DOI 10.14672/2.2025.3216

INTRODUCCIÓN

Carmen Castillo Peña, Università degli Studi di Padova
Alejandro Fajardo Aguirre, Universidad de La Laguna

A Felisa Bermejo, *in memoriam*[1]

El patrimonio lexicográfico, en su sentido más amplio, está constituido por todo texto que recoge unidades léxicas de una o más lenguas para dar cuenta de su significado, ya sea mediante definiciones, ejemplos o por medio de equivalencias intra o interlingüísticas. Esto implica que, junto a los diccionarios alfabéticos —monolingües, bilingües o multilingües, generales, terminológicos, de sinónimos o palabras afines, etimológicos, históricos, etc.—, forman también parte del patrimonio lexicográfico las nomenclaturas, los diccionarios conceptuales o cualquier listado léxico onomasiológico. Asimismo, se incluyen las listas de palabras que forman parte de los paratextos presentes en gramáticas y traducciones.

La preservación de este vasto y heterogéneo conjunto de textos es imperativa, pues los diccionarios —todos, desde los más importantes hasta los más humildes— son obras que conservan no solo el acervo léxico, sino también el patrimonio cultural, al ser testimonio de ideologías, visiones del mundo, conocimientos técnicos y científicos, así como denominaciones de objetos, especies, costumbres y hábitos de una comunidad en un momento histórico determinado.

Efectivamente, los diccionarios son productos metalingüísticos complejos, herramientas para la comprensión y el uso de las lenguas y, a la vez, expresión de la ideología dominante, de la cultura y de la historia de las comunidades lingüísticas. Además, la aproximación historiográfica a los textos lexicográficos permite observar la evolución del léxico y, junto a esta, la historia de los procesos y recursos metalingüísticos utilizados para representar el lenguaje, el devenir de las ideas sobre la lengua objeto y la trayectoria del discurso ideológico que inexorablemen-

1 Apenas concluidas las tareas de este monográfico, nos ha dejado Felisa. Consternados, los miembros del grupo TELEI, al que Felisa dedicó toda su pasión por los diccionarios, los amigos y compañeros que durante años hemos gozado de su saber, de su fina inteligencia, de su exquisita ironía y, sobre todo, de su gran humanidad, le rendimos homenaje.

Recibido: 20 de septiembre de 2025
Aceptación definitiva: 25 de septiembre de 2025

te impregna el diccionario.

Al igual que ocurre en otros ámbitos relacionados con los bienes culturales, materiales o inmateriales, los avances tecnológicos relacionados con la digitalización han supuesto un gran impulso en la preservación de los diccionarios, abriendo el camino a una proficua línea de investigación basada en las metodologías de las humanidades digitales.

En este sentido, la investigación historiográfica en lexicografía se ha beneficiado de la posibilidad de acceder a textos menores o de difícil localización. Esto es posible gracias a las reproducciones digitalizadas de los fondos de las grandes bibliotecas, como, por ejemplo, la *Biblioteca Digital Hispánica* de la Biblioteca Nacional de España, o la *Biblioteca Digital* de la Real Academia Española, pero sobre todo, gracias a las bibliotecas virtuales que catalogan fondos digitales de múltiples procedencias, permitiendo la consulta —y en ocasiones el descubrimiento— de textos digitalizados que de otro modo estarían dispersos en la red, como es el caso de la magnífica *Biblioteca Virtual de la Filología Española* <https://www.bvfe.es/es/>, creada por Manuel Alvar Ezquerra y dirigida hoy por M.ª Ángeles García Aranda o de la biblioteca de diccionarios del portal *Contrastiva* <https://www.contrastiva.it>, dirigido por Félix San Vicente.

Por otra parte, las humanidades digitales han transformado profundamente el propio concepto de diccionario, que ya no es solo un libro impreso, sino una base de datos, normalmente gratuita y de libre acceso, a la que se puede interrogar de varias formas y con diversas finalidades, adaptadas en cada ocasión a las necesidades específicas de distintos tipos de usuarios (Tarp 2019). Se trata de diccionarios proyectados como digitales, sin los vínculos macro y microestructurales que impone el papel. Con frecuencia, constituyen la única vía factible para proyectos que en versión impresa serían casi irrealizables: diccionarios de lenguas indígenas o en vías de extinción, de pares de lenguas con escasa representación, de nueva planta monolingües, terminológicos o para finalidades específicas, con nomenclaturas restringidas a un tipo particular de unidades léxicas o a variedades específicas. Son diccionarios de realización dilatada en el tiempo, que suelen estar sometidos a una actualización constante y que a menudo no cuentan con los recursos organizativos y económicos de una empresa comercial.

Sin pretensiones de agotar el listado posible, damos algunos ejemplos de diccionarios digitales del español, de distintas tipologías, alcances y finalidades:

- *Diccionario electrónico sáliba-español* (Estrada Ramírez) es un diccionario electrónico concebido para la "documentación de la lengua y la cultura" y la "protección del patrimonio idiomático" impulsado por el Instituto Caro y Cuervo.

CUADERNOS AISPI 26 (2025): 9-17
ISSN 2283-981X

- *Diccionario digital del español (DIDES)* (Fuertes Olivera), monolingüe, permanentemente actualizado, descrito por su autor como un "repositorio dinámico de datos lexicográficos (está sometido a un proceso de cambio constante)" y proyectado para el uso de la Inteligencia Artificial Generativa para describir "la realidad del español tal y como la misma se percibe analizando fuentes lexicográficas auténticas, actualizadas y fiables".
- *Diccionario panhispánico de términos médicos* (Real Academia Nacional de Medicina de España 2023), dedicado al léxico científico de la medicina y con vocación panhispánica, "se presenta como una obra de acceso libre y gratuito" elaborada por un equipo integrado por las Academias Nacionales de Medicina de España y americanas, en el que han participado "traductores, informáticos, etimólogos, lexicógrafos y especialistas en codificación".
- *Diccionario de colocaciones del español (DICE)*. Se trata de un diccionario realizado en el marco teórico de la lexicología explicativa y combinatoria, basado en el concepto de función léxica y con una nomenclatura por ahora limitada al campo léxico de los nombres del sentimiento (Alonso Ramos).
- *Diccionario de locuciones idiomáticas del español actual (DiLEA)*. Recoge el tipo de unidades léxicas indicadas en su título, en concreto 9075 entradas y 12 627 locuciones delimitadas diatópicamente en el español de España y de uso actual. El diccionario se actualiza periódicamente por lo que el número de entradas y locuciones aumenta progresivamente (Penadés Martínez 2019).
- *Diccionario de partículas discursivas del español*. Tiene una microestructura ideada para la descripción y ejemplificación del significado procedimental de las partículas. El corpus en el que se basa abarca sobre todo el español de España, pero se actualiza para ampliarlo a otras variedades (Briz; Pons; Portolés 2008).

Por último, está la digitalización de diccionarios, entendida no ya como una reproducción facsimilar digital, sino como la conversión y transformación de un diccionario impreso en un conjunto de datos legibles e interpretables por un sistema informático con la finalidad de obtener una edición digital interrogable. El producto final es similar al de un diccionario nacido digital, pero la base de partida es un diccionario impreso existente, tratado con el rigor historiográfico necesario en función de su datación, sus características lingüísticas y textuales y la finalidad misma de la edición digital, de ahí que se use el término *retrodigitalización*. Este neologismo designa el conjunto de operaciones informáticas,

ecdóticas y filológicas que permiten al investigador y al usuario interactuar con el diccionario editándolo, integrándolo con otros diccionarios, haciendo búsquedas complejas o, simplemente, consultándolo.

Son numerosos los diccionarios retrodigitalizados presentes en la red. Además de los que ha publicado la Real Academia Española, citamos los siguientes a título ilustrativo:

- *Diccionario del español de México* (en línea). Se trata de la versión electrónica de la segunda edición del diccionario (Lara 2024), con la posibilidad de búsquedas avanzadas y la presencia de recursos añadidos: listas de sufijos, conjugaciones, escritura de los numerales, el enlace a la presentación de otras investigaciones del Colegio de México y al Corpus del español de México contemporáneo[2]. Además, el usuario puede hacer propuestas o plantear dudas al equipo lexicográfico a través de un formulario.
- *Diccionario del español actual* (DEA). Esta tercera edición en versión electrónica es, a nuestro juicio, un caso híbrido, ya que parte de la retrodigitalización de las ediciones impresas del diccionario publicadas en 1999 y 2011 (Seco; Ramos; Andrés 1999), si bien el producto final se considera una edición nueva "notablemente aumentada y puesta al día" (Andrés 2023).
- *e-DRAE1884*. Con finalidad historiográfica, de sumo interés para el investigador en historia de la lexicografía y del léxico, consiste en una edición digital del DRAE de 1884, realizada a raíz de un proyecto dirigido por Gloria Clavería Nadal y Margarita Freixas Alás (Clavería Nadal; Freixas Alás).

Así pues, pueden distinguirse tres grandes tipos de productos, originados por la interacción entre las humanidades digitales y la lexicografía: las reproducciones facsimilares digitales, los diccionarios digitales y los diccionarios digitalizados o retrodigitalizados.

Con respecto a este último tipo, una de las aplicaciones que más se beneficia de la retrodigitalización de diccionarios es la elaboración de tesoros lexicográficos. En la tradición lexicográfica hispánica, el término *tesoro*, además de referirse a un cierto diccionario (como el de Sebastián de Covarrubias, que así lo indica en el

2 En el momento de la redacción de esta introducción, el enlace al corpus está inactivo porque envía a la primera versión del Corpus, que ha sido actualizada con una segunda versión publicada en una nueva URL <https://cemcii.colmex.mx/index.html>. Mencionamos la circunstancia, muy común por otra parte, no para disminuir el valor de este excelente diccionario en línea, sino como ilustración anecdótica de uno de los grandes problemas de las humanidades digitales, como es la necesidad de una actualización constante —en especial la de los enlaces—.

CUADERNOS AISPI 26 (2025): 9-17
ISSN 2283-981X

título: *Tesoro de la lengua castellana, o española*) se utiliza para denominar un *corpus glosariorum* (Nieto Jiménez; Alvar Ezquerra 2007: IX), esto es, un diccionario de diccionarios. Estos repertorios compilan y ordenan de forma exhaustiva, normalmente por criterios cronológicos, la información lexicográfica de una lengua. Su función principal es descriptivo-documental, con aplicaciones historiográficas y metalexicográficas que permiten rastrear tanto de las influencias de unos diccionarios en otros, como de las innovaciones. Entre los principales tesoros lexicográficos de la lengua española se encuentran los siguientes, los tres primeros, publicados solo en formato impreso, y los restantes, en versión digital:

- *Tesoro lexicográfico (1492-1726*) de Samuel Gili Gaya (1960). Obra de gran trascendencia para la historia de la lexicografía, a pesar de haber quedado inconclusa, pues solo llega hasta la letra E, de estar cronológicamente limitada a los diccionarios anteriores a la publicación del denominado *Diccionario de Autoridades* de 1726 y de no ser exhaustiva, pues son muchos los diccionarios no incluidos en el corpus del insigne lexicógrafo.
- *Nuevo Tesoro Lexicográfico del español (s. XIV-1726)* de Lidio Nieto Jiménez y Manuel Alvar Ezquerra (2007). Es una obra monumental, con un corpus de 145 repertorios, unas cien mil entradas, que llegan a casi un millón si se cuentan variantes según autor y más de seiscientas mil referencias de diferentes textos. Su corpus alberga diccionarios alfabéticos y conceptuales, monolingües, bilingües y multilingües, repertorios de ámbito diatópico, diccionarios de lenguas especiales, glosarios, etc.
- *Tesoro lexicográfico del español de Canarias* de Cristóbal Corrales Zumbado, Dolores Corbella Díaz y M.ª Ángeles Álvarez Martínez (1992). Además de recoger las palabras de la tradición diccionarística del español de Canarias, incluye las aportaciones de estudios lexicológicos, especialmente la del *Atlas Lingüístico y Etnográfico de las Islas Canarias* (Alvar 1975-78).
- *Nuevo Tesoro Lexicográfico de la Lengua Española* (NTLE) de la Real Academia Española. Es, probablemente, el *Tesoro* más conocido. A pesar de sus indiscutibles méritos, su corpus es bastante limitado, al menos en relación con el *Tesoro* de Nieto y Alvar, solo se pueden buscar los lemas, no ofrece la posibilidad de la búsqueda avanzada y los resultados que proporciona consisten en las imágenes facsimilares de los diccionarios.
- *Tesoro lexicográfico médico (TeLeMe)* de Bertha Gutiérrez Rodilla. El Corpus comprende ocho diccionarios terminológicos de medicina originales, no traducidos, publicados entre mediados del XVIII y primeros años del XIX. Permite búsquedas avanzadas y, como el *NTLE* de la RAE, ofrece

como resultado el facsímil de la página en la que se encuentra la voz.

- *Tesoro lexicográfico del español de Puerto Rico en línea* de la Academia Puertorriqueña de la Lengua Española (2020). Actualmente contiene unas treinta mil palabras y frases puertorriqueñas, provenientes de un corpus de sesenta y cinco textos lexicográficos, pero también de investigación lexicológica, sobre el español de Puerto Rico escritos entre 1788 y 2022. Está basado en la retrodigitalización del *Tesoro lexicográfico del español de Puerto Rico* (Vaquero; Morales 2005), a la que se han añadido fuentes nuevas en un proceso de actualización constante.

El proceso de retrodigitalización pasa por dos grandes fases operativas: la primera consiste en la transcripción de los textos y la segunda en su codificación en un lenguaje de marcado XML Para esta tarea se suele usar TEI, el estándar más extendido en el ámbito académico. Si bien la digitalización de un diccionario prevé unas normas de transcripción y codificación coherentes con su datación y del número de ediciones, la elaboración de un tesoro implica, además, un proceso de armonización y estandarización de criterios. Esto es esencial, ya que las normas adoptadas deben ser válidas para todos los diccionarios que lo conforman.

Los artículos de esta sección monográfica describen estos procesos —métodos para la transcripción automatizada, codificación en XML y armonización de esta para la construcción de un tesoro— ya que todos ellos plantean reflexiones y resultados en relación con dos proyectos de investigación en curso dedicados a la elaboración de tesoros digitales: el *Tesoro lexicográfico del español de América* (TLEAM) y el *Tesoro digital de la lexicografía bilingüe español-italiano* (TELEI). De estos trabajos se desprende una idea central: la retrodigitalización de diccionarios antiguos no consiste en una mera conversión de formatos (del papel a la pantalla). Al igual que cualquier otro texto sometido a un proceso de edición filológica, requiere un estudio crítico previo y, en el caso de los diccionarios, también un análisis metalexicográfico para determinar la naturaleza de los paratextos, las características de la macroestructura, la nomenclatura, y, sobre todo, la microestructura.

Varios de los artículos derivados del proyecto del TELEI se centran en la fase de transcripción. Alemany Martínez, De Beni y La Manna analizan con detalle las fases del trabajo necesario para la digitalización del *Vocabolario italiano e spagnuolo* (1620) de Lorenzo Franciosini, desde la conversión en formato texto con el uso del software de OCR Transkribus, hasta las decisiones editoriales relativas a la puntuación, la acentuación o el uso de las mayúsculas o el análisis interpretativo

CUADERNOS AISPI 26 (2025): 9-17
ISSN 2283-981X

del enunciado lexicográfico necesario para el etiquetado en XML TEI. Por su parte, Ferrante, Valente y de Hériz exponen con abundancia de detalles el minucioso proceso de análisis, ecdótico y metalexicográfico, previo a la digitalización del *Diccionario de faltriquera* de Cormon y Manni de 1805. Las autoras explican una serie de casos emblemáticos en los que se observan con claridad las dificultades que presenta la transcripción automática de un texto lexicográfico, incluso del siglo XIX, realizada con el software Transkribus. Asimismo, Bermejo Calleja, Lanteri, Valero Gisbert y Zaccone dedican una buena parte de su ensayo al uso de *Transkribus* en la transcripción del diccionario bilingüe de Ambruzzi y al análisis de fenómenos macroestructurales de este diccionario. Se enfocan en el uso de las abreviaturas, que no son lo uniformes que se esperaría en un diccionario de mediados del siglo XX, proponiendo soluciones para estandarizar la marcación.

La transcripción requiere un análisis crítico del texto que, como se apuntaba antes, no tiene solo un carácter estrictamente lingüístico (estado de lengua, grafemas, puntuación, erratas y errores, variantes morfoléxicas, etc.) sino también metalexicográfico (tipos de lematización, tipos de definición o de equivalencias, marcación, abreviaturas utilizadas, etc.). Este tipo de enfoque es el que ofrecen los artículos de Lombardini y de Peñín Fernández, ambos dedicados a la lexicografía bilingüe no alfabética. Lombardini estudia la marcación del género gramatical en la nomenclatura incluida en *L'italiano istruito nella cognizione della lingua spagnuola* de Francisco Marín (1833), ya que, como se sabe, las nomenclaturas tienen una microestructura mínima, normalmente reducida al lema y a la equivalencia interlingüística, por lo que, la información gramatical, si la hay, es irregular y de expresión variable, ciertamente no normalizada con una abreviatura. Peñín Fernández, por su parte, dedica su artículo a un análisis crítico completo del *Vocabulario Espanyol Italiano y Tudesco* de Ignacio de Boria (1719), una nomenclatura trilingüe español-italiano-alemán que aporta la peculiaridad de su condición inédita.

La fase de codificación y cómo resolver la necesaria homogeneización para la constitución de un tesoro digital son los puntos clave de los cuatro artículos siguientes; el primero está dedicado a un diccionario del TLEAM y los otros tres, a diccionarios del corpus de TELEI.

Díaz Rodríguez, tras reseñar brevemente la finalidad, las características y las fases de elaboración del TLEAM, se concentra en el proceso de homogeneización de los materiales que constituyen el corpus de este Tesoro, a partir del estudio de caso del *Diccionario de voces americanas* de Manuel José de Ayala (1777), tratando en particular la normalización de la ortografía, la lematización geminada, la du-

plicidad de lemas y la codificación de las unidades poliléxicas. A este último tema, está también dedicado el trabajo de Dalle Pezze y Sartor, quienes analizan con detalle una de las cuestiones más relevantes del *Vocabolario* de Franciosini, como es la fraseología, o el discurso repetido, siguiendo la terminología coseriana utilizada por las autoras. El artículo da cuenta de la dificultad que presenta el etiquetado metalexicográfico de este tipo de estructuras, difícilmente sistematizables en los diccionarios antiguos. Para ello, las autoras trabajan con un subcorpus piloto en el que analizan la técnica lexicográfica empleada por Franciosini para lematizar y establecer equivalencias, así como las condiciones que debe cumplir su etiquetado en XML-TEI Lex0, que es el estándar de codificación utilizado en TELEI. En la misma línea, García Jiménez y Pérez Vázquez describen las características específicas del *Vocabulario italiano-español* del *Diccionario marítimo* de O'Scanlan publicado en 1831 y discuten las propuestas de codificación para los lemas pluriverbales, lemas geminados o las equivalencias definicionales de este diccionario de especialidad, bilingüe, aunque no bidireccional. Por último, también Castillo Peña aborda con detalle los problemas de codificación que presentan las nomenclaturas en TELEI, en cuanto diccionarios conceptuales cuya microestructura es difícilmente homologable a la de los diccionarios alfabéticos.

En conclusión, estos trabajos ponen de manifiesto, por un lado, la gran variedad genérica y cronológica de los textos estudiados: diccionarios del español hablado en América, diccionarios bilingües, diccionarios de especialidad, nomenclaturas, glosarios incluidos en gramáticas, textos impresos de gran trascendencia, textos menores y textos inéditos. Por otro lado, ponen de relieve la necesidad de un conjunto específico de recursos metodológicos y tecnológicos propios de las humanidades digitales. Los procesos de digitalización, retrodigitalización y armonización en la construcción de tesoros lexicográficos evidencian que la edición crítica de diccionarios requiere no solo un tratamiento filológico riguroso, sino también una reflexión metalexicográfica. Esta reflexión es fundamental para garantizar su coherencia y su utilidad para la investigación. Por tanto, estos proyectos en curso confirman la oportunidad de este enfoque para asegurar la preservación y la proyección futura de los repertorios lexicográficos, consolidando así un campo de estudio que consideramos esencial para la historiografía lingüística y para la comprensión de la tradición lexicográfica en su conjunto.

CUADERNOS AISPI 26 (2025): 9-17
ISSN 2283-981X

Referencias bibliográficas

Academia Puertorriqueña de la Lengua Española, (2020), *Tesoro lexicográfico del español de Puerto Rico en línea*. [29/03/2025] <https://tesoro.pr/>

Alonso Ramos, Margarita; Grupo DICE (dir.ª) (en línea), *Diccionario de colocaciones del español (DICE)* [26/06/2025] < http://www.dicesp.com>.

Alvar, Manuel (1975–78), *Atlas Lingüístico y Etnográfico de las Islas Canarias*, Las Palmas de Gran Canaria, Cabildo Insular de Gran Canaria.

Andrés, Olimpia (dir.ª) (2023), *Diccionario del español actual* (edición electrónica). Fundación BBVA [14/06/2025] < https://www.fbbva.es/diccionario>.

Briz, Antonio; Pons, Salvador; Portolés, José (coords.) (2008), *Diccionario de partículas discursivas del español.* [04/04/2025] <www.dpde.es>.

Corrales Zumbado, Cristóbal José; Corbella Díaz, Dolores; Álvarez Martínez, María Ángeles (1996), *Tesoro lexicográfico del español de Canarias*, Madrid, Real Academia Española.

Clavería Nadal, Gloria; Freixas Alás, Margarita (en línea) *e-DRAE 1884* [15/04/2025] <https://edrae1884.uab.cat>.

Estrada Ramírez, Hortensia (coord.) (en línea), *Diccionario electrónico sáliba-español: una propuesta de documentación de la lengua y la cultura sálibas*, Bogotá, Instituto Caro y Cuervo [24/06/2025] < https://saliba.caroycuervo.gov.co>

Fuertes Olivera, Pedro (en línea), *Diccionario Digital del Español (DIDES)*. [24/06/2025] <https://diesgital.com/>.

Gili Gaya, Samuel (1960), *Tesoro lexicográfico (1492-1726)*, Madrid, Consejo Superior de Investigaciones Científicas.

Gutiérrez Rodilla, Bertha M. (dir.ª) (en línea) *Tesoro lexicográfico médico (TeLeMe)* [en línea]. [28/04/2025] <http://teleme.usal.es>

Lara, Luis Fernando (dir.) (2024), *Diccionario del español de México* (*DEM*), México DF, El Colegio de México. Versión en línea [15/05/2025] <https://dem.colmex.mx>.

Nieto Jiménez, Lidio; Alvar Ezquerra, Manuel (2007), *Nuevo Tesoro Lexicográfico del español (s. XIV-1726)*, Madrid, Arco Libros.

Penadés Martínez, Inmaculada (2019), *Diccionario de locuciones idiomáticas del español actual (DiLEA)* [03/04/2025] <www.diccionariodilea.es>.

Real Academia Española (en línea), *Nuevo tesoro lexicográfico de la lengua española (NTLLE)*, [04/06/2025] <https://apps.rae.es/ntlle/SrvltGUISalirNtlle>.

Real Academia Nacional de Medicina de España (2023), *Diccionario panhispánico de términos médicos* [25/05/2025] <https://dptm.es/>.

Seco, Manuel; Andrés, Olimpia; Ramos González, Gabino (1999), *Diccionario del español actual*, Madrid, Aguilar.

Tarp, Sven (2019), "La ventana al futuro: despidiéndose de los diccionarios para abrazar la lexicografía", *RILEX. Revista sobre investigaciones léxicas*, 2: 5-36.

Vaquero, María; Morales, Amparo (2006), *Tesoro Lexicográfico del Español de Puerto Rico*, San Juan, Academia Puertorriqueña de la Lengua Española.