

## SOBRE LAS HUMANIDADES DIGITALES EN LA FILOLOGÍA HISPÁNICA

*Nerea Parro Gómez*  
Université de Neuchâtel

Que las humanidades digitales irrumpieron con fuerza hace años en el ámbito de la filología es un hecho indiscutible. Es numerosa la bibliografía que ahonda en la historia del uso de lo digital en el campo de las humanidades desde sus orígenes —que la mayoría de los autores establece en el proyecto<sup>1</sup> del jesuita Roberto Busa para la indexación de la obra completa de Tomás de Aquino con el apoyo económico y la tecnología de IBM, que dio lugar a la publicación del *Index Thomisticus* (Busa 1980; 2022 [1962])<sup>2</sup>—. Sin embargo, no es el propósito de estas páginas profundizar en la historia y evolución de las humanidades digitales ni detallar sus hitos (y mitos), sino ofrecer un panorama amplio de su campo de acción en lo que respecta a la labor del filólogo. Ahora bien, los expertos tardaron un tiempo en decidir el término que daría nombre a esta disciplina. Así se explica en la introducción al volumen titulado *Defining Digital Humanities*:

Over the past years, the field that we now refer to as digital humanities has been known by many terms: humanities computing, humanist informatics, literary and linguistic computing and digital resources in the humanities, to name but a few. Most recently it has predominantly been known as digital humanities, though other variations such as eHumanities are occasionally to be found in literature emanating from continental Europe (Terras, Nyhan y Vanhoutte 2013: 2).

---

<sup>1</sup> Aunque la mayoría de los autores afirman que este es el primer proyecto en humanidades digitales, muchos apuntan como precursora a Ada Lovelace en el siglo XIX (Hernández-Lorenzo 2022: 153).

<sup>2</sup> Un proyecto llevado a cabo por Alarcón (2002) y Alarcón y Bernot (2005) ha sido el convertir este trabajo del padre Busa en un corpus al uso en la actualidad. Puede consultarse en: <<https://www.corpusthomicum.org/it/index.age>>.

Si bien es cierto que en determinadas ocasiones los términos para el mundo anglófono no tienen necesariamente el mismo significado cuando se usan para el ámbito hispánico —Del Rio Riande (2014: 33; 2015: 15-17) insiste en la dificultad que supone transponer esta disciplina heredada de los *English Departments* (cf. Kirschenbaum 2010) a las humanidades digitales hispánicas y Priani Saisó *et al.* (2014: 11) observan una diferencia de uso entre el término en español y en inglés—, en este caso, también se ha reflexionado sobre cómo denominar esta disciplina para el español. En esta misma línea, Galina Russell (2011: 3) ha observado los siguientes términos “Recursos Digitales para las Humanidades, Cómputo para las Humanidades, Cómputo en las Humanidades, Informática Digital y Cultural e Informática para las Humanidades”, aunque a lo largo de la bibliografía publicada puede advertirse que se utilizan también otros para referirse a este tema como ‘recursos digitales’, ‘nuevas tecnologías’, ‘filología digital’, ‘informática para las humanidades’, ‘informática humanística’, ‘metodología informática’, ‘*humanities computing*’ y demás términos que algunos autores prefieren mantener en inglés, hasta decidirse por el sintagma ‘humanidades digitales’ o ‘*Digital Humanities*’<sup>3</sup>.

Ahora bien, lo complicado en esta disciplina —o línea de trabajo, etiqueta, metodología o campo (Rockwell 2013 [1999]: 13-33; Rodríguez-Yunta 2013; Del Rio Riande 2015: 17), porque parece que también hay dificultad para determinar su naturaleza ontológica— no fue cómo denominarla, ya que observábamos más arriba cómo la multiplicidad de términos desembocó en ‘humanidades digitales’ o ‘*Digital Humanities*’. Lo realmente complicado para los expertos parece ser el hecho de proponer una definición que concrete qué son. En general, las propuestas de los autores se caracterizan por la cautela a la hora de dar una definición de lo que son las humanidades digitales —e incluso de lo que no son—, aunque, a veces, siguen siendo poco concretas y demasiado vagas, quizá debido a que se sitúan en medio de muchas disciplinas.

Un ejemplo de la disparidad de definiciones podría ilustrarse mediante la existencia de una web (imagen 1) que recoge algo más de ochocientas definiciones de lo que diferentes expertos entienden por humanidades digitales o *Digital Humanities*. Estas citas, unas más extensas y elaboradas y otras más

---

<sup>3</sup> Kirschenbaum (2010: 56-57) propone el origen de esta denominación en una discusión entre los editores sobre el título que le darían al volumen que acabó llamándose *A Companion to Digital Humanities*, publicado en 2004.

breves y concretas, fueron recogidas entre 2009 y 2014 con motivo del Día de las Humanidades Digitales.

## What Is Digital Humanities?

New quote

DH is the perfect cocktail of computer science and rationalization of culture and humanities. **Elena González-Blanco**

---

NB: Click the [New quote](#) button to display a new random quote. Quotes were pulled from participants from the [Day of DH](#) between 2009-2014. As of January 2015, the database contains 817 rows and randomly selects a quote each time the page is loaded. If you want to do something cool with the data, I am providing the code and data [here](#).

Made by [Jason Heppler](#). Problems? Questions? I'm [@jaheppler](#) on Twitter.

Imagen 1<sup>4</sup>

*Entrada de la página web “What is Digital Humanities?” (Heppler 2015)*

Centrado en el ámbito hispánico, Del Rio Riande proporciona toda una serie de definiciones aportadas por diferentes hispanistas de ambos lados del Atlántico (2014: 21-25), todas en la línea de la enunciada por Rojas Castro (2013a: 79), que fue una de las primeras propuestas acerca de lo que se entiende por humanidades digitales para este ámbito, quien las describe como un “conjunto de principios, valores y prácticas en donde convergen múltiples objetos de estudio y saberes cuyas fronteras se encuentran en continua negociación”. Con todo, tal vez lo relevante no sea insistir en establecer una definición exacta, sino enfocarse en identificar cuáles son las características comunes que se repiten a lo largo de las diferentes descripciones. Una propuesta parecida a esta idea es la de Hernández-Lorenzo (2022: 142-152), dado que destaca algunas de las características que también están presentes en la mayoría de definiciones, como: la interdisciplinaridad; el trabajo colaborativo entre especialistas de distintos ámbitos de estudio; la internacionalización; la difusión en acceso abierto de códigos y métodos; la institucionalización gracias a diferentes organismos, centros académicos o laboratorios; o el trabajo

---

<sup>4</sup> Se ha escogido una definición al azar de las ochocientas diecisiete disponibles en mayo de 2025, cuya autora es Elena González-Blanco.

con grandes cantidades de datos también llamados *big data*. De lo anterior se desprende, sin querer aportar una nueva definición, sino simplemente recoger los puntos en común orientados al campo de la filología —ya que ese es el propósito de este volumen—, que se trata del punto de encuentro del saber filológico con la metodología y tecnología digitales que daría como resultado nuevas herramientas para la edición y estudio de fuentes para la investigación en lengua y literatura y otros recursos útiles para el análisis y presentación de datos que, además, pueden compartirse con toda la comunidad investigadora y pueden difundirse en acceso abierto.

Aunque cada vez menos, esta nueva proyección de los estudios filológicos puede suponer un cambio de mentalidad que parecería chocar con el concepto tradicional del filólogo que se encierra solitario en una biblioteca, rodeado de libros y fichas e investigando de manera aislada en sus manuscritos (González-Blanco 2013). Ese filólogo mostraría ahora otra faceta como filólogo digital, que ha adquirido formación tecnológica e informática, cuyo trabajo es individual y a la vez colaborativo, ya que tras la edición de textos o creación de corpus textuales —partiendo de la base de que el texto es una de las fuentes principales de estudio del filólogo— suele haber todo un equipo trabajando en conjunto, y que, además, está conectado con otros investigadores para poder compartir los métodos de trabajo y resultados. Eso sí, la única manera de poder llevar a cabo este tipo de proyectos sería contando con una sólida formación filológica: en palabras de Fradejas (2022: 40) “*primum philologia, deinde computatrum scientia*, no se puede ser editor digital sin una sólida base de filología y crítica textual”. En efecto, la figura del filólogo tradicional se revela como un componente esencial en la conformación del filólogo digital. En este sentido, solo aquel que es filólogo de manera “analógica”, también puede serlo en el mundo digital, toda vez que las decisiones sobre el objeto de estudio deben ser tomadas desde una formación consolidada que garantice el rigor científico en el paso desde el papel hasta el formato digital.

Si bien este escenario representa el ideal del quehacer filológico unido a las posibilidades que ofrece el mundo digital, uno de los desafíos actuales es precisamente la formación tecnológica básica de los investigadores. En la mayoría de los casos, el autoaprendizaje ha supuesto la única manera para la adquisición de competencias digitales tales como conocimientos en programación, herramientas de análisis, procesamiento y presentación o visualización de datos, etiquetado de textos, servidores web, creación y gestión de bases de datos, etc. Precisamente porque esta formación tecnológica se ve algo

limitada, es habitual formar equipo junto a técnicos o ingenieros informáticos. Esta simbiosis de ciencias y letras es, en muchas ocasiones, la única combinación que hace posible la realización del proyecto pensado primeramente por el filólogo y que se plasma en el mundo digital gracias al ingeniero. A pesar de esto, la formación tecnológica de los filólogos es cada vez más completa y especializada y la formación en humanidades digitales se va implementando paulatinamente en los planes de estudios y cursos destinados a investigadores.

Retomando una de las ideas anteriores, el texto es una de las fuentes principales del filólogo como objeto de estudio, y por esta razón, no es de extrañar que también fuera el objeto de los primeros proyectos que se desarrollaron desde la implementación de lo digital en el campo de la filología hispánica, que comenzó su andadura a finales de los años 80 (Rojas Castro 2013b; Spence y González-Blanco 2014). Así, a partir de esta época comenzó a publicarse bibliografía sobre este tema<sup>5</sup> y, unos años más tarde, empezaron a digitalizarse archivos y bibliotecas al mismo tiempo que se creaban repositorios digitales de textos que ya nacían digitales<sup>6</sup>. Estos tipos de edición digital podrían corresponderse con las dos primeras tipologías de edición digital descritas por Pierazzo (2019a: 179) —también por Torruella y Llisterri (1999: 51), aunque enfocadas únicamente hacia la lingüística de corpus—: el *archivo editorializado* o “archivo/colección informatizado” (podría ser, según lo descrito por los autores, una colección de imágenes digitales, provista de metadatos y que puede contar con una catalogación del contenido) y la *edizione di lettura* o “Biblioteca de Textos Electrónicos” para Pierazzo (2019a: 179) tendría el énfasis en la simplicidad de acceso al texto y la fiabilidad científica ya que, para este tipo de ediciones normalmente se dispone de las imágenes facsimilares o de una transcripción con criterios explícitos y notas al texto<sup>7</sup>, mientras que para Torruella y Llisterri (1999: 51) esta colección de textos no seguiría un riguroso criterio de selección. Sin embargo, el objeto de estudio iría hoy más allá de la digitalización para la difusión y acceso abierto de ciertas obras, lo que nos

---

<sup>5</sup> Por ejemplo, uno de los primeros libros publicados fue el de Marcos Marín en 1994 y un artículo de Lucía Megías en 1998.

<sup>6</sup> Podemos destacar la digitalización de la Biblioteca Nacional de España o de archivos históricos nacionales y provinciales y otros proyectos nacidos digitales como el Archivo Digital de Manuscritos y Textos Españoles (ADMYTE), la Biblioteca Virtual Miguel de Cervantes, el Corpus de Referencia del Español Actual (CREA) y el Corpus Diacrónico del Español (CORDE), entre muchos otros.

<sup>7</sup> Adaptación del original en italiano (Pierazzo 2019a: 179-180).

llevaría al tercer tipo que la autora distingue: la *edizione arricchita*. Esta “edición enriquecida” sería la manera en la que se hace hincapié en la calidad del paratexto, la posibilidad de realizar diferentes presentaciones del mismo texto (paleográfica, crítica, diplomática, facsimilar, etc.) y puede ir acompañado de anotaciones filológicas, históricas, de un aparato crítico, etc. (Pierazzo 2019a: 180). Resulta interesante cómo la propuesta de esta autora intenta establecer tres tipologías generales que serían clasificables sea cual sea el acercamiento metodológico o teórico del investigador al preparar la edición<sup>8</sup>. En este sentido, parece esencial establecer unos criterios claros que permitan separar de manera explícita la edición del propio texto de los añadidos del editor. Es precisamente la edición enriquecida —o corpus enriquecido—, que es la que se aplica generalmente a los textos literarios y documentos históricos, lo que le interesa generalmente al filólogo, dado que esta implica la adición de metadatos, marcado estructural y anotaciones lingüísticas que incrementan su valor para la investigación (búsquedas, análisis, presentación de datos, etc.) y la conservación digital.

Si se tiene en consideración el marco anterior, un corpus lingüístico o literario podría explicarse —de manera sumamente simplificada ya que para cada una de las ramas de esta disciplina existirían diferencias ligadas a los intereses del estudio— como una colección de textos debidamente recopilados en función de unos criterios marcados por el objeto de estudio. Sería pertinente realizar una aclaración, y es que un corpus no necesariamente debe cumplir el requisito de estar alojado en un soporte digital, pues siguen existiendo y teniendo gran utilidad los corpus realizados de manera tradicional:

[...] el término corpus se relaciona casi indefectiblemente con lo digital [...]. Elaboración de corpus e informática han confluído, pero no siempre se han identificado, pues hay corpus elaborados para edición en papel. Es evidente, sin embargo, que ahora no pueden concebirse por separado (Sánchez-Prieto Borja 2012: 29).

Sin embargo, sigue existiendo un problema y es que, aunque se marquen los criterios que ha seguido el proceso de edición de un texto, ¿cómo puede saberse sin ninguna duda si un corpus es fiable? Señalaba Le Roy Ladurie (1982 [1975]: 10) que “toute étude historique doit ou devrait commencer par

---

<sup>8</sup> Sobre este tema resulta esclarecedor el manual de Pierazzo (2015) acerca de los diferentes acercamientos teóricos y metodológicos y modelos de edición digital de textos.

une critique aux sources”<sup>9</sup>, pero quizá hoy sería conveniente añadir la coletilla ‘y a cómo estas han sido editadas’. Aquí entraría en juego la importancia de la estandarización que requiere el empleo de unos criterios digitales coherentes de edición que sean claros para que el investigador conozca cómo ha sido el proceso de edición digital y sea capaz de recuperar la información que le interesa. Este fue precisamente el objetivo principal del consorcio Text Encoding Initiative<sup>10</sup> (TEI), cuya misión es desarrollar y preservar directrices para la codificación de textos literarios y lingüísticos y el intercambio de textos electrónicos. Esta estandarización es posible gracias a, por un lado, el uso del lenguaje XML (eXtensible Markup Language), que permite la preservación, estructuración y almacenaje de los datos y metadatos; y, por otro, al etiquetado TEI, que se ha convertido en el estándar para el etiquetado de textos porque posibilita recuperar la información textual y no textual, es decir, cualquier otra información que pueda ser interesante para el investigador en el documento. Esto es, permite realizar una anotación lingüística del texto (por ejemplo, una anotación morfosintáctica, semántica, lematización, etc.) y no textual (añadir, por ejemplo, la data crónica y tópica, el género textual, si hay cambio de manos, el autor, etc.). No quisiéramos detenernos más en las posibilidades del etiquetado, ya que se explica con mayor profundidad y a partir de ejemplos de proyectos filológicos en los siguientes capítulos. En resumen, el lenguaje de marcado XML-TEI permite representar datos textuales con precisión estructural y etiquetarlos con procesamiento informático. Ciertamente, una parte de ese etiquetado se realiza de manera automática, aunque requiere una revisión manual. Esto, que parece algo intrascendente, sería oportuno tenerlo en cuenta puesto que, en función del fenómeno, sería necesario revisar los datos recuperados de cada uno de los corpus por si la máquina no hubiese sabido interpretar algún aspecto del etiquetado y el humano lo hubiera pasado, sin querer, por alto. En cualquier caso, este lenguaje de marcado ofrece una total transparencia, pues generalmente el corpus se presenta junto con la hoja de ruta que se ha seguido en el proceso de edición. Igualmente, esta transparencia y difusión permiten la interoperabilidad, un punto realmente interesante en tanto en cuanto el hecho de compartir la metodología seguida en la edición permitiría su replicabilidad por otros investigadores:

---

<sup>9</sup> “Todo estudio histórico debe o debería comenzar por un estudio crítico de las fuentes” (traducción mía).

<sup>10</sup> TEI Consortium disponible en <<https://tei-c.org/>>.

El papel de la codificación TEI como formato de archivo e intercambio es una ventaja para la preservación de ediciones digitales a largo plazo y para su documentación: el lenguaje TEI fue diseñado para modelar y codificar tanto el texto, por ejemplo, de una fuente existente, como la intervención académica de los editores durante la transcripción y edición. Esto convierte a un documento TEI en un registro importante del trabajo editorial en sí mismo. En otras palabras, una ficha TEI colabora sustancialmente con la documentación de un proyecto de edición digital (Del Rio Riande 2024: 52).

Si bien este planteamiento se muestra como la gran solución a la estandarización de los diferentes proyectos en humanidades digitales, la realidad muestra que la financiación constituye un factor importante en la estabilidad y continuidad de los proyectos para la creación de corpus digitales. Esta dependencia de los recursos económicos para la disponibilidad del soporte técnico y humano paradójicamente se revela como una contradicción frente a los principios de preservación y acceso abierto. En este sentido, no existe una única solución que garantice la conservación y continuidad de un proyecto en el tiempo y todo depende del tipo de proyecto (cf. Pierazzo 2019b sobre las ediciones tipo *haute couture* y *prêt-à-porter*) y su dependencia de los fondos institucionales. En caso de pérdida de financiación que no permitiese la continuidad del proyecto, hoy es posible el almacenamiento de los flujos de trabajo y la preservación de los datos gracias a sitios web estáticos y repositorios especializados que no tienen apenas mantenimiento denominados *minimal computing* (Del Rio Riande 2024: 50-54) que van ampliando sus capacidades.

En las páginas anteriores se ha querido mostrar un panorama general de las humanidades digitales en el ámbito hispánico y poner de relieve que las tecnologías digitales configuran activamente las formas de generar y difundir el conocimiento. No obstante, aún queda mucho camino que recorrer desde la necesidad de enfoques interdisciplinarios más sólidos hasta la continuación, pervivencia y reconocimiento académico de los corpus o ediciones digitales. Pese a todo lo anterior, se puede ser optimista en cuanto al futuro de las humanidades digitales en la filología hispánica puesto que, en muy poco tiempo, se han desarrollado numerosos proyectos de los que se han ido empapando y aprendiendo los más recientes, lo que da cuenta de la progresiva reducción de la brecha entre lo tecnológico y lo filológico. Quién sabe. Quizás algún día se resuelvan todas estas cuestiones, dejemos de hablar de humanidades digitales y volvamos a hablar solo de ‘humanidades’, que tendrá implícita la parte de lo digital.

La idea de realizar este volumen surge a partir de las reflexiones intercambiadas durante el XXI Congreso de la Asociación Internacional de Hispanistas celebrado en Neuchâtel a propósito de un simposio que versaba sobre el etiquetado TEI como herramienta metodológica en la elaboración de corpus y cómo estaba siendo su implementación y desarrollo en diferentes proyectos en el ámbito de las humanidades, especialmente en lengua y literatura hispánicas<sup>11</sup>. Sin embargo, es importante señalar que el presente volumen no se corresponde con las actas de dicho encuentro<sup>12</sup> ni del simposio mencionado. Si bien esta sección dentro del encuentro académico fue el punto de partida para muchos de los debates aquí desarrollados, las comunicaciones originalmente presentadas durante este simposio no entraron finalmente en el volumen. No obstante, gracias a ese intercambio se observaron una serie de preocupaciones comunes entre muchos investigadores a este respecto, así como un creciente interés en torno al uso de sistemas de etiquetado como herramienta para la creación de bases de datos empíricas para la investigación en el campo de la filología hispánica. Justamente, de esta inquietud manifiesta nace este volumen, *La investigación en filología hispánica: humanidades digitales, corpus y estándar TEI*, que reúne una serie de trabajos firmados por un grupo heterogéneo de expertos humanistas digitales que respondieron a la invitación de participar, con el objetivo de dar a conocer las diferentes facetas del uso del etiquetado TEI en la confección de corpus textuales y las posibilidades para la realización de estudios a partir de estos. En conjunto, esta selección de trabajos permite articular una perspectiva plural y actualizada sobre los desafíos y posibilidades de la codificación y etiquetado textual para la investigación en hispánicas. Por esta razón, se ha pretendido aportar un modesto acercamiento en las páginas anteriores sobre los tres puntos clave que vertebran el monográfico: las humanidades digitales, los corpus y el estándar TEI.

El libro se articula a partir de un enfoque filológico integrador, que actúa como hilo conductor de los contenidos. Con este enfoque se pretende dar una visión diferente puesto que la bibliografía en humanidades digitales da cuenta muchas veces de la fragmentación monotemática en las publicaciones: lite-

---

<sup>11</sup> Agradezco las intervenciones, presentaciones e intercambio de ideas a todos los investigadores que participaron en este simposio que organizamos junto a mis compañeras Elena Padrón Castilla y Maialen Blázquez González en la Université de Neuchâtel en julio de 2023.

<sup>12</sup> Las actas del encuentro pueden consultarse en (Sánchez Méndez, Padrón Castilla y Parro Gómez 2025).

ratura frente a lingüística y frente a historiografía<sup>13</sup>. Por este motivo, en los capítulos que componen este volumen se abordan las disciplinas fundamentales para la investigación en filología hispánica, entendida como una ciencia humanística interdisciplinar —el análisis literario, el estudio lingüístico y la contextualización histórica—, a la vez que imbrican de manera transversal la metodología propia de las humanidades digitales. Esta convergencia metodológica y de disciplinas abre el camino a una aproximación filológica renovada que enriquece los procedimientos de análisis y presentación de datos además de la edición y sistematización de los textos. Pese a no contar con separaciones que aúnen por temática los diferentes capítulos del volumen, la estructura que imbrica la organización en bloques de los capítulos es la siguiente:

## 1. Proyectos, ediciones y corpus etiquetados con TEI

Esta primera sección recoge diferentes proyectos que se están llevando a cabo en la actualidad o han concluido recientemente, en los que los autores explican las fases, metodología del proceso, desafíos y toma de decisiones en la creación de diferentes corpus literarios e históricos.

Así, en la primera parte, Elena Cano Turrión muestra una revisión crítica de las posibilidades metodológicas que ofrece la codificación XML-TEI en el ámbito de la investigación en literatura centrado especialmente en el estudio de los Siglos de Oro. La autora realiza una revisión historiográfica de diferentes proyectos que se han desarrollado en la creación digital de corpus literarios: el proyecto pionero SILEM, que actualmente va por su III edición, que sentó los cimientos para futuros proyectos y otros dos corpus posteriores con textos de época áurea, PRESOLO y PARANOBA. El artículo recorre las ventajas y limitaciones de cada uno de estos corpus etiquetados con TEI, además de ofrecer las diferentes decisiones metodológicas que se tomaron para

---

<sup>13</sup> No obstante, algunas de las últimas publicaciones en este campo están enfocadas desde una temática concreta como, por ejemplo, el papel de las humanidades digitales en la enseñanza de la literatura hispánica (Martínez Cantón, Ortuño Casanova y Huertas Morales 2023); humanidades digitales aplicadas al estudio de la historiografía lingüística hispánica (Battaner Moro y López Iniesta 2023); ventajas y desventajas de la realización de estudios a partir de corpus de distinto tamaño: desde el pequeño dominio hasta los *big data* (Calderón y Vaamonde 2024); humanidades digitales enfocadas en los textos del español medieval y moderno (González Zalacain y Vaamonde 2025), entre otras.

la edición, anotación de los textos, buscadores y programas empleados en la creación de cada uno de ellos, en función del objeto de estudio de cada uno de los proyectos. Sin embargo, todos estos proyectos abogan por una integración metodológica coherente del estándar TEI en la investigación filológica que les permite articular descripciones estructuradas, enriquecer los textos con metadatos, facilitar su análisis y mecanismos de búsqueda, además de, por supuesto, la preservación y difusión de los textos de cada corpus gracias al entorno digital.

El segundo capítulo, escrito por Soledad Castaño Santos, ofrece una reflexión crítica sobre los desafíos metodológicos encontrados durante el proceso de creación de un corpus conformado por Biblias romanceadas, desarrollado en el marco del proyecto “Mapping the Scriptures in Western Sephardi Literature”, que se está llevando a cabo en la Universidad de Basilea. El objetivo principal del capítulo es justificar y explicar las decisiones tomadas para solventar diversos problemas metodológicos derivados de la complejidad del objeto de estudio. Entre estas dificultades destacan la característica longitud y subdivisiones de los textos seleccionados, las decisiones de segmentación y transcripción y el anotado de distintas particularidades que se muestran en los documentos, como la intervención de la función litúrgica o fragmentos en otras lenguas, además de las ornamentaciones de los textos. Cada desafío ha implicado soluciones diversas por parte de los autores, quienes han abogado por los criterios de adaptabilidad, interoperabilidad y permeabilidad de manera que la metodología que han desarrollado pueda adaptarse a otros proyectos futuros y a la particularidad de los textos sefardíes romanceados. Además, a lo largo del capítulo, la autora da cuenta de la reflexión filológica y documental previa a todo el proceso de creación del corpus —anterior a la transcripción, segmentación y etiquetado con TEI— y que permite establecer criterios homogéneos para la selección y tratamiento de los textos.

Juan Esteban Hincapié Atehortúa presenta la edición crítico-genética digital de *Frutos de mi tierra* (1896) de Tomás Carrasquilla, un proyecto en curso en humanidades digitales pionero en la filología colombiana. El capítulo revisa los fundamentos teóricos de la crítica textual, reflexiona sobre los desafíos que supone la edición digital y plantea el paso del texto impreso a la plataforma digital a partir de las fases tradicionales de la ecdótica. El autor explica con detalle todos los pasos que sigue para la realización del proyecto: la elección y sistematización de los testimonios de esta novela; la identificación de variantes textuales y genéticas; la codificación y anotado del texto en

XML-TEI que complementa añadiendo etiquetas para la creación de un glosario de dialectalismos y otras notas explicativas. Esta metodología permite la representación dinámica en la visualización de la edición de las variantes textuales, manuscritas e impresas, las intervenciones del autor y los procesos de corrección. La adopción del estándar TEI no solo asegura la preservación y accesibilidad de los textos, sino también hace que el proyecto sienta una base de posibilidades metodológicas para la investigación y el desarrollo de ediciones críticas y genéticas digitales futuras.

## **2. Estudios lingüísticos a partir de corpus etiquetados con TEI**

Los siguientes capítulos intentan dar respuesta a la pregunta: ¿qué ventajas ofrecen los corpus etiquetados con TEI? o, más específicamente, ¿cómo emplearlos en una investigación lingüística de corte diacrónico? La pertinencia de los corpus etiquetados, lematizados y anotados morfosintácticamente para el estudio histórico de fenómenos en lengua española es muy ventajosa en tanto en cuanto permiten la búsqueda y recuperación automática de datos que, en caso de los corpus de cierto tamaño, posibilitan análisis cuantitativos a partir de los cuales poder observar la diacronía de un determinado fenómeno lingüístico.

Este quinto capítulo trata el recorrido histórico del diminutivo *-ico* en español peninsular entre los siglos XVI y XVIII (1500-1799). Para este estudio Miguel Calderón Campos atiende especialmente a dos factores de cambio lingüístico: el origen geográfico y el tipo textual. Para ello, confronta los datos de dos regiones, Madrid-Sevilla frente a Granada-Almería, y analiza dos tipos de género textual, inventarios de bienes y epistolarios privados. Además de esto, la clasificación de los diminutivos con base en las tres funciones básicas de este fenómeno (denotativa, afectiva y atenuadora) le permiten determinar las diferentes etapas evolutivas del diminutivo *-ico* en su cronología. Para la obtención de los datos, el autor acude a los corpus en línea *Post Scriptum*, *Oralia Diacrónica del Español (ODE)*, *Corpus Léxico de Inventarios (CorLexIn)* y *Corpus de Documentos Españoles Anteriores a 1900 (CODEA +2022)*. Tanto *Post Scriptum* como *ODE* son corpus etiquetados con TEI y están construidos en la plataforma TEITOK. Esta plataforma incluye un etiquetador morfosintáctico que permite realizar búsquedas en CQL (*Corpus Query Language*) y recuperar así todos aquellos sustantivos o adjetivos etiquetados como disminu-

tivos acabados en -ico y sus derivados. El estudio, que se apoya en un análisis cuantitativo de los datos, permite determinar que, a partir del siglo XVII, se reduce territorialmente el uso de -ico en la franja Madrid-Sevilla y se limita funcionalmente a las funciones afectivas o atenuadoras, más presentes en las cartas que en los inventarios.

Inmaculada González Sopeña y Gael Vaamonde buscan un doble propósito en su artículo sobre la anotación y el análisis diacrónico de arabismos en el español clásico y moderno. Por una parte, explicar la metodología que han implementado en el proceso de anotación y recuperación automática de arabismos en el corpus ODE; y, por otra, realizar, a partir de esta anotación, un análisis cuantitativo sobre la evolución y frecuencias de uso de este tipo de léxico que les permita proponer una periodización de este fenómeno. Para ello, los autores describen el procedimiento y estándares que han seguido para la marcación y codificación (XML-TEI) y la anotación morfosintáctica (EAGLES) de los arabismos en el corpus, además de su recuperación para el análisis mediante sintaxis CQL. Para la anotación de los arabismos, han acudido a un leuario extraído de una herramienta externa (IEDRA) y explican detalladamente el desarrollo del *script* que les ha permitido anotar como arabismos aquellos lemas del corpus que aparecían en esa lista. Este proceso les ha facilitado realizar un estudio cuantitativo en el que se analizan las frecuencias por millón del uso de arabismos, que presentan en gráficos estadísticos. Los datos muestran una periodización de los arabismos dividida en cuatro etapas para el periodo entre 1500 y 1900, con especial atención a la segunda mitad del siglo XVI y al siglo XVIII como periodos de resurgimiento de este tipo de léxico.

### 3. Paralelismos históricos

Es curioso cómo el intento de establecer unos mismos criterios para estandarizar grandes cantidades de documentación ha sucedido otras veces en el pasado. En este último capítulo Andrés Felipe Echavarría Peláez y Bárbara Santiago Medina realizan una analogía entre dos sistemas normativos de estandarización documental que, si bien distan cronológicamente en su creación, ambos responden al objetivo de regular y formalizar las prácticas de producción documental: el *Orden de processar* de Pablo García (1568) y las *Guidelines* del Consorcio TEI. En cuanto al primero, Pablo García, secretario de tribunal de distrito de la Inquisición, se basaría en sus visitas de inspec-

ción a los diferentes tribunales para elaborar un manual práctico con el fin de unificar la práctica procesal y documental de los procesos seguidos por la Inquisición en cada uno de sus tribunales. Esta propuesta de homogeneización acabaría siendo implantada de manera general en todos los tribunales de distrito por el Consejo de la Suprema. Con este mismo objetivo se crea el estándar internacional TEI, en aras de promover la estandarización de diferentes tipos documentales. De esta manera, puede observarse un paralelismo de cómo estas dos iniciativas terminan convirtiéndose en modelos de trabajo finalmente respaldados por instituciones oficiales en aras de la unificación de criterios normativos. Además, los autores ejemplifican el etiquetado de distintos elementos TEI al adaptarlo al esquema jurídico inquisitorial que siguen los expedientes por procesos de fe.

En conclusión, el conjunto de estudios que aquí se presentan ofrecen diversas posibilidades de creación, edición y estudio de los corpus con etiquetado TEI a partir de las experiencias propias de los investigadores. Por una parte, se reúnen diferentes proyectos que dan cuenta de la profunda reflexión teórico-metodológica que se esconde tras el empleo de este etiquetado a lo largo de cada una de las fases del proceso de edición y creación de corpus. Por otra parte, se visibilizan de manera especialmente significativa las ventajas que ofrece el etiquetado TEI para la investigación lingüística y, en particular, para el estudio de la lengua en perspectiva histórica. También se muestra cómo los corpus codificados con TEI, cuando son complementados con lematización y anotación morfosintáctica, se convierten en instrumentos de análisis diacrónico de alta precisión, capaces de ofrecer datos sistemáticos y recuperables a gran escala.

A pesar de la versatilidad del etiquetado TEI, sobresale en todos los artículos de este volumen la necesidad de compartir las decisiones metodológicas, los desafíos técnicos y las implicaciones epistemológicas que se han experimentado en cada uno de los proyectos en aras de fomentar la interdisciplinariedad y la colaboración de manera que futuros filólogos puedan servirse de la experiencia de investigaciones anteriores y así adaptarla a sus propios proyectos y objetivos.

En definitiva, este volumen pretende, modestamente, ilustrar el estado actual de las prácticas de etiquetado textual digital en filología hispánica. El uso del etiquetado TEI seguirá siendo, previsiblemente, un punto de encuentro entre la tradición filológica y la tecnología digital. Por último, este volumen sugiere algunas vías que podrán ser desarrolladas en un futuro como, por ejemplo, la investigación en relación a la representación de fenómenos; el desarrollo de es-

trategias para la anotación semántica enriquecida; las posibilidades que ofrece la inteligencia artificial para el etiquetado y transcripción de textos, especialmente manuscritos, y las herramientas de reconocimiento y procesamiento del lenguaje; o la evolución de las prestaciones de las plataformas en línea en las que se albergan los corpus. Además, se pone de relieve la necesidad de adaptación del estándar TEI a tipos documentales específicos y de extender su aplicación a géneros textuales no canónicos. Estas futuras líneas de investigación tienen en común un horizonte interdisciplinar en el que la estandarización textual mediante el etiquetado TEI se convierte en impulsor de integración metodológica.

No podría poner fin a estas páginas introductorias sin manifestar mi agradecimiento a todos los autores por su infinita paciencia, su interés y el rigor de sus trabajos. Además, quisiera agradecer también la labor de Anne Wigger y de los revisores externos que han evaluado con imparcialidad cada uno de los trabajos de este volumen. Asimismo, la publicación de este monográfico es posible gracias a una subvención del Fonds national suisse de la recherche scientifique, por lo que aprovecho la ocasión para reconocer a esta institución su apoyo económico.

Sin la ayuda de Juan Pedro Sánchez Méndez y sin la de Elena Padrón Castilla no habría existido este volumen. Tampoco habría llegado a buen término sin el respaldo de Antonio Corredor Aveledo, Cipriano López Lorenzo y Mariela de La Torre. A ellos les doy de corazón las gracias por su apoyo constante.

## Referencias bibliográficas

- ALARCÓN, Enrique (2002): “El proyecto *Corpus Thomisticum*: descripción y perspectivas”, *Anuario Filosófico*, 35(3), pp. 791-801, DOI: <<https://doi.org/10.15581/009.35.29445>>.
- ALARCÓN, Enrique y BERNOT, Eduardo (2005-): *Corpus Thomisticum*, edición digital de las obras de santo Tomás de Aquino. <<https://www.corpusthomisticum.org/it/index.age>>.
- BATTANER MORO, Elena y LÓPEZ INIESTA, Juan Alonso (eds.) (2023): *Humanidades digitales e historiografía lingüística hispánica: proyectos de presente y retos de futuro*. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.
- BUSA, Roberto (1980): “The Annals of Humanities Computing: The Index Thomisticus”, *Computers and the Humanities*, 14(2), pp. 83-90.
- (2022 [1962]): “El análisis lingüístico en la evolución mundial de los medios de comunicación” (introd. Stefano Bazzaco, trad. Soledad Castaño Santos), *Historias Fingidas*, número especial 1, pp. 19-38, DOI: <<https://doi.org/10.13136/2284-2667/1228>>.

- CALDERÓN CAMPOS, Miguel y VAAMONDE, Gael (eds.) (2024): *Linguistic Corpora and Big Data in Spanish and Portuguese*. Berlin/Boston: De Gruyter. DOI: <<https://doi.org/10.1515/9783110781465>>.
- DEL RIO RIANDE, María Gimena (2015): “Humanidades Digitales. Mito, actualidad y condiciones de posibilidad en España y América Latina”, *ArtyHum. Revista Digital de Artes y Humanidades*, 1, pp. 7-19.
- (2024): “Acerca de la sostenibilidad, archivo y preservación de proyectos de edición digital de textos hispánicos”, *Philologia Hispalensis*, 38(2), pp. 41-57, DOI: <<https://doi.org/10.12795/PH.2024.v38.i02.02>>.
- FRADEJAS RUEDA, José Manuel (2022): “De editor analógico a editor digital”, *Historias Fingidas*, Numero Speciale 1: *Digital Humanities e studi letterari ispanici*, pp. 39-65, DOI: <<https://doi.org/10.13136/2284-2667/1108>>.
- GALINA RUSSELL, Isabel (2011): “¿Qué son las Humanidades Digitales?”, *Revista Digital Universitaria*, 12(7), pp. 3-10.
- GONZÁLEZ-BLANCO GARCÍA, Elena (29/10/2013): *Humanidades Digitales y Filología. Nuevos retos para el hispanismo*, UNED TV (III Seminario del Máster en Formación e Investigación Literaria y Teatral en el Contexto Europeo). <<https://canal.uned.es/video/5a6f615fb111f54168b45b8>> [consultado 10/05/2025].
- GONZÁLEZ ZALACAIN, Roberto J. y VAAMONDE, Gael (eds.) (2025): *Digital Humanities in Medieval and Early Modern Spanish Texts: Current Perspectives and Approaches*. London/New York: Routledge. DOI: <<https://doi.org/10.4324/9781003393771>>.
- HEPPLER, Jason (s.f.): *What Is Digital Humanities?* <<https://whatisdigitalhumanities.com/>> [consultado 10/05/2025].
- HERNÁNDEZ-LORENZO, Laura (2022): “Introducción al panorama internacional de las Humanidades Digitales”, *Revista de Humanidades*, 46, pp. 137-166.
- KIRSCHENBAUM, Matthew G. (2010): “What Is Digital Humanities and What’s It Doing in English Departments?”, *ADE Bulletin*, 150, pp. 55-61.
- LE ROY LADURIE, Emmanuel (1982 [1975]): *Montaillou, village occitan de 1294 à 1324*. Paris: Gallimard.
- LUCÍA MEGÍAS, José Manuel (1998): “Editar en Internet (che quanto piaci il mondo è breve sogno)”, *Incipit*, 18, pp. 1-40.
- MARCOS MARÍN, Francisco (1994): *Informática y humanidades*. Madrid: Gredos.
- MARTÍNEZ CANTÓN, Clara I., ORTUÑO CASANOVA, Rocío y HUERTAS MORALES, Antonio (eds.) (2023): *Las humanidades digitales en la enseñanza de las literaturas hispánicas. Aplicaciones prácticas*. Berlin: Peter Lang. DOI: <<https://doi.org/10.3726/b21145>>.
- PIERAZZO, Elena (2015): *Digital Scholarly Editing: Theories and Practices*. London/New York: Routledge.
- (2019a): “Edizione documentaria digitale: rinuncia intellettuale o opportunità scientifica?”, *Ecdotica*, 16, pp. 174-185, DOI: <<https://doi.org/10.7385/99302>>.
- (2019b): “What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter”, *International Journal of Digital Humanities*, 1, pp. 209-220, DOI: <<https://doi.org/10.1007/s42803-019-00017-3>>.
- PRIANI SAISÓ, Ernesto; SPENCE, Paul; GALINA RUSSELL, Isabel; GONZÁLEZ-BLANCO GARCÍA, Elena; ALVES, Daniel; BARRÓN TOVAR, José Francisco; GODÍNEZ BUSTOS, Marco Antonio y PAIXÃO DE SOUSA, Maria Clara (2014): “Las humanidades

- digitales en español y portugués. Un estudio de caso: DíaHD/DiaHD”, *Anuario Americanista Europeo*, 12, pp. 5-18.
- ROCKWELL, Geoffrey (2013 [1999]): “Is Humanities Computing an Academic Discipline?”, en Melissa Terras, Julianne Nyhan y Edward Vanhoutte (eds.), *Defining Digital Humanities. A Reader*. London: Routledge, pp. 13-33. DOI: <<https://doi.org/10.4324/9781315576251>>.
- RODRÍGUEZ-YUNTA, Luis (2013): “Humanidades digitales, ¿una mera etiqueta o un campo por el que deben apostar las ciencias de la documentación?”, *Anuario ThinkEPI*, 7, pp. 37-43.
- ROJAS CASTRO, Antonio (2013a): “Las Humanidades Digitales: principios, valores y prácticas”, *Janus*, 2, pp. 74-99.
- (2013b): “El mapa y el territorio: una aproximación histórico-bibliográfica a la emergencia de las Humanidades Digitales en España”, *Caracteres. Estudios Culturales y Críticos de la Esfera Digital*, 2(2), pp. 10-53.
- SÁNCHEZ MÉNDEZ, Juan Pedro; PADRÓN CASTILLA, Elena y PARRO GÓMEZ, Nerea (eds.) (2025): *Sendas del hispanismo: márgenes, centros y convergencias. Actas del XXI Congreso de la Asociación Internacional de Hispanistas* (Neuchâtel 2023). Madrid/Frankfurt am Main: Iberoamericana/Vervuert. DOI: <<https://doi.org/10.31819/9783968696935>>.
- SÁNCHEZ-PRIETO BORJA, Pedro (2012): “La Red CHARTA: proyecto global de edición de documentos hispánicos”, en María Jesús Torrens Álvarez y Pedro Sánchez-Prieto Borja (eds.), *Nuevas perspectivas para la edición y el estudio de documentos hispánicos antiguos*. Bern: Peter Lang, pp. 17-44.
- SCHREIBMAN, Susan; SIEMENS, Ray y UNSWORTH, John (eds.) (2004): *A Companion to Digital Humanities*. Oxford: Blackwell.
- SPENCE, Paul y GONZÁLEZ-BLANCO GARCÍA, Elena (2014): “A Historical Perspective on the Digital Humanities in Spain”, *H-Soz-Kult*, [7 de noviembre de 2014]. <<https://www.hsozkult.de/debate/id/fddebate-132278>>.
- TEI CONSORTIUM (s.f.): Text Encoding Initiative. <<https://tei-c.org/>> [consultado 10/05/2025].
- TERRAS, Melissa; NYHAN, Julianne y VANHOUTTE, Edward (eds.) (2013): *Defining Digital Humanities. A Reader*. London: Routledge. DOI: <<https://doi.org/10.4324/9781315576251>>.
- TORRUELLA, Juan y LLISTERRI, Joaquim (1999): “Diseño de corpus textuales y orales”, en José Manuel Bleuca, Gloria Clavería, Carmen Sánchez y Joan Torruella (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Barcelona: Universidad Autónoma de Barcelona/Milenio, pp. 45-81. <<https://portalrecerca.uab.cat/es/publications/dise%C3%B1o-de-corpus-textuales-y-orales/>>.