com/kmkhoriuchi/849274>.
- HUGGING FACE. (2023). *Hugging Face*. <https://huggingface.co/hugging-face>.

- ISASI, Jennifer (2021). Análisis de sentimientos en R con «Syuzhet». *The Programming Historian en español*, 5. <https://doi.org/10.46430/phes0051>.
- JOCKERS, Matthew L. (2020). *Introduction to the Syuzhet Package*. The Comprehensive R Archive Network. <https://cran.r-project.org/web/packages/Syuzhet/vignettes/Syuzhet-vignette.html>.
- JORDÁN ENAMORADO, Miguel Ángel (2017). *Análisis del estilo literario de Jane Austen*. [Tesis doctoral, Universitat de València]. Disponible en: <http://hdl.handle.net/10550/59048>.
- LIU, Bing (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers. DOI: <https://doi.org/10.2200/S00416ED-1V01Y201204HLT016>.
- LUU, Chi (2018). *Jane Austen's Subtly Subversive Linguistics*. JSTOR Daily. <https://daily.jstor.org/jane-austens-subtly-subversive-linguistics/>.
- McKENZIE, Paige (11 de enero de 2018). Exploring Jane Austen novels with text analysis. *Paige McKenzie (blog)*. <https://p-mckenzie.github.io/2018/01/11/Jane-Austen/>.
- MOHAMMAD, Saif M. y Peter D. TURNEY (2012). Crowdsourcing a Word-Emotion Association Lexicon. *Computational intelligence*, 29 (3), 436-465. <https://arxiv.org/pdf/1308.6297.pdf>.
- MOURE, Teresa y Joaquim LLISTERRI, (1996). Lenguaje y nuevas tecnologías: El campo de la lingüística computacional. En Milagros Fernández Pérez (Ed.), *Avances en lingüística aplicada* (147-228). Universidade de Santiago de Compostela.
- NEUWIRTH, Erich (2022). Package «RColorBrewer». *The Comprehensive R Archive Network*. <https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf>.
- OWEN, David (2018). The European Reception of Jane Austen's Works. *Quaderns: revista de traducció*, 25, 15-27.
- PAGE, Norman (2011). *The Language of Jane Austen (Routledge Revivals)*. Londres: Routledge.
- PASCUAL, Federico (2 de febrero de 2022). Getting Started with Sentiment Analysis using Python. *Hugging Face*. <https://huggingface.co/blog/sentiment-analysis-python>.

- PÉREZ, Juan Manuel, *et al.* (2023). pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks. *arXiv (Cornell University)*. DOI: <https://doi.org/10.48550/arXiv.2106.09462>.
- Project Gutenberg (s.f.). *Project Gutenberg*. <https://www.gutenberg.org/>.
- REDONDO FIGUERO, Carlos G. (2017). *El programa R, herramienta clave en investigación*. Santander: Editorial de la Universidad de Cantabria.
- SILGE, Julia y David ROBINSON (2016). tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *Journal of Open Source Software*, 1 (3), 37. DOI: <http://dx.doi.org/10.21105/joss.00037>.
- SWAFFORD, Joanna (2015). Problems with the Syuzhet Package. *Anglophile in Academia: Annie Swafford's Blog (blog)*. <https://annieswafford.wordpress.com/2015/03/02/Syuzhet/>.
- WALDRON, Mary (1999). *Jane Austen and the Fiction of her Time*. Cambridge: Cambridge University Press.
- WICKHAM, Hadley, *et al.* (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4 (43), 1686. DOI: <https://doi.org/10.21105/joss.01686>.
- WYETT, Jodi L. (2015). Female Quixotism Refashioned: Northanger Abbey, the Engaged Reader, and the Woman Writer. *The Eighteenth Century*, 56 (2), 261-276.

ANÁLISIS CUANTITATIVO PRELIMINAR DE LA TRADUCCIÓN INGLESA DE LA NOVELA EL DIARIO DE SATÁN, DE LEONID ANDREEV

Patricia González Almarcha

Resumen

El presente artículo emprende, desde parámetros digitales, un análisis exploratorio de la versión inglesa del opúsculo ruso del escritor modernista Leonid Andreev, *El diario de Satán*. Se seleccionan, con el criterio de análisis temáticos y semánticos previos en torno a la significación de la figura del diablo en esta novela, los términos cuya frecuencia y distribución serán rastreadas computacionalmente en el corpus. Se utilizan para las mediciones herramientas informáticas como R, Sketch Engine o Voyant Tools, a fin de verificar una serie de hipótesis circulantes en torno a la temática y semántica del texto. Los resultados cuantitativos perfilan un sistema de trama artística que opera en el “juego de espejos”, según el cual el propio jugador, Satán, se halla jugando manipulado por otro jugador de un ulterior “gran juego”, que, por lo general, detenta el estatus de “modelo ontológico del mundo”, el ser humano, en una inversión del mito fáustico.

Palabras clave: Leonid Andreev, Satán/Satanás, mito, análisis cuantitativo, literatura rusa.

Patricia González Almarcha
Universidad Complutense de Madrid-ITEM
patgon22@ucm.es
ORCID: 0009-0007-0192-6007

Abstract

This article undertakes, from digital parameters, an exploratory analysis of the English version of the Russian work by the modernist writer Leonid Andreev, *Satan's Diary*. They are selected, with the criteria of prior thematic and semantic analysis around the significance of the figure of the devil in this novel, the terms of which frequency and distribution will be tracked computationally in the corpus. They are used for measurements computer tools such as R, Sketch Engine or Voyant Tools, in order to verify a series of circulating hypotheses regarding the theme and semantics of the text. Quantitative results outline a system of artistic plot that operates in the “game of mirrors”, according to which the player, Satan, finds himself playing and being manipulated by another player of a later “great game”, which, in general, holds the status of “ontological model of the world”, the human being, in an inversion of the Faustian myth.

Keywords: Leonid Andreev, Satan, myth, quantitative analysis, Russian literature.

1. INTRODUCCIÓN

Los procesos creativos del escritor ruso del siglo XX, Leonid Andreev, constituyen uno de los fenómenos más complejos en la historia de la literatura rusa. La inconsistencia en sus principios ideológicos y estéticos, sus conceptos filosóficos, la síntesis y transformación específicas de los temas e imágenes universales de la literatura mundial despiertan el interés de los investigadores en su patrimonio artístico literario. En este sentido, la emergencia de las Humanidades digitales permite abordajes novedosos de su creación literaria. Son de notar, en el espectro ruso y en el campo de las humanidades tradicionales, los trabajos de investigación de Achátova, Guseva (Нечепова, 2019; Nikiforová y Mekhasiuk 2018, p. 127). Asimismo, se ha trabajado desde la correlación de las formas de crear la imagen del diablo en las novelas *El diario de Satán* de Andreev y *El maestro y Margarita* de Bulgákov, por ejemplo, en la investigación de Ковалева (2014), que compara las imágenes del diablo en las obras antes mencionadas, desde una posición litera-

ria. Sin embargo, desde el punto de vista lingüístico computacional, estas imágenes se hallan poco investigadas.

En cualquier caso, la interpretación del diablo en la cultura siempre ha gozado de atención, aunque hayan ido variando los términos y formas de tratar esta temática. De ahí que la crítica se pregunte si han sido sólo los motivos religiosos los que han impulsado a escritores como Andreev a incluir esta imagen en sus obras. Este trabajo de investigación persigue abundar en la terminología empleada por el novelista para resignificar la leyenda. Ahora bien, como novedad, se propone abordar la traducción inglesa del opúsculo, para sentar las bases de futuras investigaciones orientadas a cuestionarse hasta qué punto hay una dependencia del idioma específico para trasladar ciertas nociones (Lozano Zahonero, 2020). La originalidad del género de la novela datada en 1919 está asociada con matices mito poéticos. Según diversos investigadores (Московкина 2005; Каманина 2024), *El diario de Satán* pertenece al género de la llamada novela neomitológica. Московкина (2005) señala que la base de la novela es “la interpretación original de la historia de Cristo el Anticristo”. Каманина habla de una síntesis de los mitos en torno a Satanás (Goethe) con los de la literatura rusa de principios de siglo y la Edad de Plata: el Superhombre (Dostoyevsky), el Anticristo (Merezhkovsky), la belleza caída Blok (1963), Cristo encarnado (mito gnóstico simbolista) (Каманина, 2024).

1.1. OBJETIVOS Y METODOLOGÍA

El objetivo general de esta aportación consiste en determinar, de un modo preliminar y desde parámetros digitales, la configuración propia y original de la imagen satánica en la versión inglesa de *El diario de Satán*, de Leonid Andreev¹, a fin de profundizar en presuposiciones temáticas acerca de una novela que gira hipotéticamente en torno al diablo encarnado. Los objetivos específicos

1 El texto en inglés se extrae de la plataforma Project Gutenberg (Andreev, 2024).

pasan, en este estadio iniciático del estudio, por identificar palabras frecuentes e infrecuentes mediante herramientas computacionales disponibles en el entorno de programación R (2020), a fin de verificar o refutar las hipótesis de partida. En la metodología se sigue una ruta mixta, que combina el análisis cualitativo del comparatismo con elementos de la crítica mitológica y con la estilometría computacional (Jockers, 2014). Los resultados esperables han de reflejar la frecuencia léxica y la distribución verbal de conceptos fundamentales en la identificación de etapas en la encarnación de Satán, y han de contribuir a sostener varias vertientes interpretativas en la conformación del mito, como la que juzga la obra a la manera de trasunto del mal social (Грушина и Журчева, 2022), en tanto que lucha de opuestos tanatológicos y eróticos o como trayecto de transformación (Kolmakova, 2019) y dualidad o especularidad entre el yo y el tú (Bushmina, 2013), e incluso actualización o subversión de mitos indoeuropeos (Косенко, 2001).

En concreto, se utilizan las Humanidades digitales en la problemática de identificar los componentes semióticos del “neomito” acerca de “Satanás hecho humano”, apelando, entre otros, al modelo metamórfico interdisciplinario que se sirve de herramientas del método existencial, en continuidad con estudios como el de Гарипова (2020). Además del elemento digital, otra de las innovaciones de la contribución viene determinada precisamente por el análisis de la semiótica anidada en los conceptos “Hombre/Persona/Ser humano-Satán”, en réplica del ámbito contextual de la “corporeidad ontológica viva”, tal como se construye en el sistema del modelo metamórfico interdisciplinario de Zaltman, cuya técnica, según Гарипова (2020, p. 7) se basa en la extracción del inconsciente de imágenes metafóricas que influyen en la comprensión semántica de la información entrante. Con acuerdo a Гарипова (2020, p. 7). Se escogen las concepciones lingüísticas y culturales “Hombre/Persona/Ser humano-Satán” como símbolos clave de un modelo metamórfico “Satanás encarnado”, que refleja las categorías y actitudes más importantes de la visión antropológica y religioso filosófica de Andreev.

Según investigadores como Косенко (2001, p. 1), la novela *El diario de Satán* no goza de tanta popularidad como otras obras de Andreev. En todo caso, proyecta una imagen especular del estado de ánimo de toda una época en una original forma artística. El autor se refiere al motivo universal de la tragedia de Goethe: el pacto entre el hombre y el diablo. Estos dos son los personajes principales de la novela. Sendas fuerzas conforman dos polos que chocan e interactúan para resolver el eterno interrogante: de qué es capaz una persona, cuáles son sus capacidades. Ahora bien, y según diversos autores, los análisis en términos del léxico poco frecuente son novedosos y están justificados por la evolución de técnicas, entre otras, como la estilometría aplicada a la atribución autoría (Faltýnek y Matlach, 2020), de ahí que se rastree ocasionalmente el peso de lexemas inexistentes o infrecuentes como *betrayal*, *sacrifice*, *incarnated*, *fallen* y *angel* (al albur de los análisis temático semánticos de Косенко, 2001 y de Гарипова, 2020).

1.2. PROCEDIMIENTO

En relación con ello, el criterio para la selección de términos que se prevén exiguos y no se hallan finalmente en el cómputo mecánico, como *sacrifice* o *betrayal*, más allá de la significativa ausencia denotativa o referencial, radica en varios aspectos: por un lado, la ambivalencia moral de la personalidad del héroe, quien, a través del sacrificio y la traición simboliza la altura y la caída éticas. Léxicamente el descenso y la deslealtad se explicitan en la raíz de la unidad terminológica *lie*, mucho más frecuente y connotativa de la abnegación o la apostasía. Por otro lado, adjetivos como *incarnated* se rastrean en tanto síntomas de adquisición por parte del demonio de cualidades y emociones humanas, lo que se recoge en los monólogos interiores y en el uso del tiempo presente, pero no en la adjetivación aludida. En su lugar, la tierra como receptáculo de la encarnación satánica, es un término, por el contrario, frecuente. A medida que Satán penetra más en el mundo humano, en la esencia de la vida, más aparentemente asquerosos se le vuelven sus ideales. Por último, la colocación semántica

fallen angel sí se analiza como rareza y a la luz de razones similares: el desarrollo paradójico de una trama donde el protagonista y su séquito de personalidades aparecen como gente común, al tiempo que retienen sus propiedades sobrenaturales, sus inclinaciones demoníacas e infernales, su omnipotencia y omnisciencia.

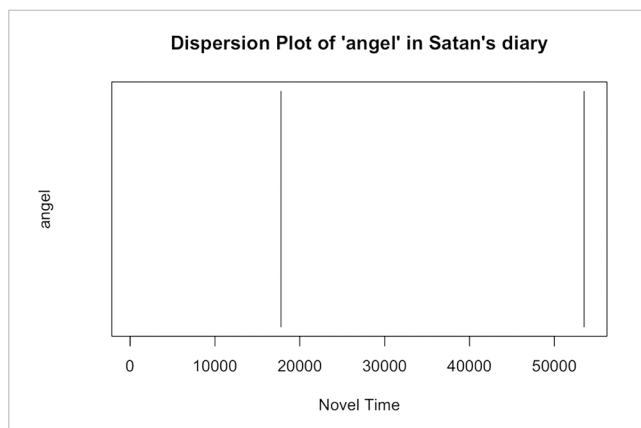


Figura 1. Distribución de la palabra *angel* en el corpus.

Fuente: elaboración propia²

En contacto con el adjetivo “alado” (*winged*), Satán aparece como símbolo de lo antitético, al frente de un automóvil imposible, lento como una tortuga. El cuestionamiento de la velocidad y la interrelación del vuelo con otros animales como el pez abundan en el carácter contradictorio de la fisionomía satánica³.

- 2 Esta y el resto de las figuras que se recogen en el trabajo son de elaboración propia y han sido realizadas con la aplicación R (2020). Se pueden obtener asimismo visualizaciones idénticas de la distribución con la herramienta Sketch Engine (Kilgarrieff *et al.*, 2014).
- 3 Posición 1 (según Sketch Engine, Kilgarrieff *et al.*, 2014): *I shall remain a schooner with sails lowered.* </s><s> *I shall be _All_ and I shall be _Nothing_.* </s><s> *Only once did a slight breeze stir my sails, but only for a moment: that was when Maria pointed out the Tiberian road to me, cutting the green hills like a white thread, and asked whether I had ever traversed it before.* </s><s> “Yes, occasionally, Signorina.” </s><s> “I often gaze upon this road and think that it must be extremely pleasant to traverse it by automobile.” </s><s> “Have you a swift car,

En relación con el epíteto “caído” (*fallen*), apelativo que le da el cardenal, la representación satánica recupera la tradición bíblica y pone en duda su nombre, evaluándolo como gesto artístico. De esta manera, la escasa frecuencia del “ángel”, sea cual fuere el contexto lingüístico de aparición, podría coadyuvar en la tesis de una percepción diabólica más escorada a su humanidad que a su rasgo infernal. En este sentido, la aparición especialmente baja del vocablo *humanity* (“humanidad”) puede interpretarse alineada como un rechazo de la designación directa del humanismo en una condena de una sociedad inhumana, o incluso en tanto que ironía predictiva de la falta de humanidad de los pobladores de la tierra, sean nominados o innominados⁴.

Signor?” </s><s> “Oh, yes, Signorina, very swift! </s><s> But those,” I continued in gentle reproach, “who are themselves limitless distances and endlessness are in no need of any movement.” </s><s> Maria and an automobile! </s><s> A winged angel entering a trolley car for the sake of speed! </s><s> A swallow riding on a turtle! </s><s> An arrow on the humpy back of a hod carrier! </s><s> Ah, all comparisons lie: why speak of swallows and arrows, why speak of any movement for Maria, who embraces all distances! </s><s> But it is only now that I thought of the trolley and the turtle. </s><s> At that time I felt so calm and peaceful, I was deep in such bliss that I could think of nothing except that countenance of eternity and undying light! </s><s> A great calm came upon Me on that day and nothing could disturb my endless bliss. </s><s> It was not long before Thomas Magnus returned, and a flying fish, gleaming for a moment above the ocean, could no more disturb its blue smoothness than did Magnus disturb.

4

Posición 2 (según Sketch Engine, Kilgariff et al., 2014).

<s> He said he is Satan. </s><s> Throw him out, Signor Magnus. </s><s> This is sacrilege!” </s><s> “Is that so?” </s><s> smiled Magnus politely: “I did not know that Satan, too, belonged to the heavenly chair...” </s><s> “Satan is a fallen angel,” said the Cardinal in an instructive tone. </s><s> “And as such he is in your service? </s><s> I understand,” Magnus bowed his head politely in acceptance of this truth and turned smilingly to me: “Do you hear, Wondergood? </s><s> His Eminence is irritated by your audacity.” </s><s> I was silent. </s><s> Magnus winked at me slyly and continued with an air of artificial importance: “I believe, Your Eminence, that there must be some sort of misunderstanding here. </s><s> I know the modesty and well-informed mind of Mr. Wondergood and I suppose that he utilized the name of Satan merely as an artistic gesture. </s><s> Does Satan ever threaten people with the police? </s><s> But my unfortunate friend did. </s><s> And, in general, has anybody ever seen _such_ a Satan?” </s><s> He stretched his hand.

La traición, el sacrificio, la encarnación, la caída angelical no dominan la trama de una forma manifiesta al menos, lo que redunda en el valor de las implicaturas en ciertas nociones vinculadas con la carnalidad y la espiritualidad. Se prefiere implicar estas nociones por medio de réplicas de los personajes en que se desdobra: el personaje desdoblado Magnus reprocha a Satán que ha llegado (encarnado) tarde, que la tierra ha crecido y ya no necesita de sus talentos. También se aprecia en los monólogos interiores de otro de los desdoblamientos diabólicos, Wondergood, quien admite una tarde sentirse plácidamente humano, para añadir un comentario cáustico sobre la mujer y un vestido confeccionado a su medida y acabar posicionándose en el limbo entre dos conciencias, al reconocer que era sensible, bueno y amable, que sólo quería jugar, pero se había inclinado hacia esa pesada tragedia.

Tabla 1. Número de ocurrencias y aparición de términos relacionados con el ángel caído

PALABRA	NÚMERO DE OCURRENCIAS	PESO RELATIVO
Betrayal ⁵	0	0
Lie	23	0.0004256343
Sacrifice	3	5,551752e-05
Incarnated	0	0
Fallen	3	5,551752e-05
Angel	2	3, 701168e-05
Heaven	8	0.0001480467

Fuente: Elaboración propia

El discurso en dos capas del héroe narrador, Satán, alterna construcciones concisas con otras más complejas, donde pide que no lo piensen como bromista, que ha llegado a sentirse muy mal dentro de Wondergood. Añade que, si por el día logra vencer a su encarnación, sobreponerse, cada noche ella le pone contra las cuerdas. Esta persona en la que habita le envía la oscuridad a los

5 Sí aparece el verbo *betray*, “traicionar”, con una frecuencia de 8 veces, según Sketch Engine (Kilgarriff *et al.*, 2014).

ojos, con sueños tontos, con memorias absurdas. Lo que sueña se vuelve muy ridículo a su juicio, y Satán admite que Wondergood se adueña de él, acorralándolo como un perro.

Precisamente el proceso de personificación de Satanás es lo que anida en la base argumental de la novela y correlaciona con el motivo del camino, que se expresa en imágenes temporales y espaciales que organizan el cronotopo de la forma novelística de diario. Las fechas de esta crónica marcan las etapas de la transformación de Satanás en un hombre que camina su *via crucis*. Esto se refleja en la semántica de los términos de los registros, que coinciden con las principales festividades cristianas fijadas en el calendario ortodoxo. El tiempo en la obra tiene una estrecha relación con las imágenes espaciales (el mar o la isla de Capri, etc.). El código cristiano utilizado por Andreev en la novela no cumple una función religiosa, sino cultural y moral, que denota la posición preexistencialista del escritor, por lo que aparece subvertido. El análisis de la frecuencia y distribución de algunos topónimos en el texto abundaría en esta idea inicial.

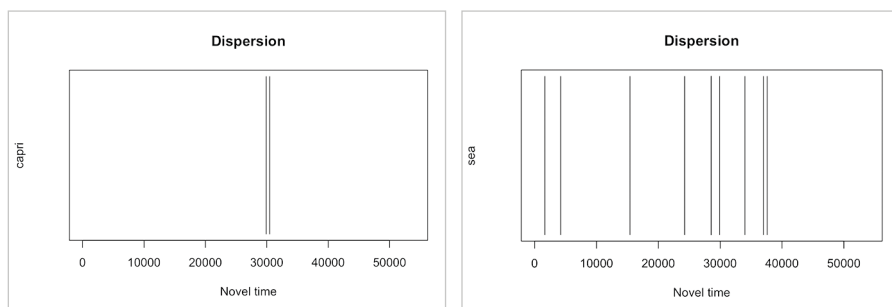


Figura 2. Distribución en el corpus de antropónimos significativos: *sea* y *Capri*. Fuente: elaboración propia

Se presenta una visión bastante única de Satanás, porque el diablo viaja del infierno a la tierra para huir del aburrimiento, encarnando por completo, de modo que los hombres lo reconocen como igual a ellos. Cuanto más tiempo pasa en el planeta, más penetra Satán en la esencia humana. Sin embargo, el mundo

dibujado por Andreev en las personas que lo habitan quiere destruir todo lo que de humano se manifiesta en el ente diabólico de la manera más cruel. El principal problema radicaré, entonces, en definir la percepción que tiene de lo satánico la gente que reside en el planeta, en concreto en Rusia. Estadísticamente importan el número y distribución de términos alusivos a la población, como *people* o *humanity*, ya que trazan un patrón alusivo al componente poblacional como más diabólico que el propio diablo, puesto que oponen, en principio, lo social a lo existencial. En su juego, Satán afirma que no sabe qué pensar sobre un débil y quejica mentiroso que, tembloroso, se coloca en la cabeza la corona de un zar teatral. Dice que le da vergüenza y le pide al ser humano que le dé aquello con lo que alimenta a sus amigos y sus bromas; no hay humanidad, pero sí mucho pueblo, en línea con lo aludido en torno a la errática percepción de lo demoníaco como muy personal, aunque poco humano. La gente corriente se arranca de la civilización o de la abstracción de una cualidad positiva anexada al humanismo:

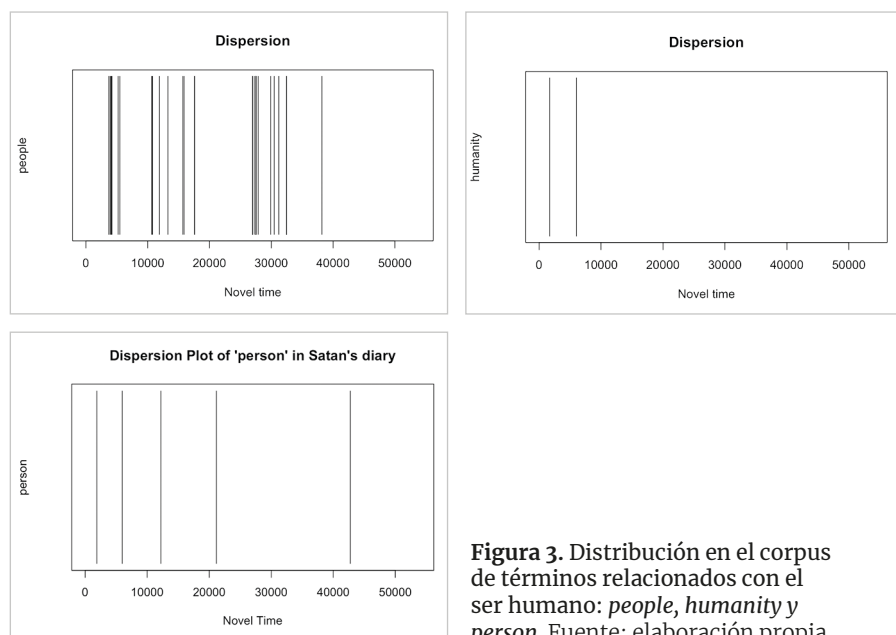


Figura 3. Distribución en el corpus de términos relacionados con el ser humano: *people*, *humanity* y *person*. Fuente: elaboración propia

Tabla 2. Número de ocurrencias y peso relativo de palabras relacionadas con el ser humano

PALABRA	NÚMERO DE OCURRENCIAS	PESO RELATIVO
People	15	0,0002775876
Humanity	2	3,701168e-05
Person	5	9.252919e-05

Fuente: elaboración propia

La primera etapa de la encarnación de Satanás en un ser terráqueo está relacionada con el cambio de su apariencia. Nuestras ideas clásicas sobre la aparición satánica se ven ridiculizadas por el mismo héroe, que adopta características de tramposo, lo que se evidencia en sus máscaras. Habla de sus apariciones en su diario, género literario novelístico de signo autobiográfico, al menos en apariencia. Lo importante estriba en que Satán ha venido a la tierra, obviamente, no como una persona común, no a vivir una vida humana tópica o a ser un representante de las masas. Con todo, acaba por representar una imagen colectiva, convirtiéndose en un símbolo de la burguesía, del mundo capitalista, que está plagado de amenazas para la humanidad y la cultura. Por ridículo que parezca, hasta el mismo diablo manifiesta su desilusión por la vida y las costumbres humanas. Algunos investigadores soviéticos sostienen que “[...] La novela de Andreev *El diario de Satán* se torna la “voz” profética del autor sobre el triunfo global del principio del mal en el orden mundial moderno...” (Hecrepova 2009, 92). La frecuencia relativa de términos relacionados con el universo léxico semántico inglés de la degradación y de la animalización y su distribución lo corroboran: “borracho”, “anticristo”, “fiera” y “gusano”. Se escogen estos lexemas en vínculo con la confesión final de Satán en el diario, quien dice haber llegado a la tierra para jugar y reírse, haberse preparado para todo mal, haber mentido. Sin embargo, se queja de un tú, un gusano peludo, que le ha mordido el corazón, que se ha aprovechado de que tiene un corazón humano y lo ha atacado. De este modo, revela su fiasco, su actuación. Se visualizan estos términos dentro de una misma categoría, en agrupación semántica.

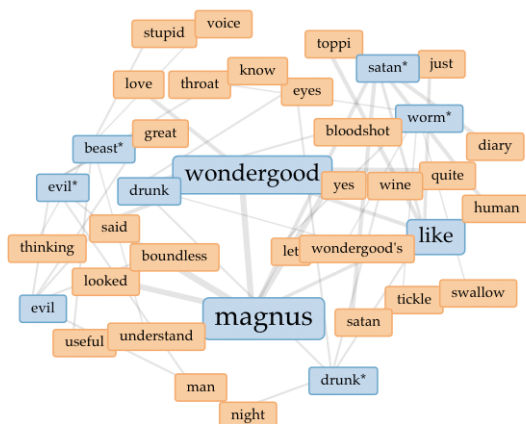


Figura 4. Red semántica de términos relacionados con el diablo y su animalidad. Fuente: elaboración propia⁶

Resulta reseñable la ligazón de la parte bestial (*beast*) del diablo con las nociones de grandeza (*great*), amor (*love*), garganta (*throat*), estupidez (*stupid*), fonación (*voice*), por un lado, y de mal (*evil*), pensamiento (*thinking*), dicción (*said*), vista (*looked*), utilidad (*useful*), comprensión (*understand*), hombre (*man*) y noche (*night*). La faceta fiera del diablo, reenvía, a su vez, a los personajes de Wondergood y Magnus. En lo relativo al lado insectívoro de Satán (*worm*), se alía con la sangre (*bloodshot*), el diario (*diary*), el vino (*wine*), la afirmación (*yes*), lo humano (*human*), así como lo bastante o suficiente (*quite*). La vertiente ebria de lo diabólico se aúna a términos como Wondergood, hacer cosquillas (*tickle*), tragar (*swallow*) y Satán (*satan*), que en conjunto se anexan nuevamente a la borrachera (*drunk*).

Andreev construye su neomito sobre un Satanás tentado por el hombre. De ahí que las significaciones de esta suerte de autoinforme y confesión se vuelven infinitamente inexpresables, como el Hombre, Satanás y Dios mismos. El escritor contempla

6 La figura se ha realizado con la herramienta Voyant Tools (Sinclair y Rockwell 2016).

el mundo como una unidad de principios antitéticos. Este pensamiento se observa en el texto no sólo en el sistema de trama y su estructuración ideológica y temática, sino también a través del modelado léxico semántico de imágenes (es decir, mediante la síntesis de antónimos existenciales: Satanás hecho hombre). En la tríada de la ebriedad, la expulsión y la animalidad, merece la pena destacar la inexistencia del término “anticristo”. Como bien notan los investigadores acerca de la mitología y poética de Andreev, el tema de la encarnación de Satán le da al escritor la posibilidad de mostrar la dualidad de la naturaleza humana. La contradicción interna de Satán-Wondergood refleja una serie de oposiciones características de la esencia humana: apariencia-alma, conciencia-inconsciencia, bien-mal, que aparecen también en otros personajes de la novela: el diablillo secretario Toppi, el cardenal X, la madona-prostituta María y el Satán personificado, Foma Magnus (gemelo, a su manera, de Satán-Wondergood). La ambivalencia es característica del personaje y prepara el terreno para la transformación del mito de la venida del anticristo en un mito sobre la venida de Cristo. El juego mitopoético de Andreev consiste en invertir los temas, imágenes y motivos bíblicos en un proceso por el cual se desmitifican los esquemas tradicionales. El Satán de Andreev, como Cristo, es un ser superior en un contorno humano, y sus sufrimientos se asemejan a los del mesías. En tanto que dios humano, experimenta el nuevo Satán persona: la tentación en la novela se realiza como una instigación inversa, de suerte que el ser humano tienta al diablo.

Cuesta Fernández (2012, p. 383), por su parte, estudia la etimología y aparición del término anticristo desde el Antiguo Testamento, pasando por el Nuevo y hasta llegar a la literatura cristiana primitiva. Se trata de un étimo de origen griego, compuesto por el prefijo “anti” (suplantación y sustitución y oposición), lo que, según Cuesta Fernández, se traduce en “contra Cristo” o “en lugar de Cristo”. La asociación del anticristo con algún personaje histórico específico, en especial los caracterizados por pésimos gobiernos y por modelizar la maldad, resulta frecuente. Andreev se vale

de Nerón entre sus protagonistas. El anticristo no constituye una encarnación de Satanás al final de los tiempos, sino que se analiza como entidad diferente, fingidora de lo mesiánico y ávida de abocar a quienes lo sigan a la destrucción corpórea y anímica. Satanás, por ello, se conforma como el Anti Dios.

Creer en que llega inminentemente el Anticristo supone un escape para pueblos o personas sometidos a la dura opresión de injusticias, hambrunas, enfermedades, etc. Este personaje se interpreta en el ámbito de la escatología como correspondiente a un tiempo futuro por llegar, de modo que, cuando ocurra, en la mentalidad de los que creen en él, finalizarán las edades y la historia humanas. Cuesta Fernández (2012, pp. 383-384) revisa la literatura vetero, novotestamentaria y patrística, a fin de rastrear en los exponentes literarios diversos el tratamiento o la aparición del lexema en cuestión. El anticristo como encarnación del mal se forja en el Antiguo Testamento, si bien no se recoge explícitamente la palabra, sino mediante alusiones o imágenes del mal relativas a la transgresión de la ley o a conductas desordenadas. Aunque dichas imágenes, como puntualiza Cuesta Fernández (2012, p. 383) no han de tomarse necesariamente como prefiguraciones oficiales del anticristo, sí son reinterpretadas como tal por los Padres de la Iglesia.

Ya en el Nuevo Testamento surge la primera referencia explícita al anticristo como personaje escatológico, pero sin vinculación con ninguna personalidad histórica. El término es bastante genérico, ya que designa a todos los que cuestionan la divinidad de Cristo. El autor del pasaje, que se hace llamar Juan, avisa a los miembros de la comunidad de la existencia de muchos que se hacen pasar por profetas. Esta temática reaparece en la carta segunda a los tesalonicenses, atribuida a Pablo de Tarso, donde el autor da testimonio de las creencias e inquietudes de los cristianos primitivos, creyentes en la inminencia de la segunda venida de Jesucristo. Las expresiones “hijo de la iniquidad” o “hijo de la perdición”, que se suman a la cuarta bestia del libro de Daniel o a las dos bestias del Apocalipsis, se yerguen como símbolos muy utilizados para refe-

rirse al anticristo (Cuesta Fernández 2012, p. 384). En la Patrística, el fenómeno se asemeja a la literatura veterotestamentaria y no se recogen menciones explícitas al anticristo, pero sí a figuras y realidades cercanas a él. Andreev maneja diversas fuentes donde la alusión al anticristo es velada y se encubre en seres históricos como Nerón o animalizaciones como bestia. Ello explica la exigüidad de la unidad terminológica “anticristo”.

No en vano, uno de los principales síntomas de la era del romántico alemán Goethe reside en la creencia en la capacidad del hombre para cambiar el mundo con la ayuda de la razón, de acuerdo con las leyes morales. Y el comienzo del siglo XX se conforma precisamente como una época en la que la decepción ante los ideales humanistas y el poder virtuoso de la razón se convierte en el rasgo dominante de la conciencia pública. Por tanto, los análisis computacionales se tornan interesantes en la confirmación de las transformaciones de las imágenes de Goethe en la literatura de la época, ya que reflejan plenamente todos los rasgos de la cosmovisión pública del momento.

Tabla 3. Número de ocurrencias y peso relativo de palabras relacionadas con el diablo y su animalidad

PALABRA	NÚMERO DE OCURENCIAS	PESO RELATIVO
Drunk	4	7.402335e-05
Antichrist	0	0
Anti	1	1.850584e-05
Christ	6	0.000111035
Devil	37	0.000684716
Satan	63	0.001165868
Beast	3	5.551752e-05
Worm	8	0.0001480467

Fuente: elaboración propia

El motivo de la transgresión de la frontera está asociado al fenómeno del espacio de lo “inefable”, de donde proviene el héroe de la novela. Sus principales características son la abstracción,

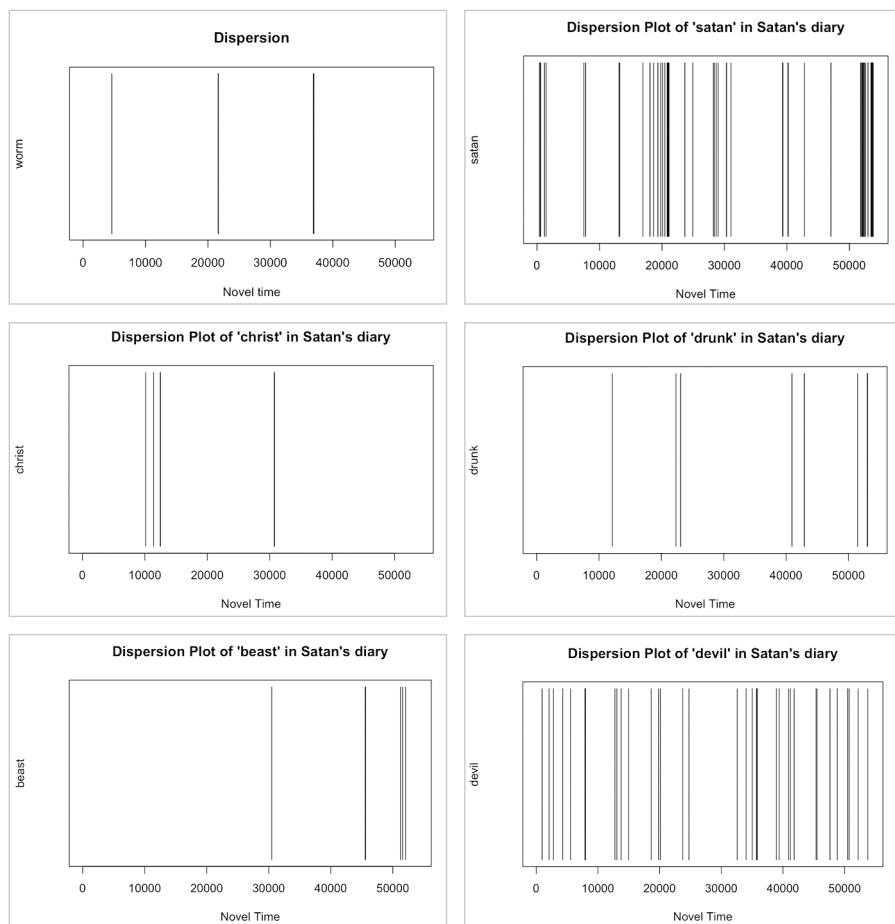


Figura 5. Distribución en el corpus de términos relacionados con el diablo y su animalidad. Fuente: elaboración propia

la metáfora, la imposibilidad de expresarse a través de una palabra humana, en línea con el misticismo renacentista. El mundo de armonía y eternidad que ha dejado Satán, está marcado en la obra con la imagen-símbolo de una goleta en medio del océano. El océano es el umbral entre lo infinito y lo finito, la barca es la oportunidad de superar las limitaciones de la realidad terrena, cuyas leyes actúan sobre el Diablo hecho hombre. Su recuerdo del mun-

do de lo “extraordinario” se desvanece lentamente. La distribución⁷ del vocablo inglés para “goleta” (*schooner*) ahonda en esta percepción.

Tabla 4. Número de ocurrencias y peso relativo de palabras relacionadas con el transporte

PALABRA	NÚMERO DE OCURENCIAS	PESO RELATIVO
Schooner	7	0.0001295409
Train	2	3.701168e-05
Carriage	2	3.701168e-05

Fuente: elaboración propia

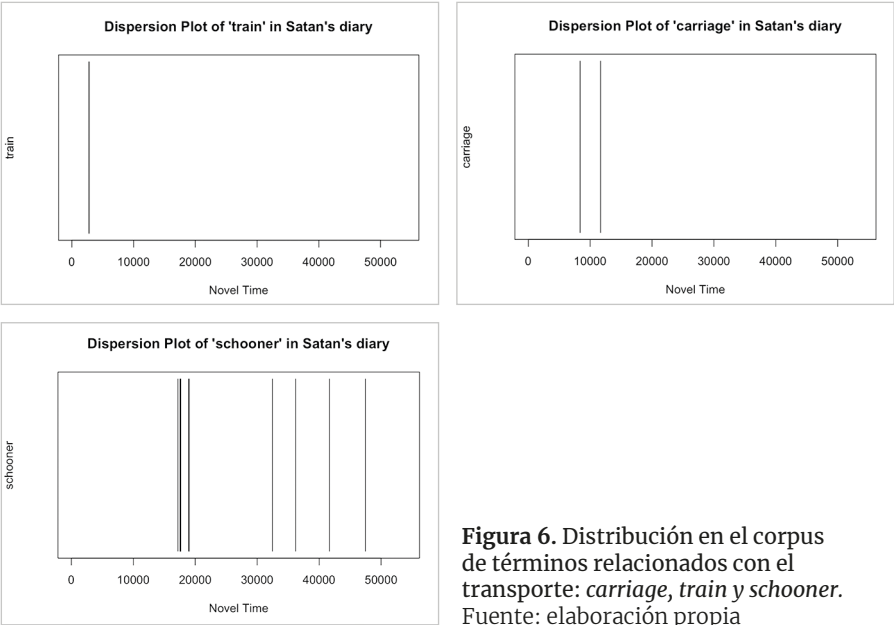


Figura 6. Distribución en el corpus de términos relacionados con el transporte: *carriage*, *train* y *schooner*. Fuente: elaboración propia

7

En estadística el concepto de dispersión es una medida que permite saber si la muestra tomada está agrupada alrededor de la media o muy desagrupada. En según qué contextos puede dar lugar a confusión, de ahí que se opte por la noción de distribución, que describe cómo se distribuyen los valores para un campo, o qué valores son comunes y poco comunes (IBM, 2024).

Por otra parte, y según Косенко (2001), la especificidad de los personajes principales de Andreev está determinada por su correlación con los protagonistas del drama de Goethe, *Fausto*, obra que a principios del siglo XX será redescubierta por los simbolistas y se verá inmersa en el centro de sus disputas literarias y morales-filosóficas. *El diario de Satán* participa de esta controversia, que supone (re)enfrentarse con la trama de Goethe y reconsiderarla significativamente desde el punto de vista de la modernidad (Московкина, 1994, p. 3). Andreev transforma con originalidad las imágenes universales de Fausto y Mefistófeles en las figuras de Satanás como Wondergood y como Thomas Magnus. La frecuencia de términos que etimológicamente van cargados de sentido: el maravilloso bien y la magnanimidad gemelar sostienen esta interpretación. Lo irónico y lo especular intervienen en estos dos nombres transparentes en su simbología, con una elevada distribución en el conjunto de la fábula, de donde se infiere una

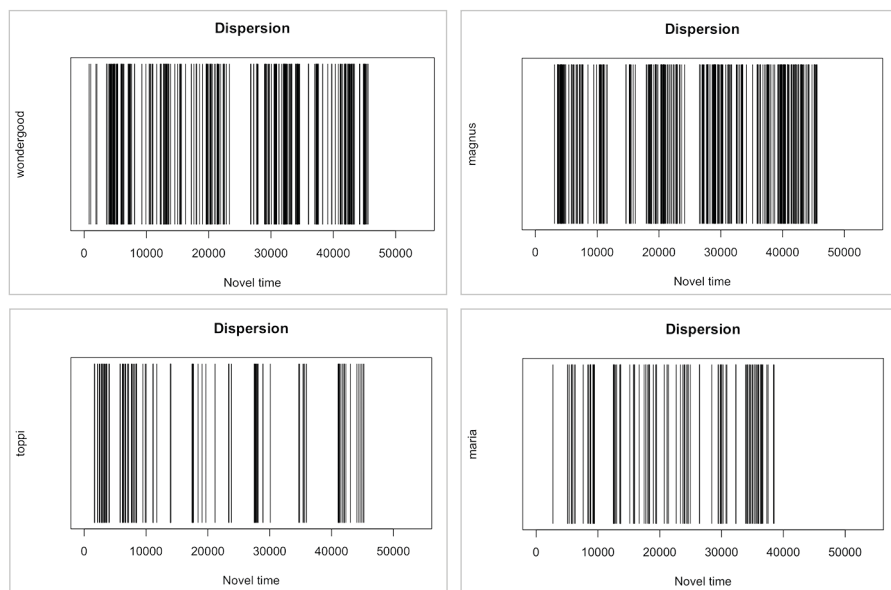


Figura 7. Distribución en el corpus de términos relacionados con las encarnaciones humanas de Satán y sus contrapuntos: *Wondergood*, *Magnus*, *Toppi* y *Maria*. Fuente: elaboración propia

Se recupera ahora, con ayuda de R, la terminología desprovista de significado léxico, a fin de evaluar la relevancia de los pronombres personales, que funcionan en especularidad y en alimento de la supuesta dicotomía entre lo humano y lo satánico o de su aparente fusión.

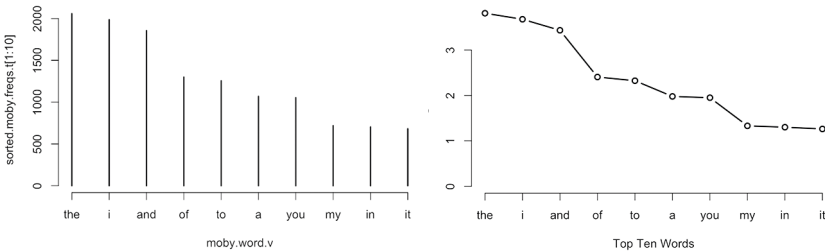


Figura 9. Términos más frecuentes sin excluir las palabras vacías de significado. Fuente: elaboración propia

Tabla 5. Número de ocurrencias y peso relativo de palabras relacionadas con lo pronominal y lo personal

PALABRA	NÚMERO DE OCURRENCIAS	PESO RELATIVO
I	1987	0.0367711
You	1054	0.01950515
They	151	0.002794382
My	719	0.0133057
He	452	0.008364639
She	73	0.001350926
Your	317	0.005866351
Their	61	0.001128856

Fuente: elaboración propia

Se infiere un predominio del “yo” sobre el “tú” y el “ellos”. Ahora bien, sea en su forma demostrativa, sea en la posesiva, el “ellos” queda más explícita o sensiblemente separado en la cuantía de apariciones de la primera y la segunda persona, de donde se deduce un refuerzo del juego de la interpelación entre lo satánico (el “yo”) y lo encarnado (el “tú”), frente a la colectividad

(el “ellos”). El desdoblamiento en un interlocutor o receptor que se describe como un tú para tomar distancia apreciativa se recalca en esta cifra pronominal bimembre.

Si se prescinde de preposiciones, conjunciones y artículos, en clave del lenguaje, el mencionado juego entre el yo y el tú se manifiesta al enfatizar la personalidad del narrador con la ayuda de un indicador de tratamiento formal, que hace que Satán se llame a sí mismo un pronombre personal “yo” con mayúscula (además de los pronombres posesivos correspondientes), en tanto que oposición egocéntrica al “tú” (representado también en los pronombres personales o posesivos del tipo de “ellos”, “tuyos”, etc.): “Estoy molesto por el reflejo plano de ti que he provocado en este escenario” (Andreev, 1989). En el futuro, el narrador-héroe (con el sentimiento de su esencia satánica que aún no se ha apartado de

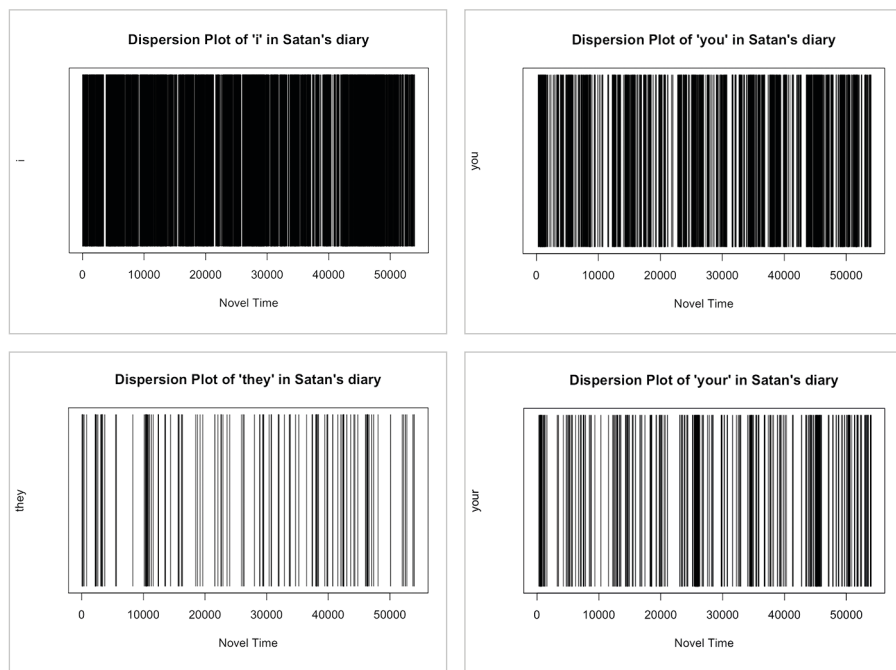


Figura 10. Distribución de los pronombres significativos en el texto: *I, you, they, your*. Fuente: elaboración propia

él) admitirá que en determinados momentos se siente desdoblado en Wondergood.

El camino de Satán-Wondergood simboliza la senda de cada persona en la novela, lo que posibilita acercarse al mito de San Andrés en torno a un hombre que el novelista habría formado al final de su vida: se trata de una persona sufriente y solitaria que no tiene noción de los límites ni orientaciones de los valores (la no presencia de Dios y el Diablo en el espacio terrenal y más allá), y que, debido a las restricciones de su naturaleza, no puede lograr la armonía a través de un avance hacia la eternidad. La pérdida de los puntos de referencia (Dios, el Diablo como leyes morales, nociones del bien y del mal) conduce, por un lado, a la libertad ilimitada, y, por otro lado, a la destrucción de la humanidad en el intento de determinar los límites de lo que está permitido a través de un experimento de apertura perspectivista.

En este sentido, el nombre Tomás nos remite al apóstol que dudaba de la resurrección del Maestro. Thomas Magnus no acepta las enseñanzas de Cristo, no pasa la prueba de su experiencia vital: solo ve decepción en las personas. María para Satanás era algo tan inexpresable como él mismo, para él ella combinó en su imagen todo lo divino y lo sobrenatural a la vez: “Era más que hermosa, era terrible en su perfecta belleza”. María no abandona los pensamientos de Satanás, a pesar de que por momentos se vea estigmatizada por sus dobles, tildada de mentirosa y amante del dinero.

En todos los personajes se van revelando las etapas de la encarnación de Satanás en humano y se va vislumbrando que el mundo creado por Andreev en el diario se nos antoja como un cosmos invertido, donde el mismo Satanás, el malvado tentador de la Biblia que paraliza el destino de las personas y las obliga a cometer pecados, se vuelve esta vez hombre. Dicho proceso coincide, precisamente, con la opuesta deshumanización de las personas; en el alma de cada persona, Satanás ha triunfado hace largo tiempo. El peso en la historia de encarnaciones como la de Nerón, César, Augusto, el gladiador, Savonarola y María lo prueban y sirven para

fundamentar las herramientas lingüísticas utilizadas por Andreev en la recreación de épocas tan diversas como la Roma antigua o el Renacimiento, puesto que estas máscaras especulares reflejan, según la crítica literaria, un cambio en la conciencia del héroe. ¿Y qué hay del carnaval y de la mascarada en todo ello? Se examina la relevancia de la voz “máscara”, en el afán de refutar o reforzar el cariz carnavalesco y de disfraz del personaje estudiado y tan ampliamente desdoblado en otras múltiples alteridades, con una incidencia relativa comentable.

Tabla 6. Número de ocurrencias y peso relativo de palabras relacionadas con la teatralidad y la mascarada

PALABRA	NÚMERO DE OCURRENCIAS	PESO RELATIVO
Mask	4	7.402335e-05
Theater	3	5.551752e-05
Mirror	3	5.551752e-05

Fuente: elaboración propia



Figura 11. Distribución de los términos relacionados con la teatralidad y la mascarada en el corpus: *mask*, *theater* y *mirror*. Fuente: elaboración propia

Según Nikiforova y Mekhasiuk (2018, p. 217), se trataría de un fenómeno de personalidad dividida que opera, en la situación de la novela, en la dualidad del propio narrador con otros personajes principales: Magnus, Toppie, María e incluso el Cardenal X, como también queda de manifiesto en la versión inglesa. El último y decisivo paso para encarnar a Satanás es el reconocimiento del hombre como igual a sí mismo. El héroe ya no escribe acerca de él en el diario con mayúscula. Satanás se vuelve consciente de la vil condición de la vida humana. Pero él acepta esta posición. Satanás pierde por completo todas sus máscaras, a diferencia de Magnus y María. La principal prueba y decepción de Satanás como persona está por venir: comprende que Magnus y María lo han derrotado en su propio juego. Satán y Magnus intercambian, así, sus posiciones. Observamos la victoria final sobre Satanás encarnado en la última escena de la novela: la irónica expulsión de Satanás.

En definitiva, los investigadores claman que la historia o trama en la obra viene conducida por “Satanás encarnado” (Исаев, 2001). Satanás, que se ha convertido en hombre y, por lo tanto, ha perdido su poder, adquiere una ventaja sobre los mortales, como se reseña en las propias palabras del personaje: “Yo soy Satanás... Pero poco puedo contarte sobre mi encarnación... En primer lugar, olvídate de tu peludo y cornudo favorito y de los diablos alados que escupen fuego y convierten fragmentos de arcilla en oro, y transforman a los viejos en jóvenes seductores, y habiendo hecho todo esto y charlado un montón de tonterías, caen instantáneamente a través del escenario” (Andreev 1989, p. 313). Hay una asociación con Mefistófeles, un rasgo “teatral”, capaz tanto de enriquecer como de rejuvenecer a Fausto. Más tarde, el Satán de Andreev indicará directamente su parentesco con Mefistófeles: “He visto en la Escala de Milán cómo mi colega Mefisto se retorció y saltaba por todo el escenario [...]” (Andreev 1989, p. 373). El narrador forma, entonces, el tejido de la historia basándose en el concepto en sí que tiene de la novela: un juego que proporciona diversas manifestaciones de la mente. La incidencia sutil de la interpretación del juego como teatro en la traducción inglesa de *El*

Tabla 7. Número de ocurrencias y peso relativo de palabras relacionadas con la espacialidad física y figurada

PALABRA	NÚMERO DE OCURRENCIAS	PESO RELATIVO
World	17	0.0003145993
Society	2	3.701168e-05
Universe	3	5.551752e-05
Earth	53	0.0009808094
Planet	5	9.252919e-05

Fuente: elaboración propia

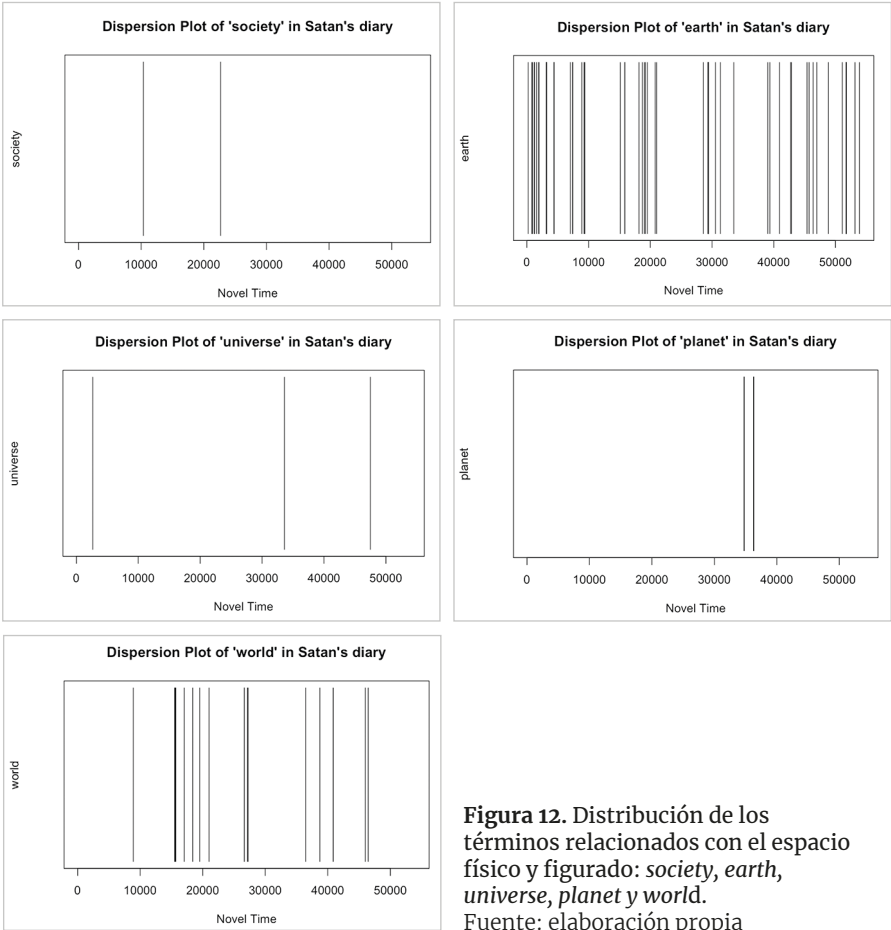


Figura 12. Distribución de los términos relacionados con el espacio físico y figurado: *society*, *earth*, *universe*, *planet* y *world*.
Fuente: elaboración propia

diario de Satán puede apoyar en algún sentido esta tesis, donde lo mundano y lo teatral aparecen representados en la ópera de Milán e implicados en la ideología barroca del gran teatro del mundo o del mundo como sueño.

Los mitos subyacentes señalados con anterioridad provocan la aparición del motivo de la dualidad. Según Тодоров (1997), en una obra de fantasía, “el doble es sólo una parte de la personalidad del héroe, una especie de encarnación de su conciencia”. El mito de la novela, ahora con acuerdo a la tipología de Миронов (2010, pp. 903-906), contribuye al hecho de que el personaje entra en lucha con el antagonista y un mundo ajeno o extranjero por su propia voluntad. Tratando de echar raíces en ese universo extraño, se cambia a sí mismo, pero no logra transformar el entorno en sí. La cantidad de veces que se recuentan vocablos como mundo, sea propio o sea ajeno, incide en este tipo de conclusiones.

No en vano, el escritor nota la fragilidad, la incertidumbre de la realidad terrenal en la que vive la persona moderna de su tiempo: los eventos de la novela se desarrollan en 1914, en vísperas de la Primera Guerra Mundial. La confianza en el contexto socio-histórico de un punto de inflexión en la era de la historia de la humanidad da a Andreev la oportunidad de mostrar el cambio de paradigmas culturales, y como consecuencia de las leyes morales en las que la humanidad se ha basado durante siglos. Los datos estadísticos certifican que la narración se lleva a cabo en nombre del Satanás encarnado. La figura de Satán sirve para plantear la cuestión de la existencia humana desde la posición de la eternidad. Al representar la conciencia de tal sujeto de visión, que vino de otro mundo trascendente que no tiene nada que ver con la realidad terrenal, se logra una novedad de una mirada a la esencia del hombre y la realidad que lo rodea. Basado en la tradición literaria cultural general, Andreev crea su propia imagen del Diablo, que es similar a las ideas de los románticos y los decadentes: Satanás es más un hombre diablo que el mismo Príncipe del Infierno. Las formas de superar la naturaleza humana, el amor y el arte, le permiten penetrar los límites del mundo real.

2. CONCLUSIONES

Como se ha expuesto, la frecuencia y distribución de determinadas palabras devuelven rasgos textuales que ayudan a generalizar características léxicas de la versión inglesa, en aras de la comparación ulterior con el original ruso, para refutar o vindicar las tesis relativas a la ambivalencia moral del personaje de Satán, a fin de corroborar o contrastar con datos objetivos todas las impresiones circulantes sobre la humanización del diablo a medida que avanza la trama de Andreev. El estudio de lo psicosocial a través del personaje diabólico puede, en un futuro, emprenderse desde la minería textual y el análisis de sentimientos, a fin de rastrear patrones estilísticos y de la dimensión afectiva subyacentes en la novela en cuestión. Particularmente relevante resultará, a partir de esta tentativa exploratoria, la aplicación de lo digital a la consideración de otras obras de este escritor ruso con la temática del diablo, desde el punto de vista de los estudios comparativos, una de cuyas tareas más importantes reside en el estudio de la evolución de las imágenes “eternas” que caracterizan cualquier “aventura legendario mitológica, folclórica y literaria” (Московкина, 1994) de la literatura del siglo XX. No en vano, las imágenes tópicas del caudal literario universal incluyen, en primerísima instancia, mitos como Don Juan, Prometeo y Fausto, entre otros. De esta manera, el comparatismo en intersección con las técnicas digitales facilita mapear cómo la mitología de raíz indoeuropea salta de un acervo literario a otro, apoyando los nexos entre intertextualidad y digitalización.

En síntesis, en todas las teorías de partida en torno a lo diabólico como eje temático semántico de la novela estudiada, se presume que la base de la estructura artística de *El diario de Satán*, de Andreev, descansa en una antítesis multifacética, en desarrollo y ramificada por todo un sistema de diferentes niveles de personalidades, e inicialmente determinada por la dicotomía externa Satán-Hombre, para después, como resultado de la metamorfosis, desplegarse en un yo interno o antinomia (Yo-hombre frente a Yo-Satanás). Se ha probado, en efecto y preliminarmente, que el relato, por un lado, es un antimito bíblico, pero por otro, obedece

a un intento de recrear estéticamente un “código corpóreo” existencialmente dado, como se pone de manifiesto en la estructura semántica de algunos lexemas analizados (externalidad: persona, cuerpo), interior (alma, personalidad). En un léxico jerárquico, y en virtud del análisis de frecuencias léxicas y distribuciones verbales, se está en disposición de percibir un primer paradigma “Hombre-Humano-Humanidad-Humanización-Deshumanización”, que se basa en la formación de palabras, y un segundo paradigma “Satanás-Diablo-Gusano-Borracho”, operante sobre el grado de aproximación cualitativa y cuantitativa al Absoluto del Mal, en una nomenclatura donde todos los componentes se definen por el onomástico anti Cristo, en línea con Гарипова (2020).

REFERENCIAS BIBLIOGRÁFICAS

- АНДРЕЕВ, Леонид. *Анатэма. Избранные произведения*. К.: Днипро, 1989.
- ANDREEV, Leonid. *Satan's diary*. Urbana, Illinois: Project Gutenberg. <www.gutenberg.org/ebooks/42665> [consulta 28/02/2024].
- Блок, Александр Александрович. *Собрание сочинений: В 8 т.* / А. А. Блок; подг. текста и примеч. М. И. Дикман. М.; Ленинград: Художественная литература, 1963.
- BUSHMINA, Irina Vasilievna. The poetics of mirror-masks in Satan's diary by L. Andreev. *Вестник КемГУ*, 1 (53), (2013), pp. 146-149.
- CUESTA FERNÁNDEZ, Jorge. “Nerón como precursor del anticristo en la literatura cristiana tardoantigua. una aproximación crítica”, *Antesteria* 1, (2012), pp. 379-390.
- FALTÝNEK, Dan y Vladimír MATLACH. Napax remains: Authorial features of textual cohesion in authorship attribution, *Preprint*, (2020), pp. 1-21.
- ГАРИПОВА, Гульчира Талгатовна. “Семиогенезис мифа о «новой экзистенции» в романе Л. Андреева *Дневник Сатаны*”, *RUDN Journal of Studies in Literature and Journalism*, 25 (1), (2020), pp. 7-16.

- ГРУШИНА, Анна Владимировна у Татьяна ЖУРЧЕВА Валентиновна. “Мир людей в присутствии сатаны (по роману Л. Андреева «Дневник Сатаны»)”, *Материалы xiviii Самарской областной научной студенческой конференции*. Том 2, (2022), pp. 11–22.
- IBM. *Statistical distribution*. <<https://www.ibm.com/docs/es/cognos-analytics/12.0.0?topic=terms-statistical-distribution>> [consulta: 28/02/2024].
- ИСАЕВ, Сергей Геннадьевич. «Сознанию неизвестная мощь...» *Поэтика условных форм в русской литературе начала XX века*. Великий Новгород: НовГУ им. Ярослава Мудрого, 2001.
- JOCKERS, Mathew. *Text Analysis with R for Students of Literature*. New York/London: Springer, 2014, pp. 3–56.
- КАМАНИНА, Елена Владимировна. Неомифологический роман Л. Н. Андреева «Дневник Сатаны» <<http://academy.cross-kpk.ru/bank/3/005/HTML/kaman.html>> [consulta 28/02/2024].
- КОВАЛЕВА, Ирина Александровна. Образ дьявола в культуре (на основе романа Л. Андреева «Дневник Сатаны» и романа М. Булгакова «Мастер и Маргарита») / И. А. КОВАЛЕВА. – Текст: непосредственный // Молодой ученый. – 2014. – № 4 (63). – С. 1211–1212. – URL: <https://moluch.ru/archive/63/10051/> (дата обращения: 12.05.2025). “Образ дьявола в культуре (на основе романа Л. Андреева «Дневник Сатаны» и романа М. Булгакова «Мастер и Маргарита»)”, *Молодой ученый*, 4, (2014), pp. 1211–1212.
- KOLMAKOVA, Olga Anatolevna. *Xronotop dnevnika v romane L. Andreeva “Dnevnik Satany”*. Natsional’nyj issledovatel’skij Tomskij universitet: Tomsk, 2019.
- КОСЕНКО, Алексей. “«Фауст» и В. Гете и «Дневник Сатаны» Л. Андреева: конфликтная связь”, *Вісник Запорізького державного університету*, 3, (2001), pp. 1–2.
- LOZANO ZAHONERO, María. “Una nueva visión de la supuesta influencia de Madame Bovary en La Regenta a través de la estilometría y el análisis de sentimientos basados en lenguaje R”, *Orillas*, 9, (2020), pp. 573–607.
- МИРОНОВ, Алексей Викторович. “Принцип исследования неомифологического романа XX века (жанровый

аспект)”, *Вестник Нижегородского университета им. Н. И. Лобачевского*, 4 (2), (2010), pp. 903–906.

Московкина, Ирина Ивановна. *Поэтика прозы Леонида Андреева. Жанровая система и художественный метод.* – К: Нац. АН Укр. Институт им. Т.Г. Шевченко, 1994.

Московкина, Ирина Ивановна. *Между «pro» и «contra»: координаты художественного мира Леонида Андреева*, Харьков, 2005.

Нестерова, Татьяна Александровна. «Русская идея» и публицистика Л.Н. Андреева 1910-х гг. [Текст]: монография, Изд-во ИГПИ им. П. П. Ершова, 2009, p. 92.

NIKIFOROVA, Svetlana Mikhailovna y Svitlana Eduardivna МЕКНАСИУК. “Stylistic methods of depicting the devil in the novels by L.N. Andreyev and M.A. Bulgakov”, *Наукові записки ХНПУ ім. Г.С. Сковороди. Літературознавство*, вип. 3-4 (89-90), (2018), p. 127.

R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

SINCLAIR, Stéfan y Geoffrey ROCKWELL. *Voyant Tools*, 2016 <<http://voyant-tools.org/>> [consulta 28/02/2024].

Тодоров, Цветан. *введение в фантастическую литературу* / Цв. Тодоров; пер. с фр. Б. Нарумова. М.: Дом интеллектуальной книги: Русское феноменологическое общество, 1997.

HUMANIDADES EN EL ESPACIO: ANÁLISIS Y VISUALIZACIÓN

Nadège Rollet, José Antonio Berenguer-Sánchez,
Patricia Sánchez-Garcías, Jan Thiele

Resumen

Siguiendo el giro espacial, la investigación en humanidades ha hecho hincapié en el papel fundamental de las representaciones espaciales para que la mente humana estructure y organice el conocimiento. Tras examinar algunos principios básicos de las conceptualizaciones mentales en términos espaciales, este artículo presenta tres estudios de caso aparentemente inconexos –(1) sobre los inicios de la cultura impresa de la literatura yemení, (2) sobre las variantes lingüísticas en los papiros griegos, y (3) sobre los topónimos y gentilicios encontrados en obras filológicas orientalistas (ss. XVI-XVII)– que se basan en un enfoque metodológico común: el análisis geográfico y la visualización de sus datos. A través de la revisión de algunos resultados de nuestros tres estudios de caso, este artículo pretende contribuir al debate sobre el modo en que el análisis espacial amplía nuestras preguntas de investigación, perspectivas y conclusiones.

Palabras clave: giro espacial, análisis y visualización de datos, geolocalización, lingüística, historia.

Nadège Rollet, LabHD, ILC-CCHS, CSIC, Madrid
nadege.rollet@cchs.csic.es; ORCID: 0000-0002-8392-1344

José Antonio Berenguer-Sánchez, ILC-CCHS, CSIC, Madrid
j.berenguer@csic.es; ORCID: 0000-0003-0038-7148

Patricia Sánchez-García, IH-CCHS, CSIC, Madrid
patricia.sanchez@cchs.csic.es; ORCID: 0000-0001-8988-1382

Jan Thiele, ILC-CCHS, CSIC, Madrid
jan.thiele@cchs.csic.es; ORCID: 0000-0002-8865-5997

Abstract

Following the spatial turn, research in the humanities has emphasized the critical role of spatial representations for the human mind to structure and organize knowledge. After surveying some basic principles of mental conceptualisations in spatial terms, this paper presents three apparently unconnected case studies – (1) about the beginnings of print culture of Yemeni literature, (2) the linguistic variants in Greek papyri, and (3) the toponyms and demonyms found in Orientalist philological works (16th–17th c.) – that build on a shared methodological approach: the geographical analysis and visualization of their data. Reviewing some results of the three case studies, this paper attempts to contribute to the discussion of how spatial analysis expands our research questions, perspectives and findings.

Keywords: spatial turn, data analysis and visualization, geolocalization, linguistics, history.

1. EL ESPACIO COMO CONFIGURADOR DE LA COGNICIÓN HUMANA

El espacio en las Humanidades se encuentra en el corazón de la investigación, en tanto que las Humanidades son las disciplinas que tienen como objeto de estudio la cultura humana y, como tales, se relacionan con la cognición humana de forma doble. Ésta es a la vez su objeto de estudio, a través de sus manifestaciones que son los datos de la investigación, y a la vez, su *modus operandi* para generar nuevo conocimiento. El concepto de espacio, por su parte, actúa como referente de la cognición humana, particularmente a través de las metáforas de espacialización que dan estructura y que organizan la mayoría de nuestros conceptos fundamentales (Lakoff y Johnson, 1980).

Un paradigma muy común de estas metáforas de espacialización es el que se basa en la oposición entre «Arriba» y «Abajo». Esta oposición se basa en la experiencia del espacio por el ser humano y actúa como metáfora de «Más» y de «Menos», según las metáforas conceptuales de «Arriba es Más» y «Abajo es Menos». A su vez, la oposición «Más/Menos» actúa como metáfora de la

oposición «Positivo/Negativo», según las metáforas conceptuales «Más es Positivo» y «Menos es Negativo», y este juego de oposiciones metafóricas rigen, por ejemplo, los movimientos convencionales del pulgar para demostrar la aprobación –pulgar hacia arriba para significar una reacción positiva–, o la desaprobación –pulgar hacia abajo para significar una reacción negativa–, en el marco de un juicio de valor que distingue entre lo «Bueno» y lo «Malo». De esta forma, «Arriba es Más/Positivo/Bueno» y «Abajo es Menos/Negativo/Malo». Esas metáforas de espacialización están tan ancladas en la cognición humana que es a veces imposible nombrar conceptos en el lenguaje sin hacer uso de ellas. En el dominio de la percepción acústica de la música, una nota se define metafóricamente por su «altura», según una dicotomía entre «nota alta» y «nota baja». Si es cierto que en español existen las palabras «agudo» y «grave» para calificar el sonido, tampoco escapan al engranaje de la oposición metafórica «Arriba/Abajo», dado el elemento metafóricamente ascendente de la punta contenido en «agudo» por una parte, y el origen latino *gravis* de sentido «pesado», es decir algo cuyo peso le tira para abajo, por otra. En las propias partituras, se reflejan convencionalmente estos conceptos de percepción por movimientos ascendentes, hacia notas más altas, o descendentes, hacia notas más graves, de la escritura.

Ahora bien, desde el punto de vista epistemológico, ¿cómo se concibe la relación entre el espacio y el conocimiento? Según la reflexión epistemológica de Deleuze, el concepto de espacio es pensado en un primer tiempo a través de la distancia y la distancia es lo que manifiesta la diferencia (Deleuze, 1969; Regnauld, 2012). La mente detecta que los objetos (las ideas) difieren porque son espacialmente disociados. Según esta aproximación, la disposición, la organización espacial y la distribución vuelven a ser un factor de producción del sentido y el espacio, en vez de estar subordinado al tiempo, se valoriza en tanto que es generador de sentido, de diferencia percibida por la distancia. Asimismo, Deleuze escribe que «el pensamiento tiene una geografía antes de tener una historia» (1969: 152).

2. EL ESPACIO COMO MARCO DEL ANÁLISIS DE DATOS EN LAS HUMANIDADES DIGITALES

Aquella geografía del pensamiento se puede expresar a su vez en términos de metáforas conceptuales organizadas por una dicotomía entre «Diferencia es Distancia» y «Similitud es Proximidad». En una publicación que propone evaluar cuál de dos métodos de visualización de matrices de proximidad (o de distancia), el *Mul-tiDimensional Scaling (MDS)* o el *Graph Analysis Layout (GLA)*, se adapta mejor para datos que llaman «de dominio cultural» y que se pueden identificar con los datos de la investigación en Humanidades, los autores defienden que la visualización por GLA es a menudo más apropiada, siempre que no son compatibles con las características de definición de las distancias euclidianas, que fundamentan la aplicación exitosa del MDS (DeJordy *et al.*, 2007). En realidad, en cualquier de estos dos casos, es interesante observar que la visualización de la proximidad o de la similitud de los datos se estructura espacialmente, o por un juego de distancias euclidianas entre los puntos repartidos en la superficie del espacio de visualización por MDS, o por las líneas entre los nodos, que trazan metafóricamente en el GLA caminos de similitud y que estructuran el conjunto.

Por otra parte, hablamos de un «*spatial turn*» en Humanidades Digitales, con referencia a menudo a Sistemas de Información Geográfica (SIG) y al uso de técnicas y herramientas geográficas por disciplinas distintas de la propia geografía, que denominamos neogeografía. Estas técnicas y el desarrollo de tales herramientas digitales han florecido gracias a la premisa por una parte de la definición del espacio, concebido como espacio mental, que se ubica en el corazón de nuestra percepción y de nuestro conocimiento, y por otra a la del estudio de relaciones simbólicas entre espacio físico y las ideas, o el conocimiento. Por último, conviene volver a la cita de Deleuze que hace hincapié en el hecho de que el espacio conceptual no está subordinado al tiempo. Es más, la percepción del tiempo también está regida por metáforas de espacialización, tal y como lo evidencia la clásica representación de

la evolución temporal en forma de flecha, que, en sistemas de lectura de la izquierda a la derecha, refleja el camino del tiempo, siendo representado, a mano izquierda, un tiempo más antiguo que, a mano derecha, un tiempo más reciente. De este modo, se aplican las metáforas «Antes es Detrás» para conceptualizar el tiempo pasado y «Después es Adelante» para conceptualizar el tiempo futuro. Estas metáforas, que estructuran la representación del eje temporal de cualquier análisis de datos en la sociedad moderna y contemporánea, ya se cristalizaban en la Antigüedad a través de la figura del dios Jano, el dios del mes de enero, representado con dos caras, una que mira hacia atrás y otra hacia delante, en tanto que se encuentra al límite del año pasado y del año nuevo. En el cuadro del «*spatial turn*» ya mencionado de la investigación en Humanidades Digitales, el eje temporal representa un parámetro de medición importante en muchas de estas metodologías de visualizaciones analíticas ya que, en conjunto con la idea de distancia en el espacio para mapear diferencias conceptuales, el tiempo a menudo tiene la capacidad de hacer surgir las diferencias, de las cuales el espacio como superficie es el receptáculo.

En los siguientes apartados, se exponen metodologías aplicadas, que se sirven de una representación geográfica de los datos de la investigación en Humanidades, para ayudar en su análisis y enriquecerlo, a través de casos de estudios distintos. El primero de ellos trata del visualizador «Early Prints of Yemeni Works», aplicado a datos de impresión de libros entre 1850 y 1950 aproximadamente, y que ejemplifica los aportes de una representación de los lugares de publicación de aquellos impresos, ponderada en función de la importancia de esta producción impresa, para determinar, entre otras cuestiones, cuáles son los polos de difusión de la tradición literaria del Yemen, y para analizar el trazado de esta difusión en un referente espacio-temporal. En el segundo caso del visualizador de variantes lingüísticas en el corpus digital de los papiros griegos, el eje temporal es de mayor importancia, para medir y visualizar la distribución y la extensión de los fenómenos lingüísticos atestiguados en dichas fuentes. En el último caso de la

representación geográfica de topónimos y gentilicios de obras filológicas orientalistas, el eje temporal se consolida por otro eje secuencial de filtros conceptuales, que son las fuentes de proveniencia de los topónimos, caracterizadas por su autor, y que permiten hacer surgir la distancia, o la diferencia, en la modelización de las diferentes perspectivas eruditas de cada autor y de la extensión de su conocimiento del árabe.

3. METODOLOGÍAS APLICADAS

3.1. «EARLY PRINTS OF YEMENI WORKS»

«Early Prints of Yemeni Works» (EPYW) es un proyecto que intenta mapear las publicaciones de libros impresos yemeníes entre la segunda mitad del s. XIX y la primera mitad del s. XX. Este marco histórico comprende una época de transición importante en la historia moderna del Yemen: Yemen siempre ha sido una región disputada entre una multitud de poderes políticos, entre ellos los imames zaydíes, líderes religiosos de una pequeña comunidad chií. Desde el siglo X, los imames zaydíes se imponían a veces con más y a veces con menos éxito como gobernadores de un territorio más o menos extendido en la región montañosa del Yemen del norte, compitiendo con otros poderes. A mediados del siglo XIX, Yemen cayó bajo dominio otomano. A principios del siglo XX, el imam Yaḥyā Ḥamīd al-Dīn (m. 1948) restableció el imamato zaydí, acabó con el dominio otomano y fundó en 1918 el reino mutawakkilí. Inspirado por el ejemplo otomano, el imam Yaḥyā hizo pasos importantes hacia la creación de un estado nacional moderno, intentando crear un control más fuerte y centralizado que sus predecesores. Hacia el exterior, el imam Yaḥyā buscaba la autonomía de su país, resistiendo a las interferencias masivas de imperios y regímenes coloniales de su tiempo y estableciendo relaciones diplomáticas con determinados estados amigos. En el interior del país, aspiraba a crear instituciones e infraestructuras modernas (por ejemplo, militar, administración, carreteras, ferrocarriles, correos y telégrafos). Las reformas de modernización incluían un programa inte-

lectual-científico-religioso con instituciones públicas, entre ellos el museo nacional, una biblioteca, la mezquita-universidad de la capital Sanaa y periódicos. La introducción de la tecnología de impresión –imprescindible para la producción de periódicos– era parte de estas reformas.

El interés del proyecto EPYW consiste por un lado en su posible contribución a una mejor comprensión del programa intelectual-científico-religioso implementado por el imam, ya que las iniciativas autóctonas para promover la impresión de libros venían en gran medida de intelectuales que eran cercanos al imam o implicados en su administración. Los libros impresos que veían la luz gracias a estas iniciativas se publicaron tanto en Yemen como fuera del país, y lo último suponía la implicación de actores no-yemeníes. Además, no-yemeníes, incluyendo intelectuales de otras partes del mundo islámico así como orientalistas occidentales, actuaban también como editores de textos yemeníes. Tras incluir las actividades editoriales fuera del Yemen, EPYW facilita también datos que permitirán abordar la pregunta sobre cuál era el interés de los no-yemeníes en la tradición literaria del Yemen.

Recientemente, la investigación sobre la transición del manuscrito al libro impreso en el mundo árabe-islámico ha visto un impulso innovador con el estudio de El Shamsy sobre el auge de la actividad editorial en El Cairo durante el siglo XIX y los inicios del siglo XX (2020). En este libro, El Shamsy retrata un grupo relativamente pequeño de editores que pusieron en marcha un ambicioso programa de renacimiento y reforma intelectual, y así influyeron profundamente en lo que se iba a considerar como el «canon» de la tradición literaria islámica. EPYW comparte preguntas centrales con el estudio de El Shamsy, centrándose en una región del mundo árabe-islámico a la cual se le atribuye el estereotipo de una cultura supuestamente anticuada, aislada y poco afectada por innovaciones procedentes de fuera. Nuestra investigación parte de algunos trabajos anteriores sobre el caso yemení. En su libro *The Calligraphic State*, Messick dedica un capítulo a la introducción de la tecnología de impresión en Yemen, contextualizando en las políticas de

reforma cultural-científica del imam Yaḥyā (1993: 115-131; retoma parte de sus resultados en Messick, 2022). Una fuente importante de Messick es un inventario de libros yemeníes impresos publicado en el año 1938 por el orientalista italiano Ettore Rossi (1938). Asimismo, Haykel pasa revista a algunas de las iniciativas para promover la impresión de libros bajo el imam Yaḥyā Ḥamīd al-Dīn en su biografía intelectual del reformista yemení Muḥammad al-Šawkānī (m. 1834) (2003: 200-210). Estos estudios ofrecen un relato valioso del contexto histórico que se resume aquí brevemente.

Las primeras imprentas fueron introducidas en Yemen en 1877 por los otomanos. La tecnología estaba destinada principalmente a la producción de periódicos que se publicaron en árabe y turco. Además, ya se usaron los primeros libros impresos en las escuelas de nivel superior de Yemen en época otomana. Cuando los otomanos abandonaron el Yemen, dejaron atrás el equipo de impresión. Rossi relata a finales de los años 1930 que la única tecnología de que disponían los impresores yemeníes eran dos imprentas que habían dejado los otomanos. Sin embargo, la falta de tecnología elaborada fue solo uno de los factores que determinaron cómo se desarrolló la cultura de imprenta para la difusión de libros en Yemen. Mientras se usaban las imprentas para la publicación del primer periódico yemení *al-Īmān* desde 1926, el imam Yaḥyā restringió el uso de libros impresos en escuelas avanzadas e incluso los prohibió para algunas disciplinas específicas (Rossi, 1938: 568, 579; Messick, 1993: 115-119). La reticencia hacia la impresión de libros en dichas disciplinas no estaba exenta de ambivalencia, ya que estos libros se publicaron en otros países árabes con la autorización del imam. ¿Estos impresos estaban principalmente destinados a una audiencia no-yemení en el mundo árabe? Efectivamente tenemos evidencia de que intelectuales egipcios que pertenecían al movimiento de reforma islámica tenían particular interés en textos específicos de la tradición yemení y que, por consiguiente, existía una audiencia que se interesaba por la tradición literaria del país fuera de él (Haykel, 2003: 206-209). Sin embargo, si la impresión de libros impresos fuera del Yemen estaba princi-

palmente destinada para no-yemeníes, ¿cómo explicamos que la impresión de libros como la primera historia nacional –que claramente está relacionada con la política de formación de un estado nacional moderno– se realizara también fuera del Yemen? Por otro lado, es llamativo que los libros que se publican con autorización del imam representan facetas variadas de la tradición intelectual y religiosa del Yemen y no permiten concluir que se promueve una corriente ideológica específica.

Una mejor comprensión de los inicios de la impresión de libros yemeníes, los motivos de sus promotores dentro y fuera del país, así como las audiencias de estos libros requiere una recopilación sistemática de los datos bibliográficos de los libros que se imprimen para encontrar respuestas a tales preguntas como: ¿Dónde y cuándo se publican? ¿Quiénes eran las personas involucradas en los procesos editoriales? ¿Cuáles son las editoriales que publican libros yemeníes fuera del país? ¿Qué títulos se publican?¹.

Para abordar estas preguntas, se han creado dos herramientas principales. La primera es una bibliografía pública Zotero Groups que recopila los datos bibliográficos, y que es libremente accesible en la web para poder acoger a más colaboradores (https://www.zotero.org/groups/4963950/early_prints_of_yemeni_works). La bibliografía se ha exportado en formato de archivo CSV, y a base de ello se ha desarrollado por un lado una base de datos SQLite, sobre la cual actúa un visor que consiste en un mapa interactivo. La herramienta está en su fase piloto y ya está accesible en la web (<https://epyw.glitch.me>). Por otro lado, se ha creado un conjunto de datos en edición TEI-XML que recoge la estructura de la base de datos para su depósito en el repositorio ins-

1 Un proyecto dirigido por Anne Katrine Bang (Universidad de Bergen) y titulado «Mprint@EAST_AFRICA. Islamic Manuscript, Print and Practice: Textual adaptations in coastal East Africa, c. 1880–2020» adopta una metodología similar de colección exhaustiva de datos bibliográficos con el objetivo de estudiar la transición del manuscrito al libro impreso para el caso de África oriental. Una descripción del proyecto se encuentra en el siguiente enlace: <https://www.uib.no/en/ahkr/143764/mprinteastfrica> [consultado el 28/01/2024].

titucional DIGITAL.CSIC, respetando los principios FAIR (Rollet y Thiele, 2023).

El usuario se puede servir de la bibliografía Zotero para hacer análisis básicos, sirviéndose de los filtros y búsquedas de la aplicación, o simplemente ordenando las entradas en su orden cronológico, por ejemplo. Los datos bibliográficos incluyen también enlaces a repositorios en línea cuando existen reproducciones digitales de los textos. Una vez exportados al formato CSV, se pueden también aplicar análisis de diferentes tipos a los datos.

Una representación cuantificada de la actividad editorial permite analizar la evolución cronológica de la impresión de los libros. Si se añade al análisis de los datos una dimensión geográfica, se ve que la impresión de libros yemeníes empieza en Europa durante la segunda mitad del siglo XIX con ediciones de textos preparadas por orientistas –coincide con la época cuando bibliotecas europeas adquieren las primeras colecciones de manuscritos yemeníes. Luego, la India se vuelve en un centro de actividad destacable², hasta que se concentra a partir de 1910 en el mundo árabe principalmente en El Cairo y Sanaa.

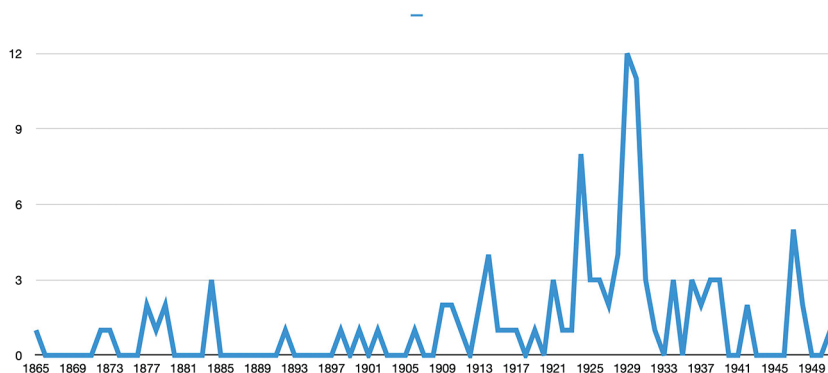


Figura 1. Trayectoria cronológica de la actividad editorial

- 2 Haykel señala el interés que al-Šawkānī recibió en la India como autoridad en la transmisión de tradiciones proféticas (2003: 82n23), pero se le parece escapar que se producen impresiones litográficas de un número significativo de sus obras en la India del siglo XIX.

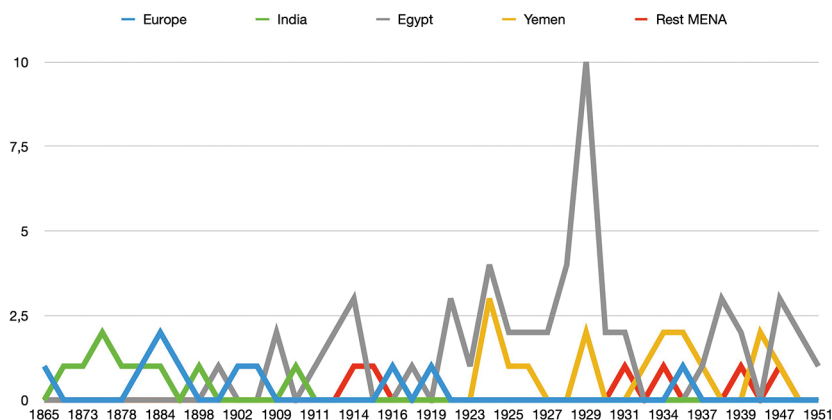


Figura 2. Trayectoria cronológica de la actividad editorial por región

El principio del visor es proyectar la actividad editorial en el espacio, y su diseño como herramienta interactiva permite al usuario hacerle preguntas analíticas más avanzadas. Según los filtros que se le aplican, se visualizan todos los libros de la base de datos, o se puede filtrar por personas específicas, por publicaciones de años específicos, o se pueden combinar los dos filtros. La actividad editorial se representa por círculos en el mapa, cuyos tamaños cuantifican la actividad. Con un clic en los círculos se puede consultar una lista de los libros, así como gráficos cuantitativos de los autores/editores implicados, de los años de publicación y de las editoriales que publican los libros.

El visor muestra claramente que el epicentro de la actividad editorial es El Cairo, lo que no nos puede sorprender, siendo la capital egipcia la ciudad del mundo islámico con la tradición más antigua de impresión de libros. Sin embargo, destaca el papel de Sanaa que a pesar de los pocos recursos técnicos –solo existían dos imprentas– da lugar a una actividad editorial importante. Las publicaciones de Sanaa se concentran en tres editoriales, que en realidad son dos, ya que una de ellas publica bajo dos nombres distintos. En comparación, la distribución de editoriales en El Cairo es muy variada.

Otro resultado destacable que proporciona el análisis cuantificable es el papel arrollador de ‘Abd al-Wāsi’ al-Wāsi’ī (m. 1960) como promotor de libros impresos. Al-Wāsi’ī era un intelectual en el entorno del imam Yaḥyā, autor de la primera historia

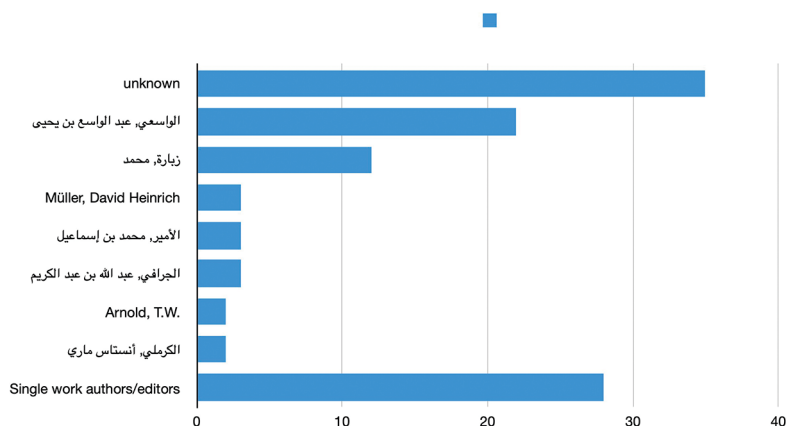


Figura 3. Autores y editores implicados en la actividad editorial

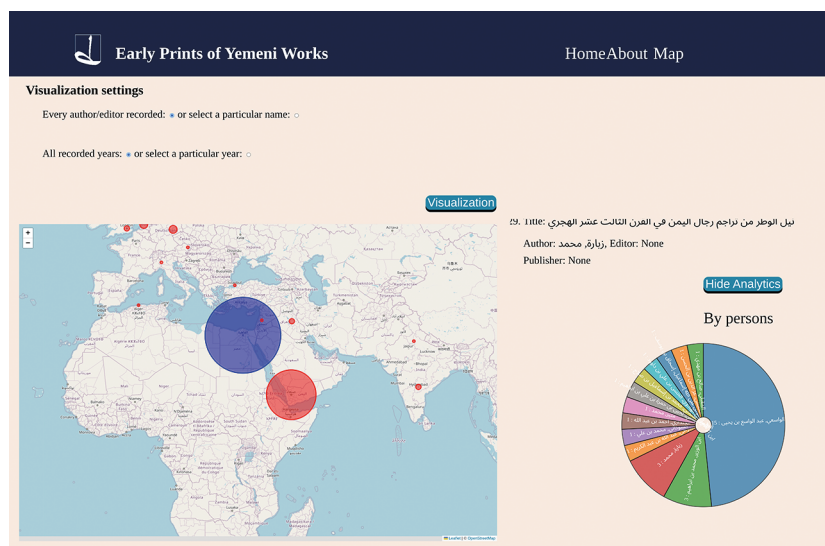


Figura 4. Autores y editores publicando en El Cairo

nacional del Yemen, y estaba implicado como autor y editor en casi la mitad de los libros publicados en El Cairo y algo menos de un cuarto de los libros publicados en Sanaa.

3.2. VISUALIZADOR DE VARIANTES LINGÜÍSTICAS DEL CORPUS DIGITAL DE LOS PAPIROS GRIEGOS

El diseño de un visualizador de variantes gráficas de textos antiguos (<https://lphd.glitch.me>) se inscribe dentro de un proyecto de investigación, actualmente en curso, cuyo objetivo fundamental es analizar la aplicación de las nuevas tecnologías al estudio de la lengua de los papiros documentales escritos en griego antiguo.

Uno de los avances más significativos que se ha llevado a cabo en los últimos años, dentro del ámbito de la investigación en lenguas antiguas, ha sido la creación del corpus digital de los papiros documentales griegos, contenido en el *Duke Databank of Documentary Papyri* (DDbDP, <https://papyri.info/docs/ddbdp>). Este corpus reúne en acceso abierto la edición filológica de más de 60.000 documentos, datados fundamentalmente entre los siglos IV a.C. y IX d.C., que contienen alrededor de 3.800.000 palabras en griego –en ocasiones se da una cifra superior, producto de no haber descontado en el cómputo los símbolos y cifras numéricas que aparecen con una gran frecuencia en muchos de estos documentos, ni un considerable número de palabras en otras lenguas (Berruenguer-Sánchez, 2023a: 10n12; 2023b: 137n27).

Por razones en las que no podemos extendernos (fundamentalmente condiciones de conservación) la inmensa mayoría de estos documentos se han localizado en Egipto. Actualmente, sus ediciones digitales ofrecen la posibilidad de su consulta tanto en formato html, como en formato xml, cuyo etiquetado ha recogido sistemáticamente las peculiaridades señaladas por los editores que estudiaron y publicaron, a lo largo de los años, la transcripción de esos papiros (Sosin, 2010; Depauw y Stolk, 2015). Su lengua se caracteriza por reflejar un alto nivel de inmediatez y espontaneidad. Hay documentos de muy diverso tipo (documentos administrativos, listas oficiales y privadas de todo tipo, cartas también oficia-

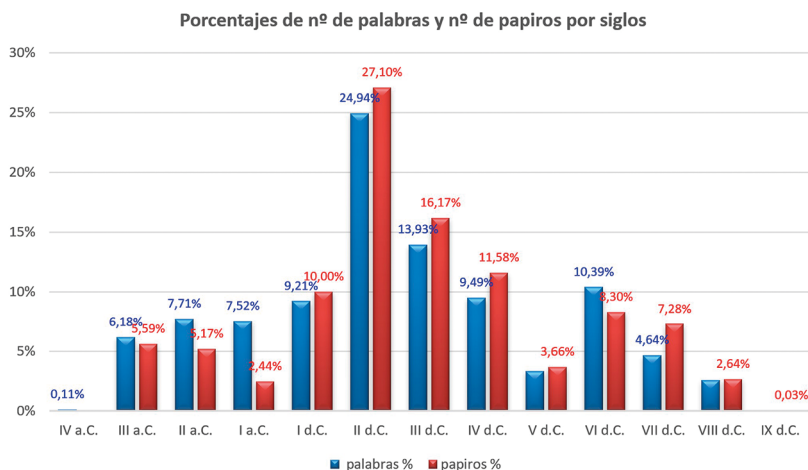


Figura 5

les y privadas, testamentos, ejercicios de escritura, incluso invitaciones de boda...). Estos documentos responden a necesidades muy diferentes y reflejan el día a día de quienes los redactaron. Por eso, en su escritura aparecen variantes gráficas que, frente a las ortografías normativas, reflejarían las características de la lengua hablada en cada momento (Browning, 1983: 23-24; Dickey, 2009: 150-151; Horrocks, 2010: 114-115). Al igual que puede observarse hoy en día en los textos no literarios, el interés fundamental de estos escritos no era el estético y los redactaban individuos con muy distintos niveles de educación y pericia en el manejo de la lengua escrita. De ahí, el gran valor que tienen sus errores gráficos cuando corresponden a cómo se hablaba en cada momento y no a cómo debía escribirse. En realidad, esto es algo que puede comprobarse también hoy en día en múltiples errores de escritura que se corresponden a la percepción que tiene quien los comete de cómo puede reflejar lo que escucha, ignorando o desconociendo la ortografía correcta. Podemos verlo, por ejemplo, en la confusión en el uso de los grafemas <b/v> en español, debido a su pronunciación equivalente, o en la omisión o inserción errónea del grafema <h> debido a que no corresponde a la pronunciación actual de ningún fonema.

Hasta ahora, la lengua griega de los papiros documentales ha sido descrita y analizada en manuales y monografías a partir del estudio y la cita de algunas decenas de estos ejemplos. Hoy en día, con las posibilidades ofrecidas por los nuevos desarrollos digitales, podemos llevar a cabo análisis exhaustivos basados en todas las variantes estudiadas en las ediciones, que en algunos casos alcanzan varios miles de ejemplos (Depauw y Stolk, 2015; Berenguer-Sánchez, 2020, 2021, 2023a y b).

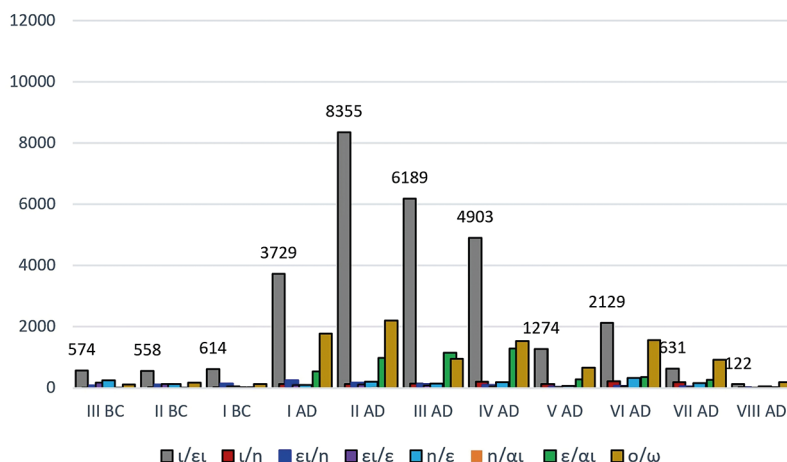


Figura 6

A ello se añade el reto de visualizar todos esos datos sobre variantes lingüísticas, aprovechando al máximo las posibilidades ofrecidas por las nuevas herramientas digitales. Así, junto al número de correcciones editoriales que corresponden a un fenómeno de alternancia gráfica como los mencionados, y debido a la desproporción en el número de palabras que tendría cada subcorpus cronológico que corresponde a un siglo, como hemos visto en el primer gráfico, podemos calcular y visualizar, mediante líneas de frecuencia, los índices de frecuencia de cada fenómeno por siglos (Berenguer-Sánchez, 2023a y b).

Dentro de las distintas opciones de visualización, nos hemos planteado recientemente la posibilidad de ver esas variantes

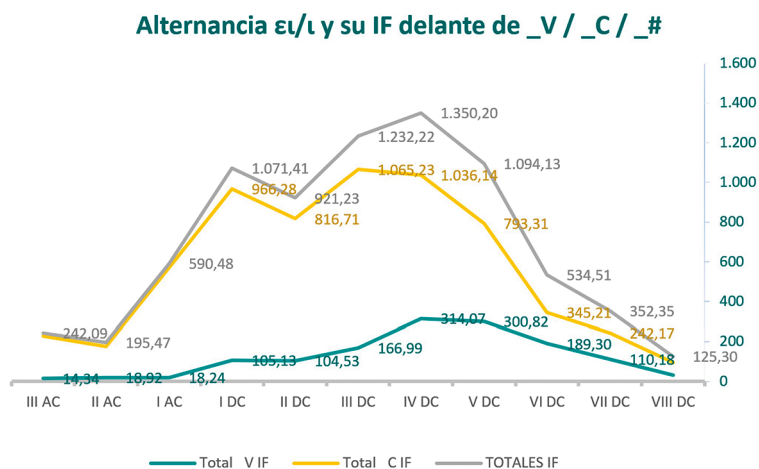


Figura 7

clasificadas, atendiendo a su localización geográfica y temporal, mediante mapas de variantes que muestren su evolución diatópica y diacrónica a lo largo de los siglos. De acuerdo con lo que hemos mencionado, el espacio fundamental sería Egipto, aunque se conservan algunos papiros en otros lugares del Mediterráneo oriental, o próximos a él, y el período abarcaría del siglo IV a. C. al IX d.C. Este enfoque resulta especialmente novedoso en el estudio de la evolución de una lengua antigua, como el griego.

Los datos sobre los cuales actúa el visualizador se recogen primariamente por tratamiento y análisis automáticos del texto propio de los papiros contenidos en las colecciones xml del *DDbDP* y de sus metadatos asociados, contenidos en las colecciones xml *Heidelberger Gesamtverzeichnis der Griechischen Papyrusurkunden Ägyptens* (HGV). A partir de los textos mismos, se recopilan las instancias de variantes, anotadas en TEI-XML por un elemento *choice*, del cual depende por una parte la forma de origen escrita en el papiro y anotada por un elemento *orig*, y por otra su forma regularizada por los editores y anotada por un elemento *reg*. A partir de los archivos de metadatos, se extrae la información de datación y de lugar, cuando existe y en su variedad.

Los datos primarios de lugar de procedencia de los papiros se completan gracias a la consulta automática de los enlaces *Pleiades*, referenciados por el atributo *ref* en los elementos *place-Name* del xml de HGV, desde los cuales se recogen las coordenadas correspondientes. En cuanto a los datos de datación, se procede a un análisis de los atributos *when* en caso de una fecha absoluta, y de los atributos *notBefore* y *notAfter* en el caso de una fecha relativa, para cada elemento *origDate* del xml del HGV, cuyos valores numéricos y estandarizados permiten su procesamiento para calcular el o los siglos de procedencia y un grado de precisión de esta datación secular.

Los datos primarios de las propias variantes no ofrecen más que las dos cadenas de caracteres correspondientes a las dos formas, la de origen y la regularizada, enteras. En otros términos, carecen de anotación analítica que aporte información cualitativa sobre el tipo de variación entre una grafía segmental y otra dentro de la palabra. Para producir esta información, se mide la diferencia entre las dos cadenas de caracteres según una métrica similar a la de Levenshtein, que basa su computación a partir de una matriz de diferencias entre la secuencia de caracteres de ambas cadenas y que permite comparar las dos cadenas y obtener un listado de las operaciones –hay tres tipos: inserción, supresión o sustitución de carácter– necesarias para la transformación en la cadena regularizada de la cadena de origen, donde se produjo el «error». Esta métrica se adapta para pasar de la lista de las transformaciones operadas sobre el carácter de la cadena a las operaciones efectuadas sobre el grafema, que puede constar de dos caracteres para anotar un único fonema, que resulta una unidad más adaptada a la descripción del tipo de error, cuando éste tiene por origen una realización fonológica «mal» percibida o que coincide con el uso de varias grafías para transcribirla. Por ejemplo, el nombre propio de Phoebe en caso genitivo, escrito <Φοίβης> (/phoĩbɛ:s/) en griego clásico, está escrito <Φύβις> en el papiro p.lond.herm.1, forma en la cual necesitamos anotar dos sustituciones, la <οι/υ> y la <η/ι>, que atestiguan el fenómeno de iotacismo según el cual el fonema

/i/, marcado por la sola <ι> en griego clásico, pasa a marcarse también por <οι>, <υ> y <η>, cuya pronunciación ya no se distingue como era el caso en época clásica –/oi/, /y/ y /ε:/ respectivamente–. En cambio, una aproximación por unidad de carácter da por resultado una serie de tres operaciones en este ejemplo, la sustitución de <ο/υ> seguida por la supresión de <ι> y la sustitución <η/ι>, en la cual las dos primeras no se reconocen como objeto de un mismo proceso.

Asimismo, la base de datos sobre la cual actúa el visualizador tiene varias colecciones enlazadas entre sí: la colección de los tipos de variación gráfica, la de los elementos *choice* donde aparecen estos tipos de variación con toda la información analítica, la de lo papiros con sus metadatos, la de siglos y la de puntos en el mapa (coordenadas). De esta manera, la herramienta permite al usuario elegir en la lista de tipos de variaciones cuál quiere visualizar y obtener el mapa donde se representan los puntos que corresponden al lugar de procedencia de los papiros en los cuales están atestiguados ejemplos del tipo de variación en cuestión. El color de estos puntos va cambiando según la concentración en el lugar de las ocurrencias de la variación gráfica y se ofrece un cursor para recorrer los datos por siglo de datación de los testimonios.

Ese visualizador es una herramienta muy útil para examinar la evolución de la distribución de un tipo de variación gráfica en el territorio cubierto por los testimonios y en el tiempo. Es cierto que son problemáticos los casos de variantes en los papiros cuya procedencia geográfica desconocemos, por una parte, porque no podemos visualizarlos en el mapa y, por otra parte, porque su datación secular es relativa, ya que en ocasiones es necesario decidir un siglo de referencia en el cual aparecen o no en el mapa. Por esta razón, más allá del estado piloto en el que se encuentra la visualización actualmente, se ofrecerá al usuario la posibilidad de configurar su búsqueda para incorporar o no resultados que tienen una datación relativa y de determinar el rango de estimación que le parece o no aceptable. Aun así, el visualizador ya permite obtener datos de interés en la investigación que estudia

los fenómenos a partir de las variantes, tal y como lo demuestra la figura 8.

En este ejemplo, se contempla analizar la confusión en los papiros entre grafemas que marcan las oclusivas sonoras con los que marcan sus equivalentes sordos en griego clásico, influida por el contacto lingüístico en Egipto entre el griego y el copto, en el cual existe únicamente la serie de oclusivas sordas. En cambio, si visualizamos los casos de uso del grafema dedicado originalmente a la sonora <γ> donde se espera el uso de del grafema dedicado a la sorda <κ>, que atestigua esta pérdida de sonoridad por hipercorrección, observamos claramente que los testimonios de este fenómeno salen fuera del marco geográfico de Egipto.

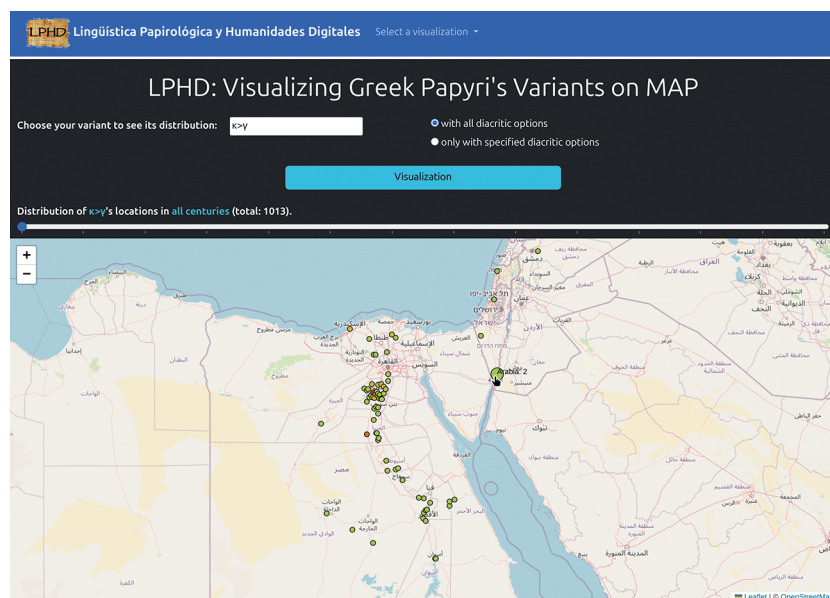


Figura 8. Distribución de la variante <γ> frente a la ortografía normalizada <κ>

3.3. ARCGIS Y PYTHON EN LA REPRESENTACIÓN GEOGRÁFICA DE TOPÓNIMOS Y GENTILICIOS DE OBRAS FILOLÓGICAS ORIENTALISTAS (SS. XVI Y XVII)

Cuando desconocemos el significado de una palabra, acudimos al diccionario. Diccionarios y gramáticas dan cabida a espacios que de algún modo son particularmente relevantes para el pensamiento o para la comunicación. Los lugares mencionados en gramáticas y diccionarios suponen, por lo tanto, un elemento significativo en la investigación del aprendizaje del árabe en Europa en los ss. XVI y XVII. Por un lado, cobran especial importancia en tanto que suponen proyecciones de un modelo mental. «The map is not the territory», como señalaba el filósofo Alfred Korzybski, incidiendo en la diferencia entre la realidad y los modelos que utilizamos para representarla (2005: XVII). Ese modelo construido resulta de mayor interés si consideramos el clima de erudición en el que se genera, un clima donde el estudio geográfico, histórico y filológico, en particular, la *geografía* y la *filología sagradas*, constituían ejercicios esenciales para la exégesis y la polémica religiosa, como demuestran obras como la *Geographia Sacra* de Samuel Bochart (1646) o estudios como el de Zur Shalev (2012).

A partir de las ideas expuestas, este trabajo toma como base el modelo de otros proyectos de georreferenciación como *Mapping Past Societies* (<https://darmc.harvard.edu/>) o, en el contexto árabe, *al-Turayyā Project* (<https://althurayya.github.io/>), y propone la representación geográfica de un conjunto de más de 400 topónimos y gentilicios procedentes de las obras de cuatro figuras: el *Arte para ligerame[n]te saber la le[n]gua arauiga* y el *Vocabulista arauigo en letra castellana* de Pedro de Alcalá (ca. 1455-post. 1508) publicadas en 1505, la *Grammatica arabica* de Marcos Dobelio (ca. 1572-1654) (BHR, Caja MS-2-051), el *Nomenclator Arabico-Latinus* de Abraham Ecchellensis (1605-1664) (BnF, ms. Arabe 4345) y el *Thesaurus Arabicae* de Tomás de León (1613-1690) (EEA, ms. 17). El análisis de estos términos vinculados a lugares permite profundizar en las fuentes que emplearon, en el modo de adaptarlas a su perspectiva, en sus métodos de análisis y compilación e incluso en su nivel de conocimiento del árabe.

La metodología propuesta para este estudio combina, a través de diferentes fases, el cotejo y edición tradicional de estas fuentes, varias de ellas manuscritas, con el uso de herramientas tecnológicas. Dentro de la parte tecnológica, se ha empleado Python, en primer lugar, para la búsqueda y el cotejo simultáneo de esos términos editados en 139 fuentes árabes (principalmente, gramáticas, diccionarios y obras geográficas) gracias al corpus de OpenITI (Nigst *et al.*, 2020). En esta búsqueda, se ha rastreado también la presencia de variantes toponímicas muy similares, empleando las métricas de similitud de Hamming. Se ha utilizado igualmente Python para la extracción automática de las coordenadas de las ubicaciones. Esta extracción automática o *web scraping* de bases de datos geográficas como GeoNames, Pleiades, OpenStreetMap (OSM) o Google Maps ha facilitado la obtención de las coordenadas necesarias para la representación geográfica mediante un volcado directo de los datos en formato CSV. En todo caso, la información geográfica se ha contrastado, además, con la que proporcionan el proyecto de *al-Turāyya* y el propio servicio de ArcGIS, el ArcGIS World Geocoding Service (AWGS).

Se ha empleado el programa de georreferenciación ArcGIS online para varios propósitos. Principalmente, se ha utilizado para la generación de cuatro mapas interactivos en los que se incluye para todas las localizaciones una ficha de información desplegable en cada punto. Podemos así visualizar los topónimos de estas localizaciones en las lenguas originales de los manuscritos (tanto el árabe y su transcripción como el latín o el castellano), acceder a notas y explicaciones de los autores con su traducción correspondiente o acceder a otra información de interés, como la relevancia del lugar o su presencia en fuentes gramaticales y lexicográficas propiamente árabes, así como recuperar la referencia concreta del manuscrito. Asimismo, se muestra el detalle de los datos geográficos consultados.

También con esta herramienta se han diseñado los mapas coropléticos o de calor a partir de las localizaciones de clara identificación para observar la concentración de los puntos. El mapa

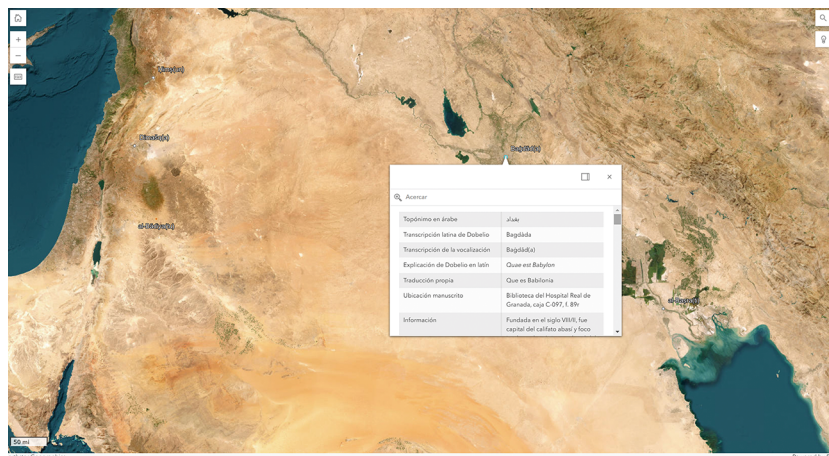


Figura 9

coroplético o mapa de calor nos permite observar con claridad en qué zonas se concentran los topónimos y gentilicios incluidos. En la obra gramatical de Dobelio, por ejemplo, los topónimos se restringen a la península arábiga y alrededores, con alguna excepción, lo que revela un uso de fuentes puramente árabe, e incluso más concretamente, de los principios de la tradición gramatical árabe, mientras que los topónimos de Pedro de Alcalá, mucho más numerosos y basados en el diccionario de Antonio de Nebrija, se concentran de manera especial en la península ibérica e Italia.

A través de *ArcGIS online*, se ha utilizado el cálculo de centroide o punto medio para ubicar de manera aproximada algún lugar de identificación imprecisa, como la ubicación de *al-Qaşà* (o *al-Quṣā*). Por ejemplo, en el diccionario de Ecchellensis se ofrecía la sucinta descripción de *al-Quṣà* como una garganta o desfiladero en Yemen. En ausencia de coordenadas y de alguna información más precisa en las fuentes, se generó el centroide o punto medio que se muestra abajo a partir de un listado de gargantas y desfiladeros ubicados en Yemen. De esta manera, aunque no disponemos de la ubicación exacta, podemos ofrecer una localización aproximada coherente con la orografía.



Figura 10

La representación geográfica de los topónimos y gentilicios nos permite apreciar detalles que en una simple lectura no resultan tan evidentes, como muestran estos dos ejemplos. El primero de ellos representa la concentración de los puntos añadidos por Pedro de Alcalá respecto a su fuente fundamental de Antonio de Nebrija que, en este caso, pone el foco particularmente en la zona de Granada, su lugar de producción, y es reflejo del interés esencial por obtener la conversión auténtica de los moriscos granadinos. En el segundo ejemplo, se observa la concentración de los puntos añadidos por Ecchellensis respecto a su fuente principal, el diccionario de *al-Qāmūs al-Muḥīṭ* de al-Fīrūzābādī. En este caso, los puntos se concentran en Túnez y, tanto por la estancia de este erudito allí como por la ausencia de estos topónimos en otras fuentes árabes, podemos especular con la idea de que parte del proceso de creación del diccionario implicase la colaboración de fuentes o informantes locales. En otros casos, estas representaciones revelan una proyección geográfica desconectada del lugar de producción o de las fuentes, mostrando una proyección mental que se relacio-

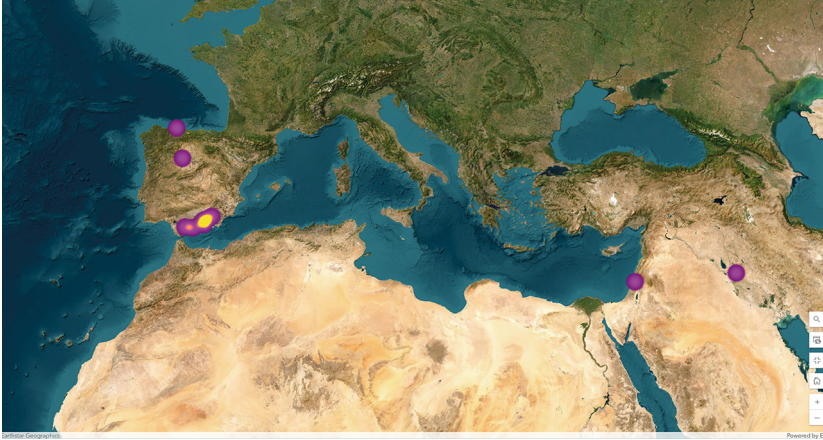


Figura 11. Concentración de las variaciones de Alcalá

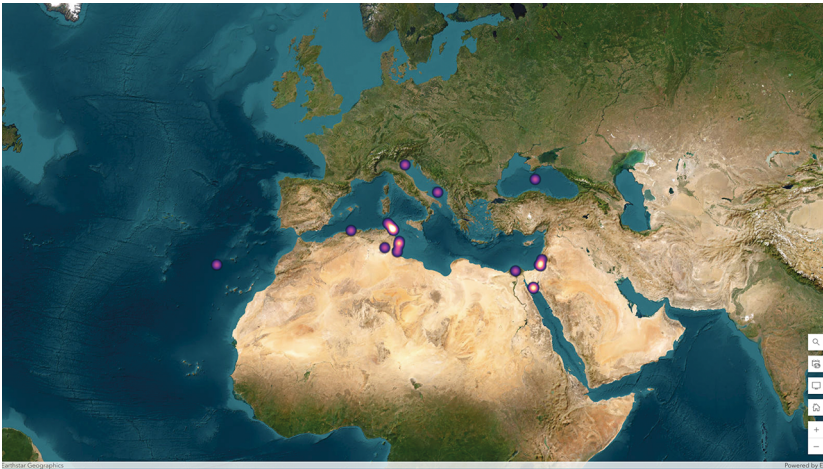


Figura 12. Concentración de las variaciones de Ecchellensis

na con una voluntad de identificación con los lugares sagrados del cristianismo.

Las herramientas empleadas nos proporcionan, de este modo, una mirada complementaria igualmente significativa para el estudio de los manuscritos. A través de estas representaciones geográficas, accesibles desde el enlace (<https://arcg.is/1T5m4v1>),

se ofrece un conjunto de datos con valor desde distintos ámbitos, como el geográfico, el histórico o el filológico desde diferentes lenguas como el latín o el árabe, con el posible reaprovechamiento para otras investigaciones.

4. CONCLUSIÓN

Tratar de las Humanidades usando el prisma del espacio afecta tanto la reflexión epistemológica, la metodología analítica y los datos representados de la investigación. En este trabajo, las metodologías aplicadas expuestas ponen el espacio en el centro de un diálogo entre estos tres componentes en sus proyectos de investigación. A través de sus diferentes visualizaciones, abren perspectivas novedosas sobre sus datos de estudio, que permiten no sólo analizarlos para contestar a las preguntas de investigación sino también crear preguntas nuevas. EPYW permite así dibujar el paisaje de un programa intelectual-científico-religioso de modernización de un estado y sus dinámicas, organizadas entre polos de difusión y agencia de sus diferentes actores a través la producción impresa de libros de tradición literaria del Yemen entre 1850 y 1950. Asimismo, pone de relieve los mecanismos de esta agencia, su evolución en el tiempo y aborda nuevas preguntas sobre los intereses que tiene aquella según el subgrupo de actores en cuestión. El visualizador de variantes lingüísticas de los papiros griegos permite no sólo interrogar los datos sobre la evolución y distribución de los fenómenos lingüísticos atestiguados, sino también contextualizarla. Permite considerar de ese modo tanto factores sociolingüísticos como el cuadro de la transmisión de las fuentes, los papiros, a la cual está sometida la muestra de los datos estudiados. Así es posible matizar cualquier cálculo cuantitativo por la visualización de las irregularidades de representación, geográficas y temporales, de la realidad lingüística que estudia por las fuentes. Por último, la representación geográfica de topónimos y gentilicios de obras filológicas orientalistas ofrece representaciones novedosas del clima de erudición de cada autor, ofreciendo

perspectivas valiosas sobre la procedencia de sus fuentes y rompe la frontera entre espacio físico y espacio conceptual, mapeando el propio modelo de conocimiento de cada uno.

FUENTES DE FINANCIACIÓN

Los trabajos presentados en este artículo se han realizado en el marco de los siguientes proyectos: (3.1) Proyecto Intramural 202210I17, financiado por el CSIC, (3.2) Proyecto PID2020-118094GB-I00, «Lingüística papirológica y humanidades digitales (LPDH)», financiado por MCIN/AEI/10.13039/501100011033, y (3.3) Proyecto PID2021-126158NB-I00, «Orientalismo en España, ss. XVII-XVIII: de la polémica religiosa a la institucionalización (ORES)», financiado por MCIN/AEI/10.13039/501100011033 y por FEDER Una manera de hacer Europa.

REFERENCIAS BIBLIOGRÁFICAS

- BERENGUER-SÁNCHEZ, José Antonio, “Nuevas herramientas digitales para el estudio de la koiné griega de Egipto”, en *Δῶρα τὰ οἱ δίδομεν φιλέοντες. Homenaje al profesor Emilio Crespo*, Luz CONTI et al. (eds.), Madrid, Universidad Autónoma de Madrid, 2020, pp. 75-83.
- BERENGUER-SÁNCHEZ, José Antonio, “Humanidades digitales, lingüística papirológica y lingüística griega”, en *Forum Classicorum: perspectivas y avances sobre el Mundo Clásico*, Jesús DE LA VILLA et al. (eds.), Madrid, Guillermo Escolar Editor, 2021, pp. 193-199.
- BERENGUER-SÁNCHEZ, José Antonio, “Vowel hiatus resolution in Koine Greek: the evidence of spelling variations for the endings -ιος, -ιον, -ίου, -ίω in documentary papyri”, *Journal of Greek Linguistics*, 23, 1, (2023a), pp. 3-35, DOI: <https://doi.org/10.1163/15699846-02301003>.
- BERENGUER-SÁNCHEZ, José Antonio, “Resolución de hiato vocálico en la koiné griega: la secuencia [i.a] en los papiros documentales”, *Cuadernos de Filología Clásica: estudios griegos e indoeuropeos*, 33, (2023b), pp. 129-149, DOI: <https://doi.org/10.5209/cfcg.84960>.
- BOCHART, Samuel, *Geographia Sacra seu Phaleg et Canaan*, Lugduni Batavorum [Leiden], apud Cornelium Boutesteyn et Jordanum Luchtmans, 1646.

- BROWNING, Robert, *Medieval and Modern Greek*, Cambridge, Cambridge University Press, 1983, DOI: <https://doi.org/10.1017/CBO9780511554148>.
- DEJORDY, Rich, Stephen P. BORGATTI, Chris ROUSSIN y Daniel S. HALGIN, “Visualizing Proximity Data”, *Field Methods*, 19, 3, (2007), pp. 239–263, DOI: <https://doi.org/10.1177/1525822X07302104>.
- DELEUZE, Gilles, *Logique Du Sens*, Paris, Les Éditions de Minuit, 1969.
- DEPAUW, Mark y Joanne STOLK, “Linguistic Variation in Greek Papyri: Towards a New Tool for Quantitative Study”, *Greek, Roman and Byzantine Studies*, 55, (2015), pp. 196–220.
- DICKEY, Eleanor, “The Greek and Latin Languages in the Papyri”, en *The Oxford Handbook of Papyrology*, Roger S. BAGNALL (ed.), Oxford, Oxford University Press, 2009, pp. 149–169, DOI: <https://doi.org/10.1093/oxfordhb/9780199843695.013.0006>.
- EL SHAMSY, Ahmed, *Rediscovering the Islamic Classics: How Editors and Print Culture Transformed an Intellectual Tradition*, Princeton, Princeton University Press, 2020, DOI: <https://doi.org/10.2307/j.ctvp2n411>.
- HAYKEL, Bernard, *Revival and Reform in Islam: The Legacy of Muhammad al-Shawkānī*, Cambridge, Cambridge University Press, 2003.
- HORROCKS, Geoffrey C., *Greek: A history of the language and its speakers*, Malden, Blackwell, 2010, DOI: <https://doi.org/10.1002/9781444318913>.
- KORZYBSKI, Alfred, *Science and sanity: An introduction to non-Aristotelian systems and general semantics*, Brooklyn, NY, Institute of General Semantics, 2005.
- KÜMIN, Beat y Cornelia USBORNE, “At Home and in the Workplace: A Historical Introduction to the ‘Spatial Turn’”, *History and Theory*, 52, 3, (2013), pp. 305–318, DOI: <https://doi.org/10.1111/hith.10671>.
- LAKOFF, George y Mark JOHNSON, *Metaphors We Live By*, Chicago, University of Chicago Press, 2003, DOI: <https://doi.org/10.7208/chicago/9780226470993.001.0001>.
- MESSICK, Brinkley, *The Calligraphic State: Textual Domination and History in a Muslim Society*, Berkeley, University of California Press, 1993.
- MESSICK, Brinkley, “Al-Khizāna al-Mutawakkiliyya: The Imam’s New Library”, en *Yemeni Manuscript Cultures in Peril*, Hassan ANSARI y

- Sabine SCHMIDTKE (eds.), Piscataway, NJ, Gorgias Press, 2022, pp. 149-170, DOI: <https://doi.org/10.31826/9781463242039-003>.
- NIGST, Lorenz, Maxim ROMANOV, Sarah B. SAVANT, Massoumeh SEYDI y Peter VERKINDEREN, “OpenITI: A Machine-Readable Corpus of Islamicate Texts (2020.2.3) [Dataset]”, *Zenodo*, 2020, DOI: <https://doi.org/10.5281/ZENODO.4075046>.
- REGNAULD, Hervé, “Les concepts de Félix Guattari et Gilles Deleuze et l’espace des géographes”, *Chimères*, 76, 1, (2012), pp. 195-204, DOI: <https://doi.org/10.3917/chime.076.0195>.
- ROLLET, Nadège y Jan THIELE, “Yemeni books printed between 1850 and 1950 [Dataset]”, *DIGITAL.CSIC*, 2023, DOI: <https://doi.org/10.20350/digitalCSIC/15701>.
- ROSSI, Ettore, “La Stampa nel Yemen”, *Oriente Moderno*, 18, 10, (1938), pp. 568-580.
- SHALEV, Zur, *Sacred words and worlds: geography, religion, and scholarship, 1550-1700*, Leiden, Brill, 2012, DOI: <https://doi.org/10.1163/9789004209381>.
- SOSIN, Joshua D., “Digital Papyrology [lightly electronic edited version of a talk that delivered at the 26th Congress of the International Association of Papyrologists, 19 August 2010, in Geneva]”, en *The Stoa: a Review for Digital Classics*, 2010, <<https://blog.stoa.org/archives/1263>> [consulta 28/02/2024].

CUANDO EL RETO NO ES DESCUBRIR YACIMIENTOS: INTEGRACIÓN SEMÁNTICA Y CARACTERIZACIÓN HISTÓRICA DE PAISAJES SERRANOS CON LiDAR

Sonia Medina Gordo

Resumen

En estas últimas dos décadas, hemos sido testigos de una creciente familiarización por parte de la comunidad académica para con los datos LiDAR. En ámbitos como la Arqueología, este acercamiento se ha fomentado con el fin de determinar el potencial que este tipo de datos poseen para el análisis histórico de los paisajes. Fruto de estas dinámicas, algunos autores han comenzado a plantearse el tipo de conocimiento que somos capaces de generar a raíz de su gestión, especialmente dentro de ciertos contextos de trabajo, como aquellos que propone la Ciencia Abierta o el denominado *Big Data*. Los trabajos desarrollados en torno al estudio de la Edad Media no son ajenos a esta realidad. También en ellos podemos apreciar un compromiso para la renovación metodológica y mejora interpretativa de esta etapa del pasado. No obstante, la naturaleza escasamente monumental de muchos de los vestigios que recuperamos apunta a la necesidad de plantearse nuevas vías con las que incorporar un aprovechamiento íntegro de los datos LiDAR, con independencia de la disciplina o la base empírica desde donde proyectemos su uso. En este trabajo, ahondaremos en cada una de estas cuestiones con el objetivo de definir y evaluar un procedi-

Sonia Medina Gordo, Universitat de Barcelona
somego12@gmail.com
ORCID: 0000-0002-3921-3025

miento de trabajo alternativo de la información generada con tecnología LiDAR. Como caso de estudio, tomaremos el paisaje de la Sierra de Gredos (España) y su desarrollo histórico a lo largo de la Edad Media.

Palabras clave: *Edad Media, construcción de conocimiento, mapeo de suelos, paisaje de montaña, Península Ibérica.*

Abstract

In the past two decades, we have witnessed a growing familiarity within the academic community towards LiDAR data. In fields such as Archaeology, this approach has been encouraged to ascertain the potential that this type of data holds for the historical analysis of landscapes. As a result of these dynamics, some authors have begun to contemplate the kind of knowledge we can generate through its management, especially within certain work contexts, such as those advocated by Open Science or the so-called Big Data. The endeavours undertaken in the study of the Middle Ages are not exempt from this reality. In them, we can also observe a commitment to methodological renewal and interpretative improvement of this historical period. However, the relatively non-monumental nature of many of the vestiges recovered underscore the need to explore new avenues for incorporating a comprehensive employment of LiDAR data, regardless of the discipline or empirical basis from which we project its use. In this work, we will delve into each of these issues with the aim of defining and assessing an alternative working procedure for the information generated with LiDAR technology. As a case study, we will focus on the landscape of the Sierra de Gredos (Spain) and its historical development throughout the Middle Ages.

Keywords: *Middle Ages, knowledge construction, soil mapping, mountain landscape, Iberian Peninsula.*

1. INTRODUCCIÓN

Desde hace un par de décadas, el aprovechamiento de los datos LiDAR ha sido integrado en el día a día de algunas disciplinas del ámbito de las Humanidades y Ciencias Sociales, como lo es la Arqueología. En estos contextos, una de las cualidades mejor valoradas es la representación de la altura o elevación del objeto de estudio (Herzog, 2022). En los estudios que analizan el denominado *paisaje histórico* (Bastian, Walz y Decker, 2013, p. 443-444), partiendo de su materialidad, la utilización de datos LiDAR ha planteado no pocos problemas, en ocasiones debido a la falta de especialización de los investigadores que los habían de manejar. En un afán por superar estas dificultades, nos hemos adentrado en un proceso de reflexión teórica sobre la naturaleza de estos datos dentro de acercamientos como el que propone la Arqueología Aérea. En este trabajo remoto, con un perspectiva ‘desde arriba’ de los yacimientos, se ha querido acotar unos estándares acerca de cómo deberían ser estos datos para fines arqueológicos, así como también qué entornos y en qué condiciones geográficas resulta verdaderamente útil su uso. Este fue el principal foco de interés de algunos proyectos concebidos a la luz de este panorama, como lo fue MALiAP¹.

Tomando como base la inercia de estos avances, algunos equipos de investigación han seguido potenciando esta línea de conocimiento y ampliándola hacia otros campos, como la Inteligencia Artificial (IA), por medio de acercamientos basados en el aprendizaje automático o *Machine Learning* (ML). Estas perspectivas están ofreciendo en la actualidad una amplia variedad de resultados y propuestas (Verschoof-van der Vaart y Lambers, 2019; Orengo *et al.*, 2020; Fiorucci *et al.*, 2022; Mazzaca *et al.*, 2022). Pero, por otro lado, también han planteado otro tipo de necesidades re-

1 *Methodological Maturity of Airborne LiDAR in Archaeology - Pilot Project*. El proyecto se desarrolló en el Instituto de Arqueología ZRC SAZU (Eslovenia) entre los años 2019-2022, y su investigador principal fue Benjamin Štular (ZRC SAZU). Los resultados del mismo se pueden consultar en su página web: <https://iza2.zrc-sazu.si/en/programi-in-projekti/methodological-maturity-airborne-lidar-archaeology-pilot-project-maliap> [consultado: 23/01/2024].

lacionadas con el mundo del Big Data y los estándares de desarrollo de proyectos a nivel europeo, particularmente en términos de transparencia y acceso en abierto de la información derivada. En este sentido, la preocupación de quienes utilizan los datos gira en torno a cómo hacer de sus resultados un tipo de información accesible y reusable para otros usuarios, sean o no arqueólogos experimentados. A tal respecto, ya existen propuestas y proyectos desarrollados dentro de iniciativas como SEADDA² o TETRARCHS³.

Dados los objetivos que se proponen cada uno de estos planes, creemos que el escenario es lo suficientemente favorable como para alcanzar un compromiso epistemológico de los datos LiDAR para con otras áreas de conocimiento, más allá del campo de la Arqueología. Como ya habíamos avanzado, las primeras décadas de uso de este tipo de datos en estudios arqueológicos y de paisaje se focalizaron en su utilidad y prestaciones. Ambos aspectos terminaron por poner en valor el aprovechamiento de la información generada con tecnología LiDAR para el *descubrimiento* de entidades patrimoniales que eran difícilmente detectables mediante fotografía aérea al quedar cubiertas bajo un manto de vegetación. Esta otra perspectiva de interpretación a distancia, aunque innovadora por entonces, ha terminado por limitar el radio de acción a situaciones muy concretas, donde la característica fundamental del

- 2 *Saving European Archaeology from the Digital Dark Age*. El programa, cuyo inicio podemos situar en el año 2019, prioriza aquellos aspectos relacionados con el archivo, la divulgación y la reutilización del dato arqueológico. El coordinador principal de la comunidad de investigadores creada en torno a estas acciones es Julian D. Richards (University of York). Se puede consultar más información en su página web: <https://www.seadda.eu/> [consulta: 23/01/2024].
- 3 *Transforming Data Re-use in Archaeology*. Este proyecto se centra en el ciclo de vida del dato arqueológico y la creación de nuevas narrativas que nos permitan su difusión. Bajo la dirección de Sara Perry (MOLA), se desarrollará entre los años 2022-2025. Con motivo de explorar los diferentes escenarios posibles, se organizaba en febrero de 2023 un seminario que abordó el tema de cómo hacer accesible la investigación arqueológica con datos LiDAR a cualquier usuario, independientemente de su nivel de conocimiento sobre el tema. Más información en: <https://www.tetrarchs.org/> [consulta: 23/01/2024].

objeto de estudio reside en una evidente monumentalidad de los restos materiales conservados en superficie (por ejemplo, Inomata *et al.*, 2020 o Fontana, 2022).

Aunque la situación parece que está cambiando, todavía son excepcionales las propuestas de análisis (Masini y Lasaponara, 2021; Lozić, 2021; Doneus, Doneus y Cowley 2022; Doneus *et al.*, 2022) que consideran otro tipo de perspectivas, más allá de la prospectiva, como pueda ser la histórica (Mathieu, 2021) o la cronométrica (Barceló y Andreaki, 2020). Esta constatación supone un problema para el avance interpretativo de muchos espacios, ya que no se termina de incorporar la noción de *temporalidad* necesaria para comprender ciertos procesos sobre el paisaje definidos a partir una interrelación de eventos. Asimismo, restringir el análisis a la búsqueda de elementos topográficos singulares sobre el terreno, como pueden ser unos restos constructivos para una buena parte de los estudios arqueológicos, limita la capacidad real que tienen este tipo de datos a la hora de caracterizar un territorio (Herzog, 2022, pp. 130-139).

En la actualidad, creemos que la proliferación de proyectos centrados en la recogida de datos sobre la geografía peninsular ibérica, así como el encuadramiento de estos mismos datos dentro de las líneas trazadas por los principios LOUD y FAIR (Thiery, 2019), nos sitúa en un escenario idóneo para avanzar en la generación de información nueva; en el particular de nuestro estudio, relacionando los datos LiDAR con otros de diferente naturaleza a través de un esquema conceptual común. Por tanto, en este trabajo pretendemos una integración de los conocimientos vertebraados por diferentes disciplinas considerando el potencial de los datos LiDAR y la perspectiva de análisis que permiten los estudios de corte histórico. Estos objetivos nos han permitido generar un procedimiento de trabajo con los datos alternativo a las propuestas actuales (Štular y Lozić, 2023). De igual modo, tomando como base esta proyección interpretativa, examinamos los resultados obtenidos en paralelo a la caracterización histórica de un espacio de montaña; en este caso, la Sierra de Gredos entre los siglos V-XI.

2. LOS DATOS LIDAR Y EL ESTUDIO DEL PASADO

Recientemente, se publicaba una nota de prensa en el diario *The Guardian* en la que se valoraba la incidencia que había tenido la tecnología LiDAR en las investigaciones arqueológicas. Entre otras cuestiones, el artículo enfatizaba en aspectos como la rapidez y la calidad con la que dicha tecnología permitía documentar grandes extensiones de territorio, a unos precios cada vez más accesibles. Si bien de manera inconsciente, los ejemplos que acompañaban a la noticia terminaban presentando un potencial de análisis en el que los datos generados podían ser la mejor y la más rápida forma de descubrir nuevas historias del pasado antes de su desaparición (Manaugh, 2023). Este caso no deja de ser un simple, pero ilustrativo ejemplo de los diversos reportajes que podemos encontrar acerca del uso que hacemos de los datos LiDAR, y que refleja con bastante claridad el imaginario que se ha ido generando en torno a ellos.

Precisamente, dicho imaginario es el resultado del proceso de maduración de este tipo de datos para el estudio del pasado; un proceso que, por otra parte, debemos vincular con los avances que ha estado experimentando la Arqueología a lo largo del tiempo. Así, en una etapa en la que los métodos más tradicionales de prospección no intrusiva, como ha sido el caso de la fotografía aérea, limitaban la exploración de zonas cubiertas por densas capas boscosas, los datos LiDAR se presentaban como una propuesta novedosa que permitía superar con cierto éxito dichos obstáculos. Además, su puesta en práctica no sólo permitía obtener la imagen de toda superficie terrestre cubierta por árboles, sino que también se mostraba eficaz sobre otro tipo de paisajes. De esta manera, el uso de estos datos ofrecía mayor seguridad a la hora de plantear proyectos, ya que parecía garantizar la obtención de la información patrimonial de un territorio con independencia de las condiciones topográficas y ecológicas del mismo.

Dadas las positivas experiencias que se obtuvieron durante esos primeros años, no es de extrañar que los proyectos pioneros centrados exclusivamente en el uso de esta tecnología

tuvieran como objetivo una optimización de las visualizaciones que se podían generar y aprovechar para fines arqueológicos⁴. Sin embargo, estos avances también respondían a las dificultades a las que se enfrentaba una buena parte de los investigadores, ya que la identificación e interpretación de las entidades arqueológicas no siempre era una tarea sencilla de acometer. Por otro lado, aún eran escasos los estudios cuyo procedimiento diera solución a los problemas y necesidades propios del trabajo arqueológico, lo que acrecentaba la curva de aprendizaje en contextos altamente especializados e impedía un aprovechamiento íntegro del potencial que presentan los datos LiDAR. De esta modo, la situación requería de un periodo de formación en el que se pudieran establecer tanto características y procedimientos (Lozić y Štular 2021) como guías sobre las que apoyarse durante el proceso interpretativo de los resultados (Kokalj y Hesse, 2017).

En el presente, esta creciente familiarización es lo que está posibilitado un afianzamiento del potencial de los datos para su uso en el análisis histórico de los paisajes. Es por eso por lo que las perspectivas de trabajo actuales se están orientado a resolver otras cuestiones, como el tipo de conocimiento que somos capaces de generar. En este sentido, la respuesta por parte de la comunidad académica se debe entender dentro de los nuevos contextos de *Open Science* y *Big Data*, a la que los trabajos que usan datos LiDAR no son ajenos, tal y como parecen evidenciar los resultados más recientes (Štular y Lozić 2023). Sobre estos aspectos, Verschoof-van der Vaart (2022: 76–81) ha ofrecido una reflexión crítica partiendo de una perspectiva de trabajo basada en la extracción activa de infor-

4 Este fue el caso del proyecto postdoctoral *Investigation of micro-relief structures with airborne lidar*. Como en el antes citado caso de MALIAP, esta investigación también se desarrolló en el Instituto de Arqueología ZRC SAZU (Eslovenia); si bien, con anterioridad, en un periodo comprendido entre los años 2011–2013. En este caso, el autor de dicho estudio fue Žiga Kokalj (ZRC SAZU), y los detalles del citado proyecto pueden consultarse en su página web: https://iaps.zrc-sazu.si/en/search_content?f%5B0%5D=content_type%3Aprojekt [consulta: 23/01/2024].

mación. En su planteamiento, el autor ha propuesto dos maneras diferentes por medio de las cuales se puede generar conocimiento; esto es, mediante la obtención de conocimiento cuantitativo y mediante la obtención de conocimiento cualitativo. Ambos acercamientos no son mutuamente exclusivos entre sí, pero sí se proyectan sobre unas metodologías y unos objetivos opuestos.

El enfoque cuantitativo se fundamenta en la generación de información *nueva*, en la medida en que los datos que aprovechamos no se conocían antes de realizar un estudio. Esta manera de afrontar el desconocimiento (Huggett, 2020) se centra en los problemas que presentan los factores epistemológicos, como puede ser la ausencia de estos mismos datos (Wylie, 2008: 184-185). Como se puede deducir, esta perspectiva de trabajo se integra sin excesivos problemas al proceso que venimos observando, ya que atiende nuevamente los requerimientos de exploración que comentábamos al principio. Por su parte, el planteamiento cualitativo se apoya en información ya existente, tratando de *entenderla*. En términos comparativos, la forma en la que se hace frente a los aspectos que no conocemos toma como base la propia complejidad de lo que cada uno entiende por “información”, y la exploración de lo desconocido va más allá de una mera ausencia de datos. Por eso, los problemas a los que esta otra vía intenta dar respuesta tienen que ver con limitaciones ontológicas (Wylie, 2008: 185), entre otras.

Dentro de este contexto de estudios en base a la tecnología LiDAR, parece estar teniendo mayor acogida y desarrollo la obtención de conocimiento cuantitativo. Sin embargo, este impulso lleva consigo un encorsetamiento de la misma dentro la Arqueología, dificultando el surgimiento de nuevas ideas planteadas desde otras disciplinas interesadas también en el pasado. Además, a largo plazo esta perspectiva cuantitativa supone una inmersión plena en lo que Levent O. Orman ha denominado la “paradoja de la información” (2015). En cambio, el marco de actuación cualitativo creemos que permite una mayor versatilidad, y sus procedimientos tienen también cabida en muchos de los problemas que se plantean

en los contextos de investigación actuales. Una parte de estos procedimientos es el que queremos presentar en este trabajo, abriendo con ello una línea de debate en cuanto a su potencial, sus limitaciones y, en cualquier caso, para deliberar sobre cómo inciden estos elementos en las Humanidades Digitales habida cuenta de que se trata de una rutina de trabajo tradicionalmente asociada a la ahora denominada Arqueología Digital (Watrall, 2016: 346–349).

3. LA SIERRA DE GREDOS DURANTE LA EDAD MEDIA

Formando parte del Sistema Central, la Sierra de Gredos (Fig. 1) es un macizo montañoso que se prolonga en dirección E-O desde la fosa del Tiétar hasta limitar con la Sierra de Béjar, con una longitud cercana a los 140 km. Este amplio sector destaca por su relieve complejo, mayoritariamente de base granítica y con unos desniveles acusados que llegan a alcanzar cotas máximas superiores a los 2500 m. s. n. m (en el pico de Almanzor) y unas mínimas cercanas a los 500 m. s. n. m. Esta morfología heterogénea, pero generalmente elevada, dispone que el paisaje se defina a partir de una climatología variable y diferenciada a lado y lado de la serra-nía. Pese a que a grandes rasgos podemos decir que se inserta en un entorno húmedo y de temperaturas bajas, las precipitaciones son ligeramente más abundantes en el sector meridional. Estos elementos, que a su vez originan una diversidad forestal y agrícola, junto con el sustrato litológico marcado por unos suelos ácidos y metamórficos, hacen de su paisaje un escenario para el desarrollo de dinámicas sociales potencialmente diferenciadas de las penillanuras circundantes.

Es por ello por lo que, desde un punto de vista interpretativo, este entorno montañoso cuenta con un variado número de estudios que han buscado su caracterización para el periodo medieval sirviéndose de datos, perspectivas y cronologías dispares. Precisamente, partiendo de esta variedad informativa es posible hoy la construcción de unos relatos históricos y unos conoci-

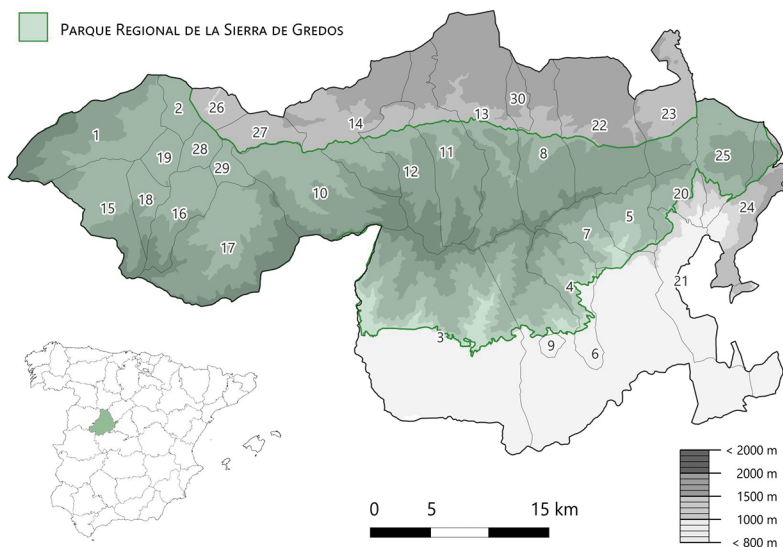


Figura 1. Mapa de la zona de estudio. El área total de la Sierra de Gredos abarca parcial o totalmente los siguientes municipios: 1. Solana de Ávila 2. La Carrera 3. Candeleda 4. Guisando 5. El Arenal 6. Arenas de San Pedro 7. El Hornillo 8. Hoyos del Espino 9. Poyales del Hoyo 10. Bohoyo 11. Navalperal de Tormes 12. Zapardiel de la Ribera 13. San Juan de Gredos 14. Santiago del Tormes 15. Puerto Castilla 16. Nava del Barco 17. Navalenguilla 18. Gil García 19. Umbrías 20. Cuevas del Valle 21. Mombeltrán 22. Navarredonda de Gredos 23. San Martín del Pimpollar 24. San Esteban del Valle 25. Villarejo del Valle 26. El Barco de Ávila 27. Los Llanos de Tormes 28. Navatejares 29. Tormellas 30. Hoyos del Collado. Fuente: elaboración propia

mientos sobre la serranía gredense en múltiples ámbitos que, en cualquier caso, todavía no están plenamente integrados. Así, por ejemplo, J. A. López-Sáez *et al.* (2009), enmarcando su análisis en las perspectivas de trabajo que nos ofrecen las denominadas *eco-dinámicas humanas* (Fitzhugh *et al.*, 2019), trata de definir el papel que desempeñaron las actividades pastoriles en la configuración del paisaje de Gredos. El uso del registro paleoambiental les permitió, no solamente establecer el proceso de deconstrucción de la narrativa del entorno ya desde la Prehistoria, sino también definir el carácter ‘marginal’ del paisaje, al menos desde un punto de vista económico (Bailey, 1989; Turner y Young, 2007).

También contamos con experiencias que tratan de recuperar la historia de Gredos desde un punto de vista social. Tomando esta perspectiva, I. Martín Viso (2018) ha propuesto una síntesis histórica del territorio apoyándose en la materialidad de la zona, complementada con las evidencias escritas. Un aspecto sobre el que llama la atención el autor es la deforestación de los paisajes de montaña durante los primeros siglos medievales, lo cual sería un indicio del aprovechamiento de estos espacios en el desarrollo de las actividades de ganadería. Pero, también, como zonas a las que las comunidades de la época acudirían para su subsistencia, dada la intensa ocupación de del territorio que evidencian los restos materiales. En conjunto, la información le ha permitido percibir una mayor resiliencia de las comunidades campesinas tras el colapso del sistema político visigodo, a diferencia de las élites, cuyo poder empezaría a regenerar muy lentamente sólo a partir del siglo X.

En lo tocante a la información LiDAR, en trabajos anteriores en los que nos centramos en la caracterización arqueológica de éste y otros espacios serranos colindantes, ya habíamos incidido en las limitaciones de los datos para incrementar el conocimiento sobre la materialidad de esta zona (Medina Gordo, 2019). Por un lado, no pudimos disponer de la calidad mínima recomendada (Doneus y Briese, 2010) para desarrollar una prospección arqueológica no intrusiva del territorio; una exploración, en nuestro caso, de carácter selectivo, en las inmediaciones de cada yacimiento. Tampoco pudimos contar con un registro de fuentes de calidad con las que retroalimentar el uso interpretativo de los datos LiDAR, ya que buena parte de los asentamientos analizados presentaban un alto grado de incertidumbre tanto a nivel de ubicación, datación, representación planimétrica como adscripción tipológica (Medina Gordo, 2019: 72-75). Por lo que, en el caso de esta investigación, el reto para desarrollar nuevas vías de investigación, y profundizar en el pasado de Gredos, recae, entre otras cuestiones, en implementar procedimientos alternativos para el aprovechamiento de estos mismos datos.

4. DE LA VISUALIZACIÓN AL MODELADO: UN CAMBIO EN EL ENFOQUE METODOLÓGICO

Tal y como se ha comentado, uno de los primeros desarrollos que experimentó el uso de datos LiDAR para el estudio del pasado fue una propuesta de mejora de sus visualizaciones. Desde esta perspectiva, el manejo de estos datos en contextos arqueológicos se ha centrado, en gran medida, en la generación de Modelos Digitales del Terreno (MDT), para los cuales se descarta la información perteneciente a la vegetación, que es aquella que impide observar, “desde arriba”, la materialidad de un paisaje. Estos modelos destacan por su capacidad para ofrecer diversas formas de representar la superficie terrestre, un aspecto que ha despertado gran interés debido a su accesibilidad y utilidad para observar y comunicar nuestros objetos de interés. Sin embargo, aunque las visualizaciones son una herramienta poderosa, no deben considerarse el objetivo principal. Más bien, representan el resultado final de un proceso mucho más amplio, estructurado en un flujo de trabajo que abarca múltiples decisiones y etapas.



Figura 2. Flujo de trabajo para el procesamiento de datos LiDAR.
Fuente: elaboración propia

Este procedimiento (**Fig. 2**) es el que recientemente se ha estado mejorando para responder mejor a las necesidades genuinas de la arqueología, con el fin de superar las limitaciones que potencialmente se presentan en cada fase. Por un lado, el uso de datos públicos supone una escasa capacidad de control sobre la calidad de la información que vamos a procesar. Pese a que el tema es cada vez mejor conocido entre los investigadores menos experimentados, conviene insistir en que la resolución de imagen va a depender en gran medida de la densidad de puntos con la que trabajemos, cuya definición debemos vincular a varios parámetros que respon-

de al tipo de vuelo y sensor utilizados. Las administraciones públicas han realizado un esfuerzo notable al poner a disposición datos de acceso libre que abarcan extensiones de terreno muy amplias, lo que supone un recurso valioso para la investigación. Sin embargo, esta amplitud de cobertura suele lograrse a costa de una menor precisión en la adquisición de los datos. En este sentido, una de las alternativas que se han comenzado a poner en práctica es el uso de drones (Risbøl y Gustavsen, 2018).

Otro aspecto relevante reside en el filtrado y la clasificación de los puntos. Para autores como Lozić y Štular (2021) esta es una fase vital de todo proyecto, ya que es en este momento en el que asignamos los puntos a unas categorías que incluyen desde el suelo hasta árboles y edificios. El tipo de modelo digital que vayamos a generar se basará precisamente en esta clasificación, lo cual ha llevado a plantear los problemas que conlleva la simple discriminación entre los puntos correspondientes al suelo de los demás. En los trabajos de prospección aérea, el principal objetivo es ser capaz de identificar anomalías sobre el territorio. Estas anomalías, en fotografía, suelen detectarse a través de lo que se conocen como *cropmarks*. En cambio, en el caso de los MDT es esencial que exista una estructura visiblemente reconocible sobre el relieve. Es por ello por lo que las propuestas más recientes (Štular, Lozić y Eichert, 2021) apuestan por un procedimiento que permita conciliar la información que tenemos del suelo y de los microrelieves.

Por último, también se ha de tener en cuenta los aspectos relacionados con la interpolación. Independientemente del modelo que queramos generar, en éstos es necesario considerar el tipo de terreno sobre el que trabajamos, ya que una triangulación de los puntos errónea puede llevar a suavizar zonas y, con ello, *eliminar* de la imagen las entidades patrimoniales (presumiblemente) conservadas. La correcta realización de cada uno de estos pasos es lo que garantiza el éxito de la visualización de la superficie. En este sentido, por tanto, podemos destacar que el núcleo central del trabajo se concentra en estos tres primeros pasos. Basándonos en estas premisas, proponemos una vía alternativa cuyo fin no sea el

de crear una visualización, sino el de *generar un modelo teórico* del territorio que nos permitan hipotetizar sobre otros aspectos que componen el paisaje y su incidencia en los procesos históricos.

4. 1. CREACIÓN DE UN MODELO TEÓRICO DEL ESPACIO

El modelo teórico que proponemos se enmarca en las formas en que las sociedades medievales concebían la propiedad de la tierra. El final del Imperio romano llevó consigo un cambio en la forma de entender y definir la propiedad, ya que es una etapa en la que se va a perder el sentido de fiscalidad. Esta problemática ha llevado a la exploración de otros aspectos que nos permitan hablar de los procesos de apropiación, ligados a nociones como la cotidianidad o la tradición. Este punto de vista, sugiere un panorama construido en torno a una idea de gestión comunitaria, cuyo protagonismo recaería en las comunidades campesinas. Siguiendo esta línea, M. Fernández Mier (2021) ha apuntado a la definición de una serie de reglas que estos sectores de la sociedad generaban para la gestión del territorio, tomando en consideración sus posibilidades ecológicas.

La idea de unas comunidades conocedoras del potencial ecológico de su entorno es un planteamiento sugerente e innovador, ya que permite mover el foco de atención de las estructuras sociales al medio físico, y realizar una evaluación contrastada de esta información con los datos que recogen los textos o el registro arqueológico, tal y como muestran otros estudios sobre el periodo (Blanco-González y Martín Viso, 2016; Martín Viso y Blanco-González, 2016). Para generar este aparato empírico, en este estudio nos hemos basado en el marco de trabajo propuesto para determinar la capacidad agrológica de un suelo. Para ello, es necesario realizar una clasificación, en cuyo proceso se evalúa las capacidades de la tierra para diversos usos, como la agricultura, la silvicultura o la conservación. Este método ayuda a entender las limitaciones y los potenciales de la tierra en términos de productividad y sostenibilidad. La generación de dicha clasificación se basa en parámetros diferentes. En nuestro caso, hemos utilizado los datos LiDAR para resaltar los rasgos morfológicos del terreno, y complementado con

información sobre suelo publicada por el Instituto Tecnológico Agrario de Castilla y León (ITACyL) (Fig. 3).

A partir de este procedimiento, hemos generado un modelo teórico del paisaje de Gredos en el que se definen las unidades de suelo arables (clases II y III, Fig. 4a), y las que no lo son (clases IV, V y VI, Fig. 4a). El planteamiento nos ha permitido obtener unos resultados preliminares que, en líneas generales, podríamos dinamizar a partir de cuestiones propias de la retrodicción; es de-

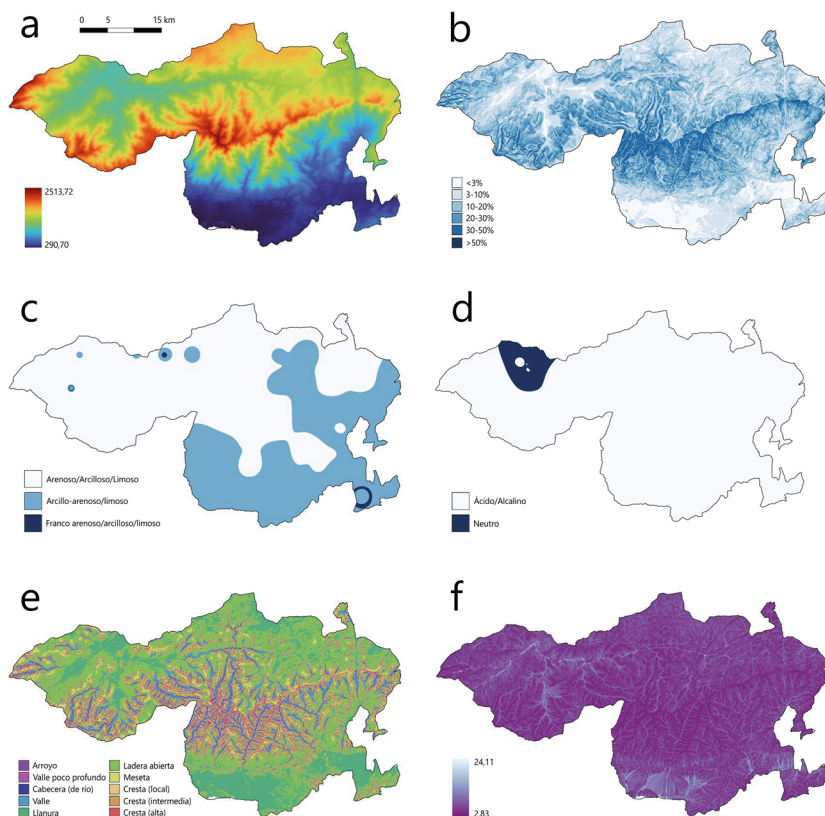


Figura 3. Los parámetros que se han utilizado en esta investigación son la pendiente (b), la textura (c) y el pH del suelo (d), el Índice de Posición Topográfica (TPI) (e) y el Índice de Humedad Topográfica (TWI) (f).

Fuente: elaboración propia

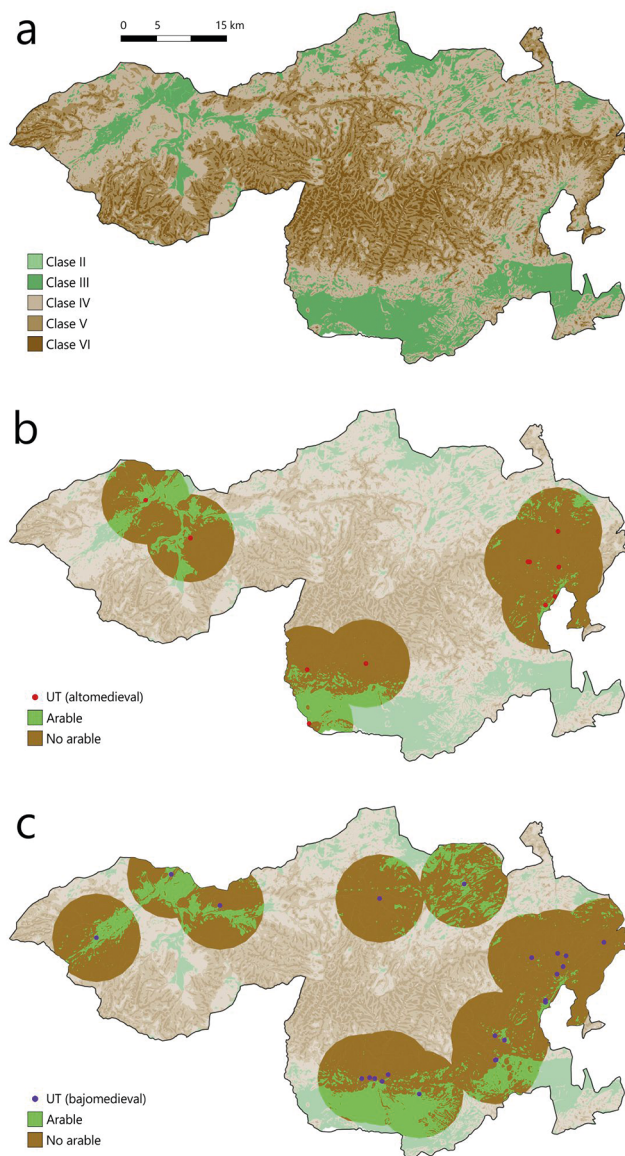


Figura 4. Clasificación de la capacidad agrológica del suelo de la Sierra de Gredos (a). Radio de acceso a los recursos más inmediatos (15km) de los asentamiento de la Alta Edad Media (b) y la Baja Edad Media (c).
Fuente: elaboración propia

cir, explicar qué ha podido ocurrir en el pasado (Quale, 2008). Así, en un primer acercamiento a los datos podemos apreciar una diferencia en cuanto a la ubicación de los asentamientos en función del tipo de suelo. En el caso de los asentamientos altomedievales, parece predominar los suelos con un equilibrio ecológico más frágil, lo que podría sugerir la necesidad de una gestión comunitaria que garantizara la vida útil de la tierra (Fig. 4b). Estos resultados abren interesantes perspectivas de análisis, especialmente en lo que respecta a las estrategias de adaptación de las comunidades altomedievales frente a un entorno con limitaciones ecológicas. Asimismo, el modelo generado cuenta con el potencial de ser una herramienta para formular nuevas hipótesis sobre la interacción entre paisaje y estructura social.

5. CONSIDERACIONES FINALES

A lo largo de este trabajo hemos presentado las limitaciones que presenta el uso de los datos LiDAR como medio exclusivo para el descubrimiento de yacimientos. En este trabajo se ha querido presentar una propuesta alternativa para el aprovechamiento de la tecnología que los genera, el conocimiento de los cual podemos potenciar en el momento en que integramos otros de distinta naturaleza. En este sentido, la búsqueda de elementos patrimoniales aparentemente desconocidos muchas veces nos lleva a acudir a otras fuentes solamente para *validar* nuestra interpretación de estos mismos datos, lo que impide una correcta integración de conocimientos. La generación de conocimiento cualitativo permite superar estos inconvenientes, y en ello juega un papel importante la denominada *Slow Science* al fomentar una actitud matizada hacia los datos, en lugar de asumir que éstos son evidentes y generalmente objetivos.

REFERENCIAS BIBLIOGRÁFICAS

- BAILEY, Mark, “The concept of the margin in the medieval English economy”, *Economic History Review*, 42, (1989), pp. 1-17.
- BARCELÓ, Juan Antonio y ANDREAKI Vasiliki, “Cronoestratigrafía. Tiempo y espacio en la excavación arqueológica”, en *Métodos cronométricos en Arqueología, Historia y Paleontología*, Juan Antonio Barceló y Berta Morell (eds.), Madrid, Dextra, 2020, pp. 163-198.
- BASTIAN, Olaf, Walz ULRICH y Annette DECKER, “Historical Landscape Elements: Part of our Cultural Heritage – A Methodological Study from Saxony”, en *The Carpathians: Integrating Nature and Society Towards Sustainability*, Jacek Kozak et al. (eds.), Berlín, Springer, 2013, pp. 441-459.
- DONEUS, Michael y Christian BRIESE, “Airborne Laser Scanning in forested areas – potential and limitations of an archaeological prospection technique”, en *Remote Sensing for Archaeological Heritage Management. Proceedings of the 11th EAC Heritage Management Symposium, Reykjavík, Iceland, 25-27 March 2010*, David C. Cowley (ed.), Bélgica, Europae Archaeologia Consilium (EAC), 2010, pp. 59-76.
- DONEUS, Michael, Nives DONEUS y Dave C. COWLEY, “Confronting Complexity: Interpretation of a Dry Stone Walled Landscape on the Island of Cres, Croatia”. *Land*, 11, (2022).
- DONEUS, Michael et al., “Stratigraphy from Topography II. The Practical Application of the Harris Matrix for the GIS-based Spatio-temporal Archaeological Interpretation of Topographical Data”, *Archaeologia Austriaca*, 106, (2022), pp. 223-252.
- FIORUCCI, Marco et al., “Deep Learning for Archaeological Object Detection on LiDAR: New Evaluation Measures and Insights”, *Remote Sensing*, 14, (2022).
- FITZHUGH, Ben et al., “Human ecodynamics: A perspective for the study of long-term change in socioecological systems”, *Journal of Archaeological Science*, 23, (2019).
- FONTANA, Giacomo, “Italy’s Hidden Hillforts: A Large-Scale Lidar-Based Mapping of Samnium”, *Journal of Field Archaeology*, 47, (2022), pp. 245-261.
- HERZOG, Irmela, “Towards Unleashing the Potential of Elevation Data for Archaeological Research. Reviewing Past and Present Appli-

- cations”, en *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies [online]*, November 2020, Wolfgang BÖRNER *et al.* (eds.), Heidelberg, Propylaeum, 2022, pp. 127–144.
- HUGGETT, Jeremy, “Capturing the Silences in Digital Archaeological Knowledge”, *Information*, 11, (2020).
- INOMATA, Takeshi *et al.*, “Monumental architecture at Aguada Fénix and the rise of Maya civilization”, *Nature*, 582, (2020), pp. 530–533.
- LÓPEZ-SÁEZ, José Antonio *et al.*, “Contribución paleoambiental al estudio de la trashumancia en el sector abulense de la Sierra de Gredos”, *Hispania*, 69, (2009), pp. 9–38.
- LOZIĆ, Edisa, “Application of Airborne LiDAR Data to the Archaeology of Agrarian Land Use: The Case Study of the Early Medieval Microregion of Bled (Slovenia)”, *Remote Sensing*, 13, (2021).
- LOZIĆ, Edisa y Benjamin ŠTULAR, “Documentation of Archaeology-Specific Workflow for Airborne LiDAR Data Processing”, *Geosciences* 11, (2021).
- MANAUGH, Geoff, “The big archaeological digs happening up in the sky”, *The Guardian*, (2023), <https://www.theguardian.com/us-news/2023/feb/23/lidar-technology-archeology-radical-thinking> [consulta: 23/01/2024].
- MARTÍN VISO, Iñaki, “Paisajes, comunidades y poderes centrales: el centro-oeste de la Península Ibérica durante la Alta Edad Media (s. VI–XI)”, *Arqueología y Territorio Medieval*, 25, (2018), pp. 195–226.
- MASINI, Nicola y Rosa LASAPONARA, “On the Reuse of Multiscale LiDAR Data to Investigate the Resilience in the Late Medieval Time: the Case Study of Basilicata in South of Italy”, *Journal of Archaeological Method and Theory*, 28, (2021), pp. 1172–1199.
- MATHIEU, Jon, “Is Historical Temporality “Heterogeneous” and “Contingent”? William H. Sewell’s Cultural Turn”, *Histories*, 1, (2021), pp. 12–21.
- MAZZACCA, Gabriele *et al.*, “Seeing among foliage with LiDAR and Machine Learning: Towards a transferable archaeological pipeline”, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 66, (2022), pp. 365–372.

- MEDINA GORDO, Sonia, *Arqueología de los Espacios Construidos. Estudio del paisaje serrano del centro-oeste peninsular durante la Alta Edad Media (siglos V-IX)* [Trabajo Final de Máster], Salamanca, Universidad de Salamanca, 2019.
- ORENGO, Héctor A., et al., “Automated detection of archaeological-mounds using machine-learning classification of multisensor and multitemporal satellite data”, *PNAS*, 117, (2020), pp. 18240-18250.
- QUALE, Andreas, “The Game of Prediction and Retrodiction: A Radical-Constructivist Perspective on the Notion of Time in Physics”, *The Journal of Educational Thought*, 42, (2008), pp. 105-125.
- RADŁO-KULISIEWICZ, Małgorzata, “Review of DTM derivatives most used in Digital Soil Mapping”, *Geomatics, Landmanagement and Landscape*, 2, (2021), pp. 87-99.
- RISBØL, Ole y Lars GUSTAVSEN, “LiDAR from drones employed for mapping archaeology – Potential, benefits and challenges”, *Archaeological Prospection*, 25, (2018), pp. 329-338.
- ŠTULAR, Benjamin y Edisa LOZIĆ, “Comparison of Filters for Archaeology-Specific Ground Extraction from Airborne LiDAR Point Clouds”, *Remote Sensing*, 12, (2020).
- ŠTULAR, Benjamin y Edisa LOZIĆ, “Executable Map Paper (EMaP) for Archaeological LiDAR”, *Journal of Computer Applications in Archaeology*, 6, (2023), pp. 79-95.
- ŠTULAR, Benjamin, Edisa LOZIĆ y Stefan EICHERT, “Airborne LiDAR-Derived Digital Elevation Model for Archaeology”, *Remote Sensing*, 13, (2021).
- THIERY, Florian, “Sphere 7 Data: LOUD and FAIR data for the research community”. *Squirrel Papers*, 1, (2019).
- TURNER, Sam y Robert YOUNG, “Concealed Communities: The People at the Margins”, *International Journal of Historical Archaeology*, 11, (2007), pp. 297-303.
- VERSCHOOF-VAN DER VAART, Wouter B., *Learning to look at LiDAR: Combining CNN-based object detection and GIS for archaeological prospection in remotely-sensed data*. [Tesis Doctoral]. Leiden, Universidad de Leiden, 2022.

- VERSCHOOF-VAN DER VAART, Wouter B. y Karsten LAMBERS, “Learning to Look at LiDAR: The Use of R-CNN in the Automated Detection of Archaeological Objects in LiDAR Data from the Netherlands”, *Journal of Computer Applications in Archaeology*, 2, (2019), pp. 31-40.
- WATRALL, Ethan, “Archaeology, the Digital Humanities, and the Big Tent”, en *Debates in the Digital Humanities*, Matthew K. GOLD y Lauren F. KLEIN (eds.), Minneapolis, University of Minnesota Press, 2016, pp. 345-358.
- WYLIE, Alison, “Mapping ignorance in archaeology. The Advances of Historical Hindsight”, en *Agnotology. The Making and Unmaking of Ignorance*, Robert N. PROCTOR y Londa SCHIEBINGER (eds.), Stanford, Stanford University Press, 2008, pp. 183-198.

VISUALIZACIÓN Y RECONOCIMIENTO DE LA MEMORIA HISTÓRICA DE LAS CAPULLANAS/CACICAS INDÍGENAS DESDE EL CAMPO DE LAS HUMANIDADES DIGITALES

Alicia Alvarado Escudero

Resumen

Este trabajo recupera la memoria histórica de 72 Capullanas o Cacicas indígenas que gobernaron en la costa norte del Perú y Ecuador entre los siglos XVI y XVII. Para ello, se ha partido de la investigación previa llevada a cabo por la autora¹. Estas mujeres desempeñaron roles de liderazgo durante la Conquista, resistiendo las imposiciones coloniales a través de herramientas legales y estrategias de negociación. A pesar del modelo patriarcal traído por los españoles, las Capullanas lograron mantener sus derechos sobre los cacicazgos heredados, demostrando autonomía y poder político en sus territorios.

Además, esta investigación contribuye con las nuevas líneas de investigación que ofrecen las Humanidades digitales. Utilizando recursos propios de las mismas, como el software *Omeka* y *Google My Maps*, se creó una base de datos y mapas interactivos que permiten localizar y analizar la información sobre estas gobernantes.

1 Se puede consultar la Tesis Doctoral en el siguiente enlace: <https://rio.upo.es/xmlui/handle/10433/6804>

Alicia Alvarado Escudero, Institución académica: UNIR
Alicia.alvarado@unir.net
ORCID: <https://orcid.org/0000-0003-0302-5847>

El presente trabajo revela que el 70.8% de las Cacicas identificadas gobernaron en su propio nombre, destacando el rol de las mujeres en contextos prehispánicos y coloniales. La mayoría de ellas se registraron en el siglo XVII y se concentraron en regiones clave como la costa norte peruana y la sierra ecuatoriana.

La metodología aplicada integra enfoques cuantitativos y cualitativos, combinando investigación histórica con tecnologías digitales para analizar y presentar los datos, facilitando el acceso a los registros históricos y evidenciando la importancia de estas mujeres y su legado, al tiempo que destaca la necesidad de seguir investigando su papel en otras regiones del Perú y América Latina.

Finalmente, este trabajo contribuye a la recuperación de una historia femenina invisibilizada, abriendo caminos para nuevas investigaciones sobre la gobernanza indígena femenina en Latinoamérica.

Palabras clave: Cacica, Capullana, Humanidades Digitales, Omeka, Google My Maps.

Abstract

This work recovers the historical memory of 72 Capullanas or indigenous Cacicas who ruled in the northern coastal regions of Peru and Ecuador between the 16th and 17th centuries. It builds upon prior research conducted on the subject. These women held leadership roles during the Conquest, resisting colonial impositions through legal tools and negotiation strategies. Despite the patriarchal model imposed by the Spanish, the Capullanas managed to maintain their rights over inherited cacicazgos, demonstrating autonomy and political power in their territories.

Furthermore, this research contributes to new lines of inquiry offered by digital humanities. By employing tools such as Omeka software and Google My Maps, an interactive database and maps were created to locate and analyze information about these female leaders.

The study reveals that 70.8% of the identified Cacicas ruled in their own right, highlighting the role of women in pre-Hispanic and colonial contexts. Most of these women were documented during the 17th century and concentrated in key regions such as the northern coast of Peru and the Ecuadorian highlands.

The applied methodology integrates quantitative and qualitative approaches, combining historical research with digital technologies to

analyze and present data. This facilitates access to historical records and underscores the significance of these women and their legacy, while emphasizing the need for further investigation into their roles in other regions of Peru and Latin America.

Finally, this work contributes to the recovery of a silenced female history, paving the way for new research on indigenous female governance across Latin America.

Keywords: Digital Humanities, Cacica, Capullana, Omeka, Google My Maps.

INTRODUCCIÓN

Las Humanidades Digitales (en adelante HD) han abierto el campo transdisciplinario de las investigaciones en Humanidades generando el desarrollo de nuevos enfoques y aportes culturales que permiten enriquecer el conocimiento humano. Este trabajo, se nutre de esta idea, y presenta una investigación interdisciplinar, siguiendo dos líneas, la *Creación digital* y la *Visualización de datos en Humanidades*. Para ello, se han tomado como referencia, los datos presentados por la autora en su Tesis Doctoral leída en 2019, *El impacto de la Conquista en las estructuras de poder femenino en la costa norte del Perú durante los siglos XVI y XVII*, donde se recogen 72 nombres de Capullanas/Cacicas de la costa norte del Perú y de Ecuador. Las Capullanas eran mujeres indígenas que habitaron en la costa norte del Perú y sur de Ecuador, antes y durante la llegada de los españoles en la época del contacto. Quedaron descritas en los Crónicas de Indias como mujeres que gobernaban sus territorios sin respaldo varonil. Lo que fascinó a los castellanos al contrastar tanto con la visión femenina exportada de la Península y con un claro cariz patriarcal heredado de la cultura romana (Alvarado, 2016) Con el tiempo, este término fue desapareciendo, siendo reemplazado por el de Cacica.

El objetivo clave de este trabajo de investigación es identificar y dar a conocer la historia de las mujeres indígenas de la costa del Perú y Ecuador, que durante los siglos XVI, XVII y XVIII, gobernaron sus territorios bajo el título y rol de «Cacicas» o «Ca-

pullanas». Con ello se pretende difundir su conocimiento al ámbito académico y al público interesado, facilitando la consecutiva ampliación de esta línea de investigación, así como generar el interés común de todo el público. Para ello, se ha creado una base de datos digital mediante el diseño de una página web con la herramienta *Omeka*, un software de código abierto cuyo fin es el desarrollo de proyectos de investigación, publicación y patrimonio cultural dentro de las HD. Posteriormente, se realiza un *Web Mapping* –conocido en español como mapa en la web o mapeo web– que permite la creación, publicación e interacción de mapas dinámicos en internet con la geolocalización de las Capullanas/Cacicas recogidas en la base de datos de *Omeka*, empleando el software de *Google My Maps*, como herramienta de visualización de datos propia de las HD, con la finalidad de analizar la información recogida desde otra perspectiva de análisis de localización espacial de las identidades cacicales femeninas, así como su análisis de concentración espacial.

En cuanto al aporte más significativo de esta investigación, los datos recogidos sobre las Capullanas/Cacicas, –o lo que es lo mismo, de las mujeres indígenas gobernadoras de ciertos territorios andinos–, consideramos que es muy importante la recuperación de la memoria histórica de estas mujeres pues tuvieron un papel fundamental en los primeros momentos de la Conquista, y posteriormente ante las nuevas reglas y normas políticas, sociales y económicas que se establecieron en la costa norte tras la imposición de la cultura castellana (sobre este tema ver los trabajos de Alvarado Escudero, 2019a; Coronel Feijoó, 2015; Daza Tobasura, 2016 y 2019; Díez Hurtado, 1988 y Graubart, 2000).

Estas mujeres tomaron un rol importantísimo al emplear una serie de herramientas de resistencia y negociación para no perder sus derechos históricos y ancestrales sobre sus Cacicazgos o terrenos de dominación ante las imposiciones de los españoles (Quispe-Agnoli, 2017), que renegaban de la mujer dejándola en un segundo plano ante el modelo de feminidad exportado de la Península Ibérica, donde no se concebía que una mujer tuviese

la «capacidad» necesaria para poder gobernar, siendo esta función fundamentalmente responsabilidad de los varones (Alvarado Escudero, 2016 y 2019; Morant Deusa y Bolufer Peruga, 1998 y Fiorentini Cañedo, 2011).

Las Cacicas, al ser conscientes de que no podían luchar en contra del nuevo modelo de imposición castellano que se estaba implantando en sus territorios, decidieron utilizar las mismas herramientas que las propias leyes castellanas les permitían, como, por ejemplo, reconocer y solicitar sus derechos de herencia y mayorazgo sobre los propios territorios heredados de sus madres (Puertas, 2007), empleando como instrumento legal las probanzas de méritos y servicios. En estos documentos justificaban mediante testigos y la presentación del árbol genealógico, su línea de consanguinidad con los gobernadores/as ascendentes de los Cacicazgos (Jurado, 2014). Aprendieron a leer y a escribir en castellano y a pleitear para solicitar sus títulos de «Cacicas» o «Capullanas»; y es gracias al empeño de mujeres por lo que hoy podemos consultar en los archivos históricos sus hazañas, sus demandas, sus identidades y darlas a conocer al mundo, para que no caigan en el olvido.

JUSTIFICACIÓN DEL TEMA

Es por tanto necesario el desarrollo de trabajos como este, mediante el marco metodológico de las HD, para recuperar la historia de estas mujeres. El uso de este lenguaje computacional permite recoger datos, analizarlos, visualizarlos y difundirlos con la finalidad de contribuir al conocimiento de una historia que durante siglos ha olvidado al género femenino.

Esta investigación es completamente necesaria para poder conocer principalmente a estas Cacicas de la costa norte del Perú y del Ecuador, aunque también se incluyen algunas Cacicas del centro y sur de Perú detectadas en la investigación previa: de los 72 datos recogidos sobre Cacicas, 45 de ellos corresponden a Cacicas que se ubican en la costa norte (62.5%), 18 Cacicas pertenecen a la región ecuatoriana (25%) y los últimos 9 datos (12.5%), hacen alusión a Cacicas que se localizan en el centro y sur de Perú,

en las regiones de Chíncha, Callao, Nazca, Cajamarca, Lima, Puno y el caso especial de Jauja, donde se localizan 3 Cacicas que ejercieron sus derechos sobre el Cacicazgo, perteneciendo a la misma familia, por lo que aquí sí podemos observar una continuidad en la tradición femenina cacical gracias a los trabajos de investigación realizados por Carlos Hurtado Ames (2017).

Además, el análisis de los datos recogidos y analizados sobre las Cacicas y su modo de gobernar –en su propio nombre, gobierna su marido o se desconoce el dato– también indican si las Cacicas ejercieron su poder de gobernación en mayor grado o, si, por el contrario, solo recibieron el título por ser esposas de Caciques. A continuación, se presentan dichos datos:

Tabla 1. Tipo de gobernación de las Capullanas/Cacicas

TÍTULO	COSTA NORTE DEL PERÚ			ECUADOR			CENTRO Y SUR DE PERÚ		
	Gobierna ella	Gobierna su marido	No hay datos	Gobierna ella	Gobierna su marido	No hay datos	Gobierna ella	Gobierna su marido	No hay datos
Señora, Capullana y Cacica	6	0	0	0	0	0	0	0	0
Señora o Cacica	2	0	4	6	0	2	1	0	0
Cacica Principal	23	1	3	7	0	2	1	2	1
Gobernadora	0	0	1	1	0	0	3	0	1
Cacica Segunda	1	2	2	0	0	0	0	0	0
TOTAL	32	3	10	14	0	4	5	2	2

Fuente: Elaboración propia (2023)

Estos datos se introducen en la herramienta *Omeka* como metadatos asociados a cada Capullana/Cacica, lo que permitirá acceder y conocer fácilmente esta información tan relevante sobre la gobernación autónoma de estas mujeres. Puesto que como se puede observar mayoritariamente –51 mujeres– gobernaron en su propio nombre, el 70.8% del total de Capullanas/Cacicas recopiladas en esta investigación. Este dato permite entender mejor que gobernaron sus territorios.

Por otro lado, esta investigación es de utilidad para que los usuarios/as e investigadores/as interesadas en las Cacicas, puedan de manera rápida y sencilla identificarlas y geolocalizarlas mediante el mapeo web elaborado en *Google My Maps* que junto con la información de la fuente que se comparte en dicho mapa, facilita la ubicación del archivo donde puede consultar la información.

METODOLOGÍA APLICADA A LAS HD

La Era digital y la irrupción del pensamiento computacional dentro de las Humanidades no solamente ha traído nuevos perfiles profesionales a este ámbito de estudio, sino que además ha generado que se formulen nuevas preguntas de investigación que ya no se pueden responder únicamente con metodologías basadas en investigaciones cualitativas. La nueva propuesta basada en el *Distant Reading* de Moretti (2013), donde los datos masivos generan nuevas preguntas y formas de dar respuesta a éstas, representa claramente una nueva ruta metodológica mixta. La metodología cualitativa o interpretativa y la metodología cuantitativa se unen para desarrollar la metodología mixta aplicable a las HD y principalmente a esta investigación.

METODOLOGÍA MIXTA Y TRANSDISCIPLINARIEDAD

En la primera fase de investigación, se revisaron las fuentes secundarias de trabajos publicados sobre el objeto de estudio. Esta fase se completó con una segunda, que consistió en la visita a los archivos históricos para realizar en primera persona la lectura de los documentos reales o fuentes primarias y comprobar de esta manera la fidelidad de los datos extraídos de las fuentes secundarias.

Por otro lado, se procedió a una explicación de la realidad femenina y su situación en la época prehispánica y medieval, así como durante la Conquista, mostrando el rol y funcionalidad mujer desde su visión más general hacia los casos particulares de estas Capullanas/Cacicas.

En esta última fase se dio un paso más y se llevó a cabo un trabajo transdisciplinar basado en una metodología mixta donde, por un lado, se agruparon, analizaron y visualizaron los datos referentes a 72 Capullanas/Cacicas recogidos en la base de datos en *Omeka*, junto el desarrollo de un mapa web, aplicando SIG, para la geolocalización de cada una de estas mujeres, así como de los yacimientos arqueológicos donde han aparecido enterramientos femeninos de élite en la época prehispánica. Este mapa será realizado con *Google My Maps*, y se correspondería con la ruta cuantitativa.

No obstante, la ruta cualitativa estará presente a la hora de analizar quiénes son cada una de estas mujeres, contextualizarlas en una época muy concreta bajo el rol político y social de Cacicas y Capullanas, y darles visibilidad para que puedan ser estudiadas en futuras investigaciones. Por lo tanto, se presenta una investigación transdisciplinar porque tanto el pensamiento computacional como el análisis histórico de las Cacicas es imprescindible para poder llevar a cabo este trabajo, no se complementan ambas disciplinas, sino que diluyen sus fronteras para generar nuevas investigaciones en el ámbito de las HD.

MARCO TEÓRICO

BREVE APROXIMACIÓN A LA HISTORIA Y DEFINICIÓN DE LAS HUMANIDADES DIGITALES

Encontrar una definición tipo para las HD en la actualidad no es una labor sencilla, pues son muchos/as los/as investigadores/as especializados/as en esta transdisciplina que todavía no aportan una definición cerrada y única, debido a la dificultad para poder caracterizarla dentro del encorsetado mundo de las disciplinas académicas (Piscitelli, 2013 y Alvarado 2012). No obstante, podemos tomar como referencia la definición de Magdalena Ziegler Delgado (2020) para aproximarnos a una idea de su constitución como transdisciplina innovadora:

Una suerte de red virtual que sintetiza las Humanidades tradicionales con la tecnología digital, logrando tejidos que ante-

riormente habrían sido imposibles. En ellas se aplica a los tradicionales y bien conocidos problemas de las Humanidades las nuevas herramientas tecnológicas. Contrario a lo que pudiera pensarse, no pretenden sustituir el cúmulo de conocimiento que hemos acumulado a lo largo de muchos siglos y que, ciertamente, es y seguirá siendo fundamental para todos. Las Humanidades Digitales tienen su propia identidad, aunque aún bajo construcción (Ziegler Delgado, 2020, 31).

No cabe duda de que las HD han sido beneficiarias del desarrollo de las nuevas tecnologías y de la incorporación del pensamiento computacional, propio de las Ciencias experimentales, dentro del ámbito de las Humanidades (Spence, 2014b; Spence, 2015; Rojas, 2015 y Rodríguez-Ortega, 2018), generando nuevas investigaciones en base a nuevas preguntas que se han formulado tras el manejo de datos masivos o datos a gran escala (Peña Pimentel y Sancho Caparrini, 2018). Es en la unión de estas disciplinas donde las HD adquieren una de sus características más identificativa, la transdisciplinariedad, superando las fronteras que encontramos entre las disciplinas únicas, y disipando áreas cerradas de trabajos y metodologías únicas para abrir un nuevo campo a la investigación, junto con un nuevo ámbito de trabajo y de formación para los humanistas, «generando un nuevo horizonte desde el que aproximarnos al análisis y la interpretación de los hechos y procesos culturales» (Rodríguez-Ortega, 2018: 3).

La transformación del humanista tradicional a humanista digital se produce tras la adquisición nuevas competencias y capacidades que le permitan la formulación de estas nuevas preguntas de investigación basadas en el uso de la tecnología, puesto que en las HD no podemos concebir un trabajo disciplinar, sino que se aboga por grupos interdisciplinarios donde diferentes áreas de investigación confluyen para dar respuesta a una pregunta en común (Galina, 2018).

A continuación, destacamos los elementos esenciales y más significativos de la HD.

METADATOS Y ESTÁNDARES Y FILOSOFÍA OPEN SOURCE EN HUMANIDADES DIGITALES

La introducción de las herramientas digitales dentro del ámbito de la investigación ha generado multitud de beneficios a la hora de poder colaborar y compartir en red proyectos transdisciplinarios publicados en abierto, convirtiéndose todo esto en la esencia de las HD. No obstante, esto no sería posible si en el uso de estas herramientas tecnológicas no se aplicaran ciertos usos y elementos sobre los datos recogidos en las investigaciones que permitieran su interoperabilidad –mediante el empleo de metadatos y siguiendo los correspondientes estándares establecidos– y la filosofía *Open Source* o código abierto que interviene para dotar a los proyectos de usabilidad y accesibilidad para todos los usuarios/as interesados/as.

El metadato es un concepto clave, puesto que es la base sobre la que se asienta la interoperabilidad, es decir, la cualidad de los recursos digitales para poder ser utilizados por diferentes sistemas de gestión de datos. Los metadatos son datos que se recogen sobre otros datos, por tanto, es toda la información adicional que podemos encontrar sobre una estructura o un contenido, de tal forma que estos metadatos nos permiten identificar, administrar y ubicar ese contenido que queremos localizar en un momento determinado.

Este trabajo ha generado *metadatos estructurados* puesto que son los que permiten recuperar, identificar y compartir toda la información recogida sobre las Capullanas/Cacicas. Además, se han empleado metadatos descriptivos encargados de identificar el recurso, así como facilitar su localización.

Por último, indicar, que en las HD se emplean estándares que permiten encorsetar y generalizar la manera de vincular los metadatos a un contenido específico. El estándar que principalmente se usa es el conocido como *Dublin Core*² para codificar los metadatos, define 15 términos con los que se puede describir un recurso digital, entre estos 15 términos podemos encontrar: título, tipo de recurso, derechos, fecha, formato, idioma, etc.

2 <https://www.dublincore.org/>.

En cuanto a la filosofía *Open Source*, traducido al español como *Código abierto*, tiene como principio la preferencia por el uso de software libre de código abierto que permita compartir datos basándose en la ampliación de conocimientos y de su colaboración entre distintos grupos de investigación de carácter interdisciplinar (González-Blanco, 2013). La finalidad de este movimiento es que cada vez más datos y más metadatos como, por ejemplo, los metadatos bibliográficos que podemos consultar en los repositorios digitales de manera fácil, accesible, y gratuita, sean una constante en el ámbito educativo. De esta manera, todos estos datos y metadatos pueden ser reutilizados por terceros, generando nuevos enfoques, reformulando nuevas interpretaciones de los paradigmas establecidos, y a su vez, nuevas investigaciones que amplíen los ámbitos del conocimiento y permitan la proliferación de investigaciones sobre temas que se encuadren en una sola disciplina.

LA RECUPERACIÓN DE LA MEMORIA HISTÓRICA DE LAS MUJERES EN LA ÉPOCA COLONIAL ANDROCENTRISMO. UNA HISTORIA CONTADA POR HOMBRES SOBRE HOMBRES

Cuando se comienza a estudiar la historia de las mujeres como objeto de estudio, sorprende la relativa actualidad de las investigaciones³. Esto mismo ocurrió cuando se inició la revisión de los trabajos para elaborar la Tesis Doctoral y se pudo comprobar como apenas existían trabajos sobre Capullanas/Cacicas exceptuando un libro que publicó en los años 60 María Rostworowski, *Curacas y sucesiones. Costa norte*; la tesis doctoral de Karen Graubart leída en 2000 y titulada *Con nuestro trabajo y sudor: indigenous women and the construction of Colonial society in 16TH and 17TH Century Perú*, en la Universidad de Massachusetts; algunos artículos de importantísima referencia como el de Chantal Caillavet (2008): “*Como caçica y Señora desta tierra...*” *Insignias, funciones y poderes de las sobera-*

3 Destacan las investigaciones de Anderson (2009), Duby (1994) y Perrot (2008).

nas del norte andino (siglos XV–XVI); el artículo de Rosario Coronel Feijóo (2015): *Cacicas indígenas en la Audiencia de Quito, siglo XVIII: las redes ocultas del poder*; y por último, el trabajo de Paula Daza Tobasura (2016), *Gobernar en tiempos de cambio. Las cacicas de la Audiencia de Quito*.

Esto llevó a investigar por qué estas mujeres tan importantes en la historia de la Conquista no estaban siendo estudiadas, ni en los libros de historia, ni en los programas universitarios, ni existían investigaciones dedicadas a su reconocimiento. Fue entonces cuando se empezó a entender que la historia se ha escrito desde una perspectiva androcéntrica, entendiendo por Androcentrismo «la visión del mundo y de las relaciones sociales centrada en el punto de vista masculino» (DLE, consultado 16 de mayo de 2022)⁴, es decir, que la historia y la ciencia se han escrito bajo un prisma y una visión masculina, construyendo un pensamiento sesgado y generando toda una tradición intelectual que invisibiliza a las mujeres (Alvarado Escudero, 2016 y Perrot, 1994). Para tratar el origen de esta desigualdad de género tendríamos que remontarnos a la época romana y a su posterior contexto histórico medieval donde se establecieron las pautas para entender cuál era el concepto de feminidad que toda mujer debía de representar (Alvarado Escudero, 2015; Morant, 2005 y Morant, 1998). Esta diferenciación entre géneros estaba justificada desde tres discursos (Duby y Perrot, 1994): el *discurso teológico*, basado en el derecho natural cristiano que recoge la castidad, pureza y virginidad como ideales de la «buena mujer». El *discurso científico* que define a la mujer como una versión «incompleta» del cuerpo masculino (Berriot-Salvadore, 1994), y por último el *discurso jurídico* asentado sobre el derecho romano castellano que situaba al hombre como el cabeza de familia y a la mujer dependiente e incapacitada para decidir sobre su propia vida (Alvarado Escudero, 2015). Teniendo como referente estas ideas, no es de extrañar que los primeros cronistas quedaran asombrados al arribar a las costas peruanas y ecuatoria-

4 <https://dle.rae.es/androcentrismo>.

nas, y observar cómo estas mujeres indígenas no encajaban en el prototipo de feminidad que ellos entendían como el correcto. Es gracias a este choque cultural por lo que hoy se pueden revisar las *Crónicas de Indias* y conocer desde la perspectiva patriarcal de estos hombres, cómo interactuaban estas mujeres indígenas de élite entre su población (ver Alvarado Escudero, 2019 y Daza Tobasura, 2019). Y por extensión, desarrollar este tipo de investigaciones tan sumamente necesarias para dar a conocer la proeza de estas mujeres, que supieron adaptarse a las normas impuestas por los castellanos para seguir manteniendo sus costumbres y sus derechos a gobernar. Fue tal su esfuerzo por ser parte de la historia y dejar constancia de ello, que en los fondos documentales se puede ver que aprendieron a pleitear bajo las normas castellanas, e incluso a leer y escribir en la lengua española, para poder reclamar sus derechos, así como los territorios que les pertenecían por herencia y tradición (Coronel Feijóo, 2015). Y gracias a este esfuerzo, hoy es posible ubicarlas en la posición histórica que ocupan y merecen, reconociendo además su valía y su identidad para que nunca más queden olvidadas en los archivos.

LAS CAPULLANAS/CACICAS EN LA HISTORIA DE LA CONQUISTA DE AMÉRICA

La historia universal que se ha aprendido durante siglos se ha visto sesgada por no incorporar en los escritos a las mujeres como protagonistas fundamentales de ésta, o por incluirlas únicamente cuando han destacado por su belleza o heroísmo, estableciendo una sociedad jerarquizada articulada entre el espacio público reservado a los hombres, y el espacio privado ocupado por mujeres (Guardia, 2018; Andreo y Guardia, 2013 y Morant Deusa, 2005).

Las fuentes primarias y secundaria toman cómo sujetos históricos y principales protagonistas de la historia a los varones, por consiguiente, las mujeres han sido invisibilizadas y silenciadas durante siglos. Hasta que en los años 70 el movimiento femenino por la emancipación y la libertad de las mujeres promovió el desarrollo de la «Historia de las mujeres», y la categoría de género

como objeto de estudio, de la mano de grandes historiadoras como son Asunción Lavrin (1985) o Michelle Perrot (2008). Comenzando así, una nueva etapa importantísima para la recuperación de la memoria histórica de éstas. Es dentro de esta historia de género actual, dónde encuadramos esta investigación en HD con el objetivo de conseguir «la reconstrucción del pasado femenino» (Guardia, 2018, 362).

En el caso de las Capullanas/Cacicas, se ha podido reconstruir su historia gracias a la documentación etnohistórica que se ha recogido tanto de las crónicas de indias como de los archivos históricos y en trabajos de fuentes secundarias. Estas mujeres aparecen citadas por 18 cronistas de los cuales 16 de ellos estuvieron presentes en el Perú. En todos los casos, estas pequeñas descripciones de las mujeres indígenas de élite corresponden al asombro de los cronistas ante su actitud, por no cumplir con el canon establecido de feminidad en la mentalidad patriarcal de la España moderna, heredada de la cultura romana (Alvarado Escudero, 2015 y 2019). Podemos afirmar que estas mujeres tomaron el rol de gobernadoras de sus territorios gracias a un derecho heredado de sus madres y abuelas que las colocaba como jefas de los territorios –Cacicazgos– que quedaban bajo su dominio y control tanto político, como social, como religioso puesto que en el área andina y mucho más en la costa norte, la sacralidad del cargo del Cacicazgo era lo que legitimaba en el poder a los Caciques y las Cacicas (Ramírez, 2002) hasta que las estructuras de poder castellanas rompieron con esta legitimidad prehispánica y sagrada e impusieron un nuevo orden donde los varones gobernaban en primer término pese a que el cargo lo recibían sus mujeres por herencia legítima (Alvarado Escudero, 2019^a: 133-140).

Pero no solo en las crónicas aparecen estas Capullanas/Cacicas. Otra fuente excelente y mucho más amplia para localizar, identificar y conocer a estas mujeres indígenas poderosas y elocuentes, son los documentos recogidos en los archivos históricos, pues los pleitos en los que participaban recogen sus datos, así como las probanzas de méritos y servicios que permiten identificarlas,

ubicar su campo de dominio, su época y su árbol genealógico, junto con las razones que las llevaron a aprender a leer y escribir en castellano para poder defender sus derechos frente a los tribunales castellanos (Alvarado Escudero, 2019a).

El análisis de estos datos muestra como la mayoría de la documentación revisada corresponde con los documentos de archivo donde estas Cacicas aparecen bajo la titularidad de «Cacica principal» –40 mujeres en total, el 58.3% de las Cacicas identificadas–. Siendo muy notable también que la mayoría de ellas, gobernaron en el siglo XVII –22 de ellas, el 30.5% de las Cacicas identificadas–. Este siglo es una época clave, puesto que se están forjando los cambios en la sociedad andina, por un lado, por la implantación de las costumbres castellanas en Latinoamérica y por el otro, por el proceso de aculturación que la población andina ya estaba viendo, hechos que se identifican en el aprendizaje de la lengua castellana y de los procedimientos legales para reivindicar sus derechos al Cacicazgo ante los tribunales (Alvarado Escudero, 2019a y Coronel Feijóo, 2015.)

Se concluye este epígrafe no sin antes mencionar que las Cacicas no solo existieron en Perú y Ecuador, sino que cada vez son más los trabajos como el de Margarita Ochoa y Sara Vicuña (2021) que demuestran que las mujeres gobernadoras en la época pre y postcolonial, estuvieron presentes a lo largo de toda Latinoamérica.

HERRAMIENTAS DIGITALES PARA LA REALIZACIÓN DE ESTE PROYECTO OMEKA

Para poder trabajar con *Omeka Classic*, se ha contado con la empresa de comunicación Umami⁵, que ha facilitado gratuitamente tanto el dominio, como *hosting* de la página web. Como ventajas se destacan:

5 <https://www.esumami.com/>.

- Mayor capacidad de almacenamiento, en principio se han ofrecido 10 GB gratuitos por la empresa Umami. La capacidad es lo más importante porque se trabajará con la subida de imágenes más adelante.
- Mayores recursos para futuras ampliaciones y modificaciones de los contenidos, así como la incorporación de plugin para realizar trabajos colaborativos.
- Soporte técnico ante cualquier problema.
- Perdurabilidad y accesibilidad inmediata.
- Emplea las 15 definiciones semánticas descriptivas de *Dublin Core* en sus metadatos.

GOOGLE MY MAPS

Se puede definir como SIG online entendiendo esta como: «la representación visual de información geográfica con el objetivo de facilitar la comunicación de conocimiento espacial» (Kraak, 2004). Además, la creación de estos mapas web interactivos, suponen el cambio de la cartografía estática por una geográfica dinámica, situada en la nube y de temática múltiple. Por esta razón, en esta investigación se ha utilizado este software gratuito para desarrollar el mapa web donde se puede llevar a cabo la visualización de la información de las Cacicas, mediante el proceso de análisis de la información y sus tres etapas (Garrido-Ramos, 2016, 16): *Pre-proceso* –recogida y preparación de datos–, *Proceso y análisis de la información* –extraer la información de los datos recolectados tras la creación del mapa web– y *Presentación* de los resultados y conclusiones tras la visualización de los datos. Lo interesante de esta herramienta no solamente son las funcionalidades que presenta de manera gratuita, sino que además se puede modificar siempre que se desee de manera fácil rápida e intuitiva, y, además, permite compartirlo fácilmente mediante un enlace⁶. Además, su carácter

6 En el siguiente enlace se puede consultar el mapa interactivo creado por la autora: <https://www.google.com/maps/d/edit?mid=1Or-PmGdPhe2heWox7f1-qCVrIPITxzGzz&usp=sharing>.

colaborativo lo convierte en una herramienta excelente para proyectar este trabajo en un futuro, donde el resto de los investigadores/as puedan participar añadiendo al mapa interactivo nuevas localizaciones de Capullanas/Cacicas que hayan encontrado en los archivos o ampliando la información de las que se encuentran ubicadas en el mapa.

Este tipo de proyectos corresponden a la inclusión de la geolocalización mediante SIG en los proyectos de HD para una mayor comprensión de los datos que se genera desde la perspectiva geoespacial ampliando a su vez un mayor número de conclusiones.

MARCO EMPÍRICO

CREACIÓN DE LA BASE DE DATOS CON OMEKA Y GOOGLE MY MAPS

Se recuperaron los datos de los 72 registros de Cacicas, organizándolos de la siguiente manera:

Título, nombre, fecha, tipo de gobernación (ella por derecho o acompañando a su esposo), territorio sobre el que gobernó, y la fuente de la que hemos sacado la información.

Estos datos específicos de cada mujer se seleccionaron en el desarrollo de la Tesis Doctoral con el objetivo de dar respuesta a la pregunta de investigación que se planteó en ella: conocer quiénes eran estas mujeres (nombre/identidad), cuándo y dónde tuvieron sus cacicazgos (fecha y ubicación). Y, por último, conocer si se había cerrado el paso a la gobernación femenina indígena a medida que los españoles imponían sus leyes y costumbres patriarcales, o si, por el contrario, emplearon estrategias de resistencia y aprendieron las costumbres castellanas, para aprovecharse de los vacíos legales existentes con el fin de mantenerse en el poder por derecho propio. La respuesta se encuentra en los datos recogidos sobre el tipo de gobernación, y en su propio nombre.

CREACIÓN DE LA PÁGINA WEB CON OMEKA

Posteriormente se definió qué tipo de base de datos se quería generar, así como cuáles fueron los metadatos que se emplearon en cada registro. Se optó por una base de datos relacional puesto que la finalidad es que desde la barra de búsqueda simple o empleando los criterios de búsqueda avanzada que aparece en la parte superior de la web, se pueda buscar por palabra clave y acceder a todos los registros de nuestra base de datos que contengan esa palabra que se está buscando.

Se elaboraron 5 colecciones (cada una correspondiente a una titularidad cacical) y, a su vez, una serie de elementos o fichas (72 en total con la información recogida de cada una de las Cacicas). Estos elementos se asociaron a cada colección, según el título que ostentó cada Cacica.

Dentro de cada colección se emplearon una serie de metadatos que identifican cada registro, esta parte fue sencilla, puesto que ya indicamos que *Omeka* se basa en los 15 campos principales de *Dublin Core*, y después, dentro de cada colección, se crearon los elementos asociados a cada una de ellas

Desde la *Home* podemos ver el acceso en pestañas a nuestro contenido, cada pestaña es una Simple Page, siendo la primera de ellas, la que corresponde a la Home, aparece como *Inicio*, donde se ha introducido la contextualización de este proyecto, así como el enlace a la Tesis, una imagen de la firma de una de las Cacicas que es representativa de todo este trabajo, y la lista de las cinco colecciones:

En la segunda pestaña se ha colocado el siguiente título: *¿Quiénes eran?* Donde se explica la historia de estas mujeres y se añaden varias imágenes sobre mujeres de élite indígenas. Otra pestaña presenta las *Colecciones*.

CREACIÓN DEL MAPA DE LAS CACICAS CON GOOGLE MY MAPS

El primer paso fue acceder mediante una cuenta Gmail a *Google My Maps*⁷ y escoger un título: *Mapa de Capullanas/Cacicas coloniales* y Descripción del recurso. Posteriormente, se siguieron los criterios empleados en la página web: es decir, el contenido se ha dividido en los bloques correspondientes a la tipología de las titularidades cacicales presentadas en las colecciones. Pero también se ha considerado conveniente añadir un apartado más en el mapa, donde se ubican los yacimientos arqueológicos femeninos característicos por contener enterramientos femeninos de poder que pueden tener alguna asociación con las Capullanas o Cacicas. Por tanto, en el mapa se presentan las seis capas correspondientes a las diferentes titularidades y en cada una de ellas, se ha introducido la ficha y geolocalización de cada Cacica.

Lo más relevante de *Google My maps* es que permite filtrar la información por capas, de tal forma, que, si a los/as usuarios/as únicamente les interesase conocer la ubicación de las Capullanas, por ejemplo, sólo tendrían que clicar en esta opción y desactivar las demás capas para centrarse únicamente en estos datos. De esta forma, se puede visualizar un mapa general con toda la información o, por el contrario, centrarse solamente en el objeto de estudio.

CONCLUSIONES

Se concluye esta investigación alcanzando la principal finalidad de esta: identificar visualizar y dar a conocer la historia de las mujeres indígenas de la costa del Perú, que durante los siglos XVI, XVII y XVIII, gobernaron sus territorios bajo el título y rol de «Cacicas» o «Capullanas». Desde el marco teórico metodológico de las HD se ha llevado a cabo la recolección, análisis y visualización de datos, mediante dos recursos de búsqueda y recopilación de la informa-

7

Se puede navegar por el mapa en el siguiente enlace: <https://www.google.com/maps/d/viewer?mid=1OrPmGdPhe2heWox7fI-qCVrI-PITxzGzz&usp=sharing>

ción –Omeka y Google My Maps– referente a estas mujeres y recogido en la página web www.CapullanasyCacicas.com, creada en Omeka, y almacenada en el servidor de la empresa Umami.

La aplicación de estas herramientas ha permitido un enriquecimiento de la investigación presentada anteriormente, puesto que el almacenamiento, análisis y recuperación de los datos cualitativos y cuantitativos permiten confirmar que, por un lado, el título de «Cacica principal», fue el más empleado por las Cacicas estudiadas en el siglo XVII: el 58.3% de las Cacicas identificadas poseyeron esta titularidad, y el 30.5% lo detentaron en el siglo XVII. Por otro lado, en el caso de la gobernación femenina cacical, en la mayoría de los casos, el 70.5%, estas mujeres gobernaron por propio derecho, sin un respaldo varonil. Por tanto, en la mayoría de los datos recogidos sobre la gobernación cacical femenina de trabajo (51 de 72), las Capullanas/Cacicas, gobernaron de forma personal e independiente al varón.

Por otro lado, la visualización de los datos sobre el mapa web empleando los SIG que incorpora *Google My Maps*, permite realizar un análisis de la localización especial de las Cacicas, así como las zonas geográficas de máxima concentración de las mismas, destacando muy por encima del resto de regiones, la costa norte del Perú y sierra de Ecuador, donde no solo se registran el mayor número de Cacicas identificadas en este trabajo, sino que además la mayoría se ubican cercanas a las zonas arqueológicas donde se han localizado mujeres de élite indígenas. Línea de investigación que ya introducía en la mencionada Tesis.

Por último, la revisión exhaustiva de las fuentes primarias y secundarias y la extracción de los datos referentes a las Cacicas que se han compartido en *Omeka* y *Google My Maps*, también pone en evidencia la falta de investigaciones realizadas en otras regiones peruanas, donde se ha podido localizar algunas mujeres de élite indígenas, pero apenas se han podido revisar trabajos sobre ellas, por lo que esta investigación también aporta ideas nuevas sobre futuras líneas de investigación que se pueden desarrollar en base al Cacicazgo femenino en el centro y sur del Perú.

REFERENCIAS BIBLIOGRÁFICAS

- ALVARADO ESCUDERO, Alicia, “El impacto de la Conquista en las estructuras de poder femenino en la costa norte del Perú durante los siglos XVI y XVII”, Tesis Doctoral, Universidad Pablo de Olavide, Sevilla, 2019a, disponible en: <https://rio.upo.es/xmlui/handle/10433/6804> [consulta: 17/12/2024].
- ALVARADO ESCUDERO, Alicia, “El adorno labial, labret, bezote, tembetá o barbote, como símbolo de empoderamiento femenino en la costa norte del Perú”, en *Género y mujeres en la historia del Perú. Del hogar al espacio público*, Claudia Rosas Lauro (ed.), Fondo Editorial Pontificia Universidad Católica de Perú, 2019b, pp.57-80.
- ALVARADO ESCUDERO, Alicia, “La Imagen de la Mujer de Élite en la Costa Norte del Perú a través de las Crónicas de Indias”, *Lex: Revista De La Facultad De Derecho y Ciencia Política De La Universidad Alas Peruanas* 14 (18), (2016), pp. 380-398.
- ALVARADO ESCUDERO, Alicia, “El Concepto de Feminidad en la Época del Descubrimiento de América”. Actas VII Congreso virtual sobre Historia de las mujeres, 15 al 31 de octubre de 2015, Asociación de amigos del Archivo H. Diocesano, 2015.
- ALVARADO, Rafael, “The digital Humanities Situation”, en *Debates in the Digital Humanities*, Matthew k. Gold (ed.), Minneapolis, University of Minnesota Press, 2012, pp. 50-55.
- ANDERSON, Bonnie y Judith ZINSSER, *Historia de las Mujeres: Una historia propia*. Madrid, Editorial Critica, 2009.
- ANDREO, Juan y Sara Beatriz GUARDIA, *Historia de las mujeres en América latina*. Murcia, CEMHAL, 2013.
- BERRIOT-SALVADORE, Evelynne, “El discurso de la medicina y de la ciencia”, en *Historia de las Mujeres, Del Renacimiento a la Edad Moderna*, George DUBY y Michelle PERROT (eds.), Madrid, Taurus, 1994, pp. 371-413.
- CAILLAVET, Chantal, “Como caçica y Señora desta tierra...” Insignias, funciones y poderes de las soberanas del norte andino (siglos XV-XVI)”, *BIFE*, 37 (2008), pp. 57-80.
- CORONEL FEIJÓO, Rosario, “Cacicas indígenas en la Audiencia de Quito, siglo XVIII: las redes ocultas del poder”, *Procesos: revista ecuatoriana de historia*, 42 (2015), pp. 9-37.

- DAZA TOBASURA, Paula, “Gobernar en tiempos de cambio. Las cacicas de la Audiencia de Quito”, *Fronteras de la Historia*, 21(2) (2016), pp. 79-103.
- DAZA TOBASURA, Paula, “Gobernar en tiempos de cambio: las cacicas de la Audiencia de Quito en el siglo XVIII”, Tesis Doctoral, FALCSO Ecuador, 2019, disponible en: <https://repositorio.flacsoandes.edu.ec/handle/10469/15241> consulta: 17/12/2024].
- DÍEZ HURTADO, Alejandro, *Pueblos y Caciques de Piura: siglos XVI y XVII*, Piura, Centro de Investigación y Promoción de Campesino, 1988.
- DUBY, Georges y Michelle PERROT, *Historia de las Mujeres, Tomo II: La Edad Media y Tomo III: Del Renacimiento a la Edad Moderna*, Madrid, Taurus, 1994.
- FAULHABER, Charles y Francisco MARCOS-MARÍN, “ADMYTE: Archivo Digital de Manuscritos y Textos Españoles”, *La Corónica*, 18(2) (1990), pp. 131-145.
- FIorentini CAÑEDO, Natalia, “Coviértete en lo que eres: construcción de la masculinidad y la feminidad en el discurso del derecho natural cristiano en la Nueva España del siglo XVI”, *Dimensión Antropológica*, 52(18) (2011), pp. 31-56.
- GALINA, Isabel, “La institucionalización de las Humanidades”, en *Humanidades Digitales: recepción, institucionalización y crítica*, Isabel Galina (ed.), Ciudad de México, Bonilla Artigas Editores, 2011, pp. 17-37.
- GARRIDO RAMOS, Beatriz, “Visualización de datos y clasificación Icon-class: un estudio de caso desde la perspectiva de las Humanidades Digitales”, *Caracteres: Estudios culturales y críticos de la esfera digital*, 6(1) (2016), pp. 10-33.
- GONZÁLEZ-BLANCO GARCÍA, Elena, “Un nuevo camino hacia las Humanidades Digitales: El laboratorio de innovación en Humanidades Digitales de la UNED (LINHD)”, *Revista Singa*, 25 (2017), pp. 79-93.
- GRAUBART, Karen, “Con nuestro trabajo y sudor: indigenous women and the construction of Colonial society in 16TH and 17TH Century Peru”, Tesis Doctoral, University of Massachusetts, 2000.
- GUARDIA, Sara Beatriz, “Historia de las mujeres: un derecho Conquista”, en *Género y mujeres en la historia del Perú. Del hogar al espacio pú-*

- blico, Claudia ROSAS LAURO (ed.), Lima, Fondo Editorial Pontificia Universidad Católica del Perú, 2018, pp. 353-367.
- HERMIDA JIMÉNEZ, Nuria, “Imago Orbis: Infraestructura de Humanidades Espaciales en el CSIC”, XII Jornadas Ibéricas de Infraestructuras de Datos Espaciales (JIIDE 2021), disponible en: <https://www.idee.es/resources/presentaciones/JIIDE21/Resumenes/01.pdf> [consulta: 17/12/2024].
- HURTADO AMES, Carlos, “Caciccas, dinastías indias y caciques en la sierra central del Perú: Élite indígena y configuración del poder en el espacio regional de Jauja”, Tesis Doctoral, México, El Colegio de México, 2017.
- JURADO, Carolina, “Descendientes de los primeros. Las probanzas de méritos y servicios y la genealogía Cacical. Audiencia de Charcas, 1574-1719”, *Revista de Indias*, LXXIV (261) (2014), pp. 387-422.
- KRAAK, Menno-Jan, “The role of the map in a Web-GIS environment”, *Journal of Geographical Systems*, 6 (2004), pp. 83-93.
- LAVRIN, Asunción, *Las mujeres latinoamericanas. Perspectivas históricas*, México, Fondo de Cultura Económica, 1985.
- LÓPEZ POZA, Sagrario, “Humanistas y Humanidades digitales. Trayectoria y proyección en la filología española”, en *Humanidades y Humanismo, Homenaje a María Pilar Cuartero*, Aurora Egido, José E. LAPLANA y Luis SÁNCHEZ LAILLA (eds.), 2020, pp. 125-159.
- LUCÍA MEGÍAS, José Manuel, La informática humanística: notas volanderas desde el ámbito hispánico, *Incipit*, XXIII, 2003, pp. 91-114.
- MORANT DEUSA, Isabel, *Historia de las Mujeres en España y América Latina. De la prehistoria a la Edad Media y el mundo moderno*. Madrid, Cátedra, 2005.
- MORANT DEUSA, Isabel y Mónica BOLUFER PERUGA, *Amor, matrimonio y familia: La construcción histórica de la familia moderna*. Madrid, Editorial Síntesis, 1998.
- MORETTI, Franco, *Distant Reading*, Londres, Verso, 2013.
- OCHOA, Margarita y Sara VICUÑA GUENGERICH, *Caciccas, The indigenous Women Leaders os Spanish America 1492-1825*. Norman, University of Oklahoma Press, 2021.
- O’PHELAN GODOY, Scarlett, *Kurakas sin sucesiones. Del Cacique al alcalde de indios Perú y Bolivia 1750-1835*, Cuzco, Centro de Estudios Regionales Andinos Bartolomé de las Casas, 1997.

- PERROT, Michelle, *Mi historia de las mujeres*, Buenos Aires, Fondo de Cultura Económica de España, 2008.
- PEÑA PIMENTEL, Miriam y Fernando SANCHO CAPARRINI, “Big Data y Humanidades Digitales”, en *Humanidades Digitales: lengua, texto, patrimonio, datos*, Isabel GALINA RUSSELL, Miriam PEÑA PIMENTEL y Ernesto PRIANI SAISÓ, (coord.), Ciudad de México, Bonilla Artigas Editores, 2018, pp. 16-39.
- PISCITELLI, Alejandro, “¿Cómo Definir a Las Humanidades Digitales? ¿O no Definirlas?”, *Educación y Ciudad* 25, (2013), pp. 25-38.
- PUERTAS, Elisabeth, “La mujer frente al poder en la sociedad colonial peruana (siglos XVI-XVIII)”, en *La mujer en la historia del Perú (siglos XV al XX)*, Carmen MEZA y Teodoro HAMPE (comps.), Lima, Fondo Editorial del Congreso del Perú, 2007, pp. 147-188.
- RAMÍREZ, Susan, *El mundo al revés. Contactos y conflictos transculturales en el Perú del siglo XVI*, Lima, Fondo Editorial del Congreso del Perú, 2002.
- QUISPE-AGNOLI, Rocío, *Women’s Negotiations and Textual Agency in Latin America, 1500-1799*, Londres y Nueva York, Routledge, 2017.
- RODRÍGUEZ ORTEGA, Nuria, “Cinco ejes para pensar las Humanidades Digitales como proyecto de un nuevo humanismo digital”, *Artnodes. Revista de arte, ciencia y tecnología* (22), (2018), pp. 1-6.
- ROJAS CASTRO, Antonio, “Las Humanidades Digitales: Principios, Valores y Prácticas”, *JANUS: Estudios Sobre El Siglo De Oro*, 2 (2013a).
- ROJAS CASTRO, Antonio, “El Mapa y El Territorio: Una Aproximación Histórico-Bibliográfica a La Emergencia De Las Humanidades Digitales En España”, *Caracteres (Salamanca)*, 2(2) (2013b), pp. 10-53.
- ROJAS CASTRO, Antonio, “De los datos al metamedio”, *Ínsula. Revista de letras y ciencias humanas* (822), (2015), pp. 15-17.
- ROSTWOROWSKI DE DÍEZ CANSECO, María, *Curacas y sucesiones: costa norte*, Lima, Instituto de Estudios Peruanos, 1961.
- SPENCE, Paul, “Centros y fronteras: el panorama internacional de las Humanidades Digitales”, en *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, Sagrario LÓPEZ y Nieves PENA (eds.), Janus, Anexo 1 (2014a), pp. 37-61.

- SPENCE, Paul, “La investigación humanística en la era digital: mundo académico y nuevos públicos”, en *Humanidades Digitales: una aproximación interdisciplinar*, Álvaro BARAIBAR (coord.), Janus, Anexo 2 (2014b), pp. 117-131.
- SPENCE, Paul, “Creación y difusión del conocimiento científico en la era digital”, *Ínsula. Revista de letras y ciencias humanas* (822), (2015), pp. 8-12.
- ZIEGLER DELGADO, María Magdalena, “El Tiempo De Las Humanidades Digitales”, *Seeci 2000*, 52 (2020), pp. 29-47.



9 788497 498951



UNIVERSIDADE DA CORUÑA



Humanidades
Digitales
Hispánicas