

Recuperación del léxico especializado del *Diccionario Tecnológico Hispano Americano* (1930). Diseño de un proyecto XML-TEI

María Rosa Castro-Prieto
Universidad Autónoma de Madrid

Recibido: 02/08/2022

Aceptado: 28/09/2022

Resumen: Los diccionarios publicados antes de la llegada del mundo digital están condenados a caer en el olvido debido al soporte en papel que los alberga. Si a esta particularidad del soporte le añadimos el componente diacrónico, tenemos el cóctel perfecto para que prácticamente desaparezca la accesibilidad a esas obras. Consideramos que el estudio de los diccionarios elaborados y publicados en otras épocas proporciona datos valiosos para estudiar el léxico desde una perspectiva histórica. La propuesta que presentamos a continuación constituye el inicio de un proyecto más amplio que pretende recuperar un diccionario publicado en 1930 y que supuso el primer intento institucional de recopilar el léxico científico y técnico en la confluencia de los siglos XIX y XX, un periodo especialmente relevante para la historia de la lengua de especialidad en español. En las siguientes páginas proponemos el diseño de un proyecto que aspira a recuperar el contenido léxico de este diccionario mediante su codificación con el estándar XML-TEI, preservarlo y ponerlo a disposición del público para su consulta.

Palabras clave: historia de la lengua española; léxico especializado; diccionario científico-técnico; XML-TEI; humanidades digitales.

Abstract: Dictionaries published before the arrival of the digital world are likely to fall into oblivion due to the paper medium. If we add the diachronic component, we have the perfect mix for the accessibility of these works to practically disappear. We believe that the study of dictionaries produced and published in other periods provides valuable data for studying the lexicon from a historical perspective. The proposal presented below is the beginning of a wider project that seeks to recover a dictionary published in 1930, which was the first institutional attempt to compile the scientific and technical lexicon at the confluence of the nineteenth and twentieth centuries, a particularly relevant period for the history of the Spanish specialty language. In the following pages we propose the design of a project that aims to recover this dictionary by encoding it with the XML-TEI standard, preserve it and make it available to the public for its consulting.

Keywords: Spanish language history; specialised lexicon; scientific and technical dictionary; XML-TEI; digital humanities.

1. INTRODUCCIÓN

El estudio del léxico especializado del ámbito de la ciencia y la técnica resulta imprescindible para completar el paisaje de la historia de la lengua española y de su lexicografía, pues desde el análisis del pensamiento tecnocientífico se puede explicar la modernización de la lengua y la sociedad. Este es el motivo principal que nos impulsa a abordar esta tarea pendiente, con la que pretendemos aportar evidencias que completen, desde un punto de vista lexicográfico, la historia del léxico especializado de la lengua española. Importantes autores (Gutiérrez Cuadrado, 1978; Garriga Escribano, 2015; Gutiérrez Rodilla, 2003; Gutiérrez Rodilla, 2017; Gutiérrez Cuadrado y Garriga Escribano, 2019) han manifestado que la historia de la lexicografía, la lengua científica o la historia de la ciencia españolas están insuficientemente examinadas. El estudio de los diccionarios científico-técnicos de épocas pasadas, en particular, ha sido de los que menos atención ha suscitado, si exceptuamos la gran labor que se lleva a cabo dentro del grupo Neolcyt¹, pues de los trabajos que se ocupan del léxico especializado la abrumadora mayoría de ellos lo hacen en el marco

¹ El grupo Neolcyt (Grupo de investigación en lengua de la ciencia y de la técnica) pertenece a la Universidad Autónoma de Barcelona. Se puede consultar más información en su página web <https://neolcyt.net/>

de los diccionarios generales y apenas se encuentran trabajos publicados que se ocupen de los diccionarios especializados.

Si tenemos en cuenta que la lengua de especialidad recorre el mismo camino que las áreas de conocimiento que simboliza y, por tanto, sufre los mismos procesos de obsolescencia, cambio o novedad que estas, los diccionarios en papel que recogen este acervo publican estas transformaciones léxicas que, de este modo, quedan recogidas de manera inalterable bajo una estructura que se constituye como representante espacio-temporal al proporcionar una fotografía del estado del pensamiento especializado, materializado mediante el léxico que lo vehicula. Los diccionarios, entonces, se pueden interpretar como recurso cultural y, contemplados históricamente, se convierten en un valioso instrumento de reproducción y usos de la lengua. Su análisis evidencia la relación que establecen con el espacio histórico y sociocultural que les corresponde, pues reflejan el conocimiento del tiempo al que pertenecen (Dubois & Dubois, 1971; Le Pesant, 2020), influyen en la sociedad (Alvar Ezquerro, 1993; Rodríguez Barcia, 2002) y encarnan una cosmovisión particular (Forgas Berdet, 2001; 2006), por lo que su estudio nos abre una puerta que nos permite entrar en otras épocas.

Recientemente asistimos a un despegue de las denominadas Humanidades Digitales, disciplina en la que convergen los estudios, métodos y perspectivas de las áreas humanísticas tradicionales atravesadas por las aplicaciones informáticas. Este cruce de disciplinas tan dispares ha producido nuevos puntos de vista metodológicos que permiten acercarse a los estudios humanísticos de otra manera y mejorar investigaciones que antes eran difíciles de abordar. Los numerosos lenguajes informáticos que están a disposición del investigador permiten realizar tareas complejas con las que observar los datos desde otras perspectivas que a la vez revelan nuevas relaciones entre ellos, convirtiéndolos en nuevos datos cuyas ramificaciones enriquecen nuestra comprensión. Desde este punto de vista, destaca un estándar de codificación textual concebido hace unos 30 años, consensuado a nivel internacional y ampliamente difundido entre la comunidad científica. La *Text Encoding Initiative* (TEI), que es como se denomina este estándar, es un complejo lenguaje de marcado y etiquetado textual que, gracias a su flexibilidad y capacidad de adaptación, facilita adoptar diferentes puntos de vista según los intereses del investigador, lo que le ha permitido ganarse la consideración de los investigadores. En este trabajo adoptaremos este estándar, pues nos permitirá articular la metodología que nos guiará en nuestro proyecto.

En las siguientes páginas nos proponemos presentar los preliminares de una iniciativa en la que convergen las Humanidades Digitales y los diccionarios científico-técnicos redactados y publicados en lengua española en la confluencia de los siglos XIX y XX. Este periodo intersecular se corresponde con una época clave para el pensamiento científico español (Sánchez Ron, 2020: 349-356), a la vez que, para la lexicografía, pues este momento de transición entre siglos constituye una de las etapas más prolíficas de la lexicografía española tanto académica como extracadémica (Ahumada Lara, 2007; Ahumada Lara, 2016). El estudio diacrónico de los diccionarios proporciona un inventario de léxico a través del que es posible observar, entre otros aspectos, qué vocablos aparecen en textos científicos o divulgativos y cuáles caen en desuso debido a los cambios que se producen en la sociedad. En el caso de los diccionarios especializados este examen es especialmente importante, pues se formalizan como una herramienta que permite medir el grado de progresión del pensamiento científico.

El objetivo del presente artículo consiste en presentar el prototipo de un proyecto de estudio y preservación lexicográfica del léxico de la ciencia y la técnica que se corresponde, *grasso modo*, con el primer tercio del siglo pasado. Abordar un proyecto tan ambicioso desde una ejecución particular es una tarea demasiado costosa y compleja, por lo que tomaremos como estudio de caso la transcripción, el modelado y la edición, en el estándar de codificación XML-TEI, del *Diccionario Tecnológico Hispano Americano*, concebido por Leonardo Torres Quevedo, redactado por la Unión Internacional Hispano-Americana de Bibliografía y Tecnología Científicas y publicado en 1930 por la Editorial Arte y Ciencia. La elección del citado diccionario se debe a que representa la intención de institucionalizar la recopilación del léxico científico español desde una perspectiva globalizadora, pues lo sitúa como primera tentativa de recopilación de léxico especializado en todas sus variedades, lo que sin duda «tiene una gran importancia en la historia de la lexicografía española, ya que se trata de un intento de establecer una pauta panhispánica en el vocabulario científico y técnico» (Garriga y Pardo 2014: 201). Consideramos también que profundizar en el estudio de una herramienta de recopilación léxica como es el diccionario especializado contribuye a reforzar los estudios de historia de la lengua y la ciencia españolas.

Se ha organizado el presente trabajo en cinco apartados, en los que se presentará el *Diccionario Tecnológico Hispano Americano (DTHA)* en su contexto sociocultural, con algunas precisiones acerca de las motivaciones de su génesis y unos breves apuntes acerca del

contenido lexicográfico de la obra. El aparato metodológico introducirá brevemente el lenguaje de marcado XML-TEI y su aplicación a otros proyectos de Humanidades Digitales orientados a la codificación y publicación digital de diccionarios. En el apartado de «Diseño» se ha articulado conforme al flujo de trabajo de la edición digital y se comentan qué etapas componen el proyecto, qué aspectos se tratarán y cómo se han comenzado a abordar algunas de ellas. Se finalizará con unos apuntes a modo de conclusión, mezcla de certezas y de deseos derivados de los objetivos del proyecto.

2. EL *DICCIONARIO TECNOLÓGICO HISPANOAMERICANO (DTHA)*

En el siguiente apartado se esboza la génesis del *Diccionario Tecnológico Hispanoamericano* desde una perspectiva histórica que, a pesar de su fecha de publicación en 1930, lo ubica en un espacio temporal a caballo entre dos siglos, pues es deudor de las ideas de progreso y de la percepción pública que se tenía de la ciencia en el siglo XIX. El interés por la unidad de la lengua española más allá del territorio español, la preocupación por la invasión de extranjerismos o el temor por la incorrecta formación de los neologismos en el ámbito científico y técnico son algunos de los motivos por los que un grupo de científicos liderados por el ingeniero Leonardo Torres Quevedo (1852-1936) se propusieron redactar un compendio que reuniera las voces de la ciencia y de la técnica en lengua española. Este conjunto de léxico especializado se articuló en torno a una selección de varias materias de conocimiento y en cuyas definiciones no faltaron referencias enciclopédicas, apuntes normativos, advertencias ortográficas, recomendaciones de uso, correspondencias georreferenciales, observaciones de autoridades en la materia, así como etimologías, marcas de clase de palabra, nombres científicos o fórmulas físicas, químicas y matemáticas.

2.1. Cuestiones socioculturales

Ampliamente analizadas desde los espacios político-sociales y literario-artísticos, las primeras décadas del siglo XX constituyen un momento sociocultural excepcional en España al que, desde algunos sectores, se ha llamado *Edad de plata* (Mainer, 1981). Otero Carvajal (2001: 151) amplía la denominación y la aplica al desarrollo que experimentó el conocimiento científico español en esa época,

pasando a nombrar al período como *Edad de plata de la ciencia*. Explorar esta época desde el ámbito de la ciencia y la técnica resulta fundamental para el estudio de la lengua española, pues el pensamiento tecnocientífico, su divulgación y su recopilación lingüística se convierten en indicadores clave para explicar la modernización de la sociedad española de aquel momento único (Castro-Prieto, 2018:39-40). En ese sentido, se puede afirmar que los redactores del *DTHA* se propusieron recoger el saber y la experiencia de la que ha sido, como afirma Sánchez Ron (2020), la última época más prolífica de la ciencia y la tecnología de producción española, por lo que cobra una especial relevancia aproximarse a esta obra considerándola receptora del saber de una época.

Como señala Forgas Berdet (2001 párrafos 1 y 2; 2006), los diccionarios cumplen con un papel cultural ineludible y se hacen eco de la sociedad y de un tiempo concreto, pues nos ofrecen un paisaje social observable a través del prisma lexicográfico, por lo que se convierten, como tantas veces se ha dicho, en reflejo de la sociedad a la que pertenecen. A este respecto, Dolores Azorín (1990: 445; 2001: 215) defiende que la incorporación del léxico de especialidad al diccionario está condicionada por «factores de índole social y cultural», a lo que añade Battaner que el valor de lo que se considera especializado puede cambiar con el tiempo al estar subordinado «a cambios históricos de valoración (atención)» (Battaner, 1996: 101). No hay que perder de vista que los diccionarios especializados se ofrecen como un recurso de apoyo a la construcción del discurso científico que, a la vez, lo acerca al gran público. Esta doble dimensión los convierte en un instrumento con la capacidad de medir el grado de avance del léxico especializado de una lengua y del conocimiento que transmite.

El interés por crear e introducir léxico científico y técnico propio en la lengua española fue un denominador común en el siglo XVIII, que decreció durante la primera mitad del siglo XIX, y que, a medida que avanza el siglo, se convirtió en una inquietud, ya que, como señala Gutiérrez Cuadrado, los eruditos de la época

se aplicaron con constancia a dos tareas diferentes, pero muy interrelacionadas: a) la discusión teórica del problema —de la que concluían la necesidad de un diccionario normativo tecnológico—; b) la redacción de diccionarios tecnológicos que defendieran la pureza de la lengua (1989: 481).

Como resultado, la incorporación de voces científicas y técnicas en los diccionarios generales fue una constante a la que se sumó la publicación de diccionarios especializados (Álvarez de Miranda, 2011: 167-184; Garriga y Pardo, 2014: 204-204), que, dicho sea de paso, tuvieron mucho éxito comercial, fruto del interés suscitado por el público general debido a la consideración que este tenía de los avances científicos. De este modo el diccionario encaja perfectamente como herramienta indispensable para lograr la difusión del conocimiento científico. En este sentido, el diccionario enciclopédico «se ofrece como instrumento polivalente, compendio de todos los saberes» (Azorín, 2001: 255) y es que el aumento del léxico científico y técnico de una lengua va de la mano de los adelantos y nuevos logros científicos. Lafuente y Saraiva (2002: 22-31) exponen que, a lo largo del siglo XIX, se desencadena un proceso de transformación de la vida urbana vertebrado a través del concepto de «progreso», mediante el cual se interviene políticamente en la vida diaria de los ciudadanos gracias a los grandes cambios proyectados a través del trabajo de ingenieros y médicos. Estos mismos autores señalan que las ciudades, en oposición a la centuria anterior, se convierten en un nuevo espacio del saber en las que los cafés, la prensa generalista, las revistas especializadas y el mundo académico funcionan como divulgadores científicos, pues la ciencia pasa a ser un asunto público.

La ciencia se hace urbana y la urbe quiere ser racional. Los científicos se convierten en personajes públicos y los políticos se rodean de sabios y técnicos. Ciertamente, el siglo XIX es la edad dorada de la divulgación, y ello en un doble sentido, pues las ideas llegan a la calle y la calle es el principal objeto de las ideas (Lafuente y Saraiva, 2002: 27).

No es de extrañar que esta visión tanto pública como académica de la ciencia tuviera también su repercusión en las obras lexicográficas y en la necesidad de difundir y ordenar el discurso científico mediante la sistematización de su léxico y la defensa de la integridad y «pureza» de la lengua frente a la incorporación de voces nuevas originarias de otras lenguas. Queda patente que estos planteamientos lingüísticos no son nuevos, pues conforman la herencia ideológica del siglo XVIII, se defienden en el siglo XIX y continúan vigentes, con matices importantes, hasta nuestros días, pues aún hoy se sigue argumentando que la mejor manera de recoger, fijar y controlar el léxico científico pasa, ineludiblemente, por la confección lexicográfica.

2.2. Necesidad de unificar la lengua española de la ciencia y la tecnología

Este razonamiento alcanza también a Leonardo Torres Quevedo (Gutiérrez Cuadrado, 1989: 492) y lo anima a convertirse en su defensor. La figura de este ingeniero ocupa un lugar relevante en la sociedad española de principios del siglo pasado (Sánchez Ron 2020: 748), no solo en el ámbito científico, sino también en el lingüístico, pues ocupó los sillones de las Reales Academias de Ciencias y de Letras (González de Posada, 1992). Se podría decir que Torres Quevedo no solo representa un puente entre las dos instituciones, sino que sirve de punto de intersección entre las necesidades denominativas de la ciencia y las inquietudes sociales y lingüísticas de la época. A propósito de su iniciativa lexicográfica y de la importancia que esta tuvo, Garriga y Pardo (2014: 202) afirman que la confección del *DTHA* «constituye un intento de elaborar un diccionario especializado a comienzos del siglo XX para una lengua como el español, receptora de terminología, que se ve en la necesidad de ordenar el vocabulario especializado que entra debido al progreso de la ciencia y de la técnica». Álvarez de Miranda (2011: 178) enmarca la propuesta de este diccionario en el ámbito de la «lexicografía de carácter corporativo o institucional» y denomina al léxico recogido «transdisciplinar o multiespecializado», es decir, «no de una especialidad, sino, reunidas, todas las especialidades, o al menos muchas de ellas» (2011: 178).

Se puede rastrear el origen de esta aspiración lexicográfica en una serie de congresos independientes entre ellos, celebrados en torno a 1900, cuyo propósito común era el de reunir a estudiosos de ambos lados del Atlántico bajo la premisa de la defensa de la unidad de la lengua española científico técnica (Gutiérrez Cuadrado, 1989: 492-495; Gutiérrez Rodilla, 2017; Gutiérrez Cuadrado y Garriga, 2019: 207). Su importancia para la elaboración del diccionario se puede rastrear en la propia Advertencia del *DTHA*, en la que se menciona que «[e]s muy de sentir que la voz de los Congresos internacionales, siquiera se vea de cuando en cuando transformada en acuerdos solemnes, enmudezca en los archivos sin haber hallado otro eco material que la letra impresa de los anales o memorias publicados sobre tales Congresos» (Advertencia, 1930: V). Aunque lo más probable es que se refiera a congresos especializados en áreas temáticas específicas, es posible que se haga un guiño a estas reuniones panamericanas entre las que destacan, precisamente por haber llegado a acuerdos solemnes, el «Congreso Literario Hispano-Americano», celebrado 1892, que

reunió tanto a especialistas de la lengua, la literatura y el arte como de las ciencias (médicos, ingenieros, botánicos, etc), y el «Congreso Científico Internacional de Buenos Aires», que se celebró en 1910, y al que acudió Leonardo Torres Quevedo como representante de la delegación española, en la que lideró la creación del proyecto de la «Unión Internacional Hispanoamericana de Bibliografía y Tecnología científicas».

En ambos congresos los participantes manifestaron su preocupación por la potencial fragmentación de la lengua española, en especial la científico-técnica, debido a la gran recepción de neologismos que admitía (Gutiérrez Cuadrado y Garriga, 2019: 207), y se buscó la complicidad de los colegas americanos para preservar la unidad de la lengua, pero contando siempre con que la variedad preeminente era la utilizada en España (Gutiérrez Rodilla, 2017: 594-596). Estas motivaciones se trasladan a la génesis del diccionario, ya que recoge las funciones redactadas en las actas de la primera junta de la «Unión Internacional Hispanoamericana de Bibliografía y Tecnología científicas», que, según Gutiérrez Cuadrado (1989: 485-490), pueden resumirse en el carácter unificador de la lengua española para ambos continentes, y en su uso y fomento en la redacción de las publicaciones científicas con el fin de conseguir la independencia nacional. Ortiz (1988: 138) sugiere razones de carácter socioeconómico ocultas detrás de esta actitud, ya que la creación de un diccionario tecnológico español, además de preservar la unidad de la lengua, «facilitaría las relaciones comerciales en la industria y el intercambio de tecnologías».

Respecto a la consecución del proyecto, su puesta en marcha y su desarrollo comenzó desde el mismo momento en el que Torres Quevedo fue miembro de número de la Real Academia Española. En efecto, el 27 de noviembre de 1920 «con el fin de facilitar las relaciones entre americanos y españoles» (Advertencia, 1930: X), se inician los trabajos de la «Junta Nacional de Bibliografía y Tecnología Científica», en cuya acta constitutiva de 1922 se puede leer cómo se insta a los representantes americanos «a que comiencen a trabajar en la forma más práctica y rápida» (Advertencia, 1930: X). Como muestra de este incansable trabajo, entre 1921 y 1926 se habían añadido 25.000 voces científicas y técnicas. El objetivo final era el de pasar de las 100 000 palabras, reunidas por «300 científicos españoles e hispanoamericanos» (Unión Internacional, 1926: 14) agrupados en diez secciones temáticas que aglutinaban más de 80 disciplinas científicas. Auténtico acontecimiento social, en el año 1926 se presentaba ante el rey Alfonso XIII el primer cuadernillo del *Diccionario Tecnológico*

Hispano Americano en un acto organizado por la Real Academia Española (Olagüe, 2006: 91; Garriga y Pardo, 2014: 208). Los discursos pronunciados en este acto tuvieron como denominador común la unidad y la pureza de la lengua española como consolidadora de la cultura hispánica e identificadora del progreso nacional, así como el reconocimiento del mismo estatus para la lengua científica del que goza la lengua común. En los años siguientes se sucede la publicación de otros cuadernillos, hasta alcanzar la cifra de seis. Los cuatro primeros se publican como primer tomo en 1930, editados por Arte y Ciencia, editorial a la que ya estaba vinculado el redactor-jefe Pelayo-Vizuette. En cambio, los cuadernillos quinto y sexto no logran formar parte del que se supone que debería haber sido el segundo tomo del proyecto (Garriga y Pardo, 2014: 210).

El diccionario, por tanto, queda inacabado debido a razones de diversa índole, entre las que se cuentan las administrativas y de organización, y la falta de medios económicos y de apoyo institucional (Álvarez de Miranda, 2011: 179; Garriga y Pardo, 2014: 228). Aunque la figura de su promotor Torres Quevedo y la de Pelayo Vizuette fueron decisivas para que el proyecto se materializara e iniciara su andadura, su continuidad, y acaso su reorganización, se vio afectada no solo por los motivos mencionados anteriormente, sino por la edad avanzada de sus impulsores y los motivos históricos que acaecieron poco después.

2.3. Selección y organización del léxico

La selección y organización del léxico especializado que contiene el *DTHA* gira en torno a los argumentos ya comentados en los párrafos precedentes y que se pueden leer en las primeras líneas de «La razón del diccionario», primero de los apartados que conforman la «Advertencia» preliminar de esta obra lexicográfica: «Pedían su aparición, por una parte, los adelantos científicos, que exigen modificaciones imprescindibles en la nomenclatura; y, por otra, el aumento del vocabulario y la diversidad de criterios en cuanto a las denominaciones técnicas» (Advertencia, 1930: V). Esta primera motivación se conecta con la evolución de la ciencia y la tecnología y el caudal de voces que vierte en la lengua, y con la necesidad de actualización de sus nomenclaturas. A lo que se añade que «todos los hombres de ciencia españoles e hispanoamericanos han reconocido, unánimemente, la necesidad de esclarecer y fijar la nomenclatura científica, unificándola al mismo tiempo y purgándola, en cuanto sea posible, de voces inadecuadas o

impropias» (Advertencia, 1930: V). Se puede percibir el recelo que esconden estos argumentos respecto a la creencia del descontrol que sufre la lengua española de la ciencia y la técnica al verse invadida por términos extraños a ella, o por formaciones que no sean normativas. Se refuerza esta opinión unas líneas más abajo cuando se menciona el mayor riesgo de todos, el de no llegar a tiempo de intervenir en los usos equivocados, pues «cuando en el uso arraiga un término recién llegado, a todas luces erróneo, o una pronunciación viciosa, es difícilísimo que desarraigue» (Advertencia, 1930: V).

Garriga y Pardo (2014: 231), a partir de una cata del 10% del diccionario, trazan una semblanza del *DTHA* que incluye sus antecedentes lexicográficos, así como su gestación, con todas las incidencias a las que se vio sometido, su planteamiento, la selección temática de voces y acepciones, y las cuestiones de microestructura más relevantes. Este porcentaje estudiado ayuda a los autores a extrapolar las cifras, comparándolas con los 67 000 lemas de la edición de 1925 del *Diccionario de la Real Academia de la Lengua Española*, y a plantear lo que hubiera podido abarcar la obra de haberse continuado:

La proyección de la muestra a la totalidad del *DTHA* proporciona los resultados recogidos en la Tabla 3.

Tabla 3: Proyección de la nomenclatura total del *DTHA*

Entradas 136 000

Subentradas 23 000

Acepciones (suma de las acepciones en entradas y subentradas) **225 000**

[...] Siguiendo la misma proyección, el *DTHA* hubiera alcanzado los 16 volúmenes (64 cuadernos, 9200 páginas aprox.). Garriga y Pardo (2014: 213).

Aunque son aproximadas y hay que tomarlas con prudencia, estas cifras dan cuenta de la envergadura del proyecto y, acaso, pueden reforzar las razones por las que quedó inacabado. Si se quiere profundizar en ellas, Garriga y Pardo (2014: 210 y 224-226) tratan con detalle los motivos que afectaron a la continuidad del *DTHA* y, por otra parte, Gutiérrez Cuadrado (2012: 49-51) ofrece un completo estudio acerca de por qué quedan truncados algunos proyectos lexicográficos, entre ellos el que nos ocupa en este trabajo.

Se puede decir que la trascendencia de este diccionario radica en su original «motivación panhispánica», aunque en el tomo publicado se recojan menos usos americanos de los esperados *a priori*. En una lectura superficial del diccionario no se observan las suficientes variantes americanas como para argumentar ese anunciado carácter unificador de todas las variantes de la lengua española de la ciencia y de la técnica. Se ha sugerido que la dificultad para «conseguir la colaboración de científicos y técnicos latinoamericanos» (Garriga y Pardo, 2014: 215) haya sido la verdadera razón de esa ausencia, hecho que se rastrea, como hemos verificado en el apartado anterior, en la «Advertencia» del propio diccionario. Desde luego, llama la atención que en la «Advertencia» del diccionario no se publique la composición de las juntas nacionales americanas —la de la española está debidamente informada— y solo se comente de una manera bastante genérica que se «cuenta con una activa cooperación de las Juntas Nacionales Americanas» (Advertencia, 1930: XIV). Debemos añadir que se completa el apartado con una descripción generalista de las voces americanas que se aspira a incluir en el diccionario y que abarca desde voces indígenas y voces creadas por naturalistas y viajeros, hasta neologismos, es decir, voces de reciente creación.

Otra cuestión significativa en este diccionario es la que se refiere a que varios de sus artículos se sancionan desde un punto de vista preceptivo. Se señalan como obsoletos (desusado, anticuado), que no atienden a la norma (voces mal escritas, ortografía viciosa, uso incorrecto), que se emplean de manera excesiva, o incluso se distinguen y señalan usos preferidos entre variantes vulgares y científicas de un mismo concepto. Estas recomendaciones están estrechamente vinculadas con la ideología prescriptiva que envuelve a este proyecto de la que ya hemos hablado en apartados anteriores y cuyas pistas se pueden encontrar en el primer apartado de la «Advertencia» del *DTHA*. También son relevantes otros datos que quedan ocultos en las definiciones, que cobran sentido vistos por el ojo experto y que a los profanos les pueden resultar indiferentes o pasar inadvertidos. Nos referimos a los nombres de autoridades citadas, a los nombres científicos, a las fórmulas químicas, físicas o matemáticas, a los puntos geográficos y a las indicaciones georreferenciales. Toda esta información semántica es esencial para el discurso científico y convierte al diccionario de especialidad en un instrumento de representación científica (Moreno Villanueva, 2021: 347) que conecta la lengua con el pensamiento y la sociedad. Las ilustraciones también forman parte de este conjunto de datos relevantes, pues, si bien apenas vehiculan información verbal,

actúan como soporte cognitivo de las entradas a las que hacen referencia y sirven para acercar al lego realidades desconocidas, por lo que adquieren un gran potencial divulgativo. Garriga y Pardo (2014: 233) indican que son las de ciencias naturales las que predominan en el *DTHA* y hacen un cálculo proporcional aproximativo de las que podría haber albergado: «Si se consideran las 323 ilustraciones del 1^{er} tomo, y si se hubiera mantenido la proporción, el *DTHA* hubiera tenido alrededor de 5.000 ilustraciones. Por ese mismo cálculo, el *DTHA* ilustra el 3,96% de las entradas».

Todas las entradas y distintas acepciones del *DTHA* están clasificadas, al menos, por una marca temática que las supedita a las mismas áreas que estas representan. Estas marcas temáticas están recopiladas en la «Advertencia» del diccionario en la sección de abreviaturas. Hemos extraído y contabilizado aquellas que hacen referencia a áreas de conocimiento con un resultado de 94 abreviaturas, aunque Garriga y Pardo (2014: 212) señalen que «entre las marcas de especialidad se incluyen 95 abreviaturas, a las que deben sumarse otras 11 que se encontraron durante la lectura de la obra». Los mismos autores señalan que «el listado de marcas de especialidad indica una perspectiva de la obra claramente técnica y científica si se tiene en cuenta que no se observan abreviaturas para campos de artes y humanidades» (2014: 213). Sin embargo, aunque compartimos la opinión de que el espíritu del diccionario es científico y técnico, consideramos que es inexacto afirmar que no hay abreviaturas que representen las artes y las humanidades. En nuestra observación hemos encontrado las abreviaturas y marcas temáticas de *antropología*, *bellas artes*, *geografía*, *lingüística*, *ortografía*, *prehistoria*, *prosodia* y *sociología*. Es cierto, por otra parte, que la presencia de algunas de las entradas clasificadas con estas marcas es testimonial, pues *ortografía* contiene 0 entradas; *prosodia* contiene 1 entrada; *lingüística* y *sociología* contienen 2 entradas; *bellas artes* contiene 3 entradas; *prehistoria* contiene 10 entradas. Mientras que los artículos y acepciones de *antropología* y *geografía* pasan de las 100 entradas cada uno. Consideramos que, aunque minúscula, la presencia de estas áreas es una demostración del ambicioso plan que se había trazado en la confección del diccionario. Que en la publicación no haya apenas evidencias de algunas áreas se debe al carácter inacabado de la obra, tal y como ya se ha mencionado. Además, tenemos que contradecir también los cálculos que se han hecho. Entendemos que hay un margen de error sobre el 10% estudiado, es decir, sobre las «cinco especialidades mejor representadas» expuestas en la tabla 4 de su trabajo (2014: 213), en

la que señalan que la *milicia* es la quinta materia mejor representada. Sin embargo, y según los cálculos del presente trabajo, la quinta especialidad se corresponde con la *agricultura* (534 casos), seguida de la *marina* (392 casos) y la *paleontología* (383 casos), mientras que la *milicia* (371 casos) ocupa el puesto octavo. Precisamente para que estas inexactitudes no queden consolidadas, el marcado con el estándar XML-TEI que propugnamos en nuestro proyecto de codificación del *DTHA* ayudará a fijar cuantitativamente todas estas cuestiones de descripción numérica relativas al diccionario, además de proporcionar otra lectura más profunda del mismo gracias a la posibilidad de poder destacar los aspectos que se encuentran más ocultos en su microestructura.

No compartimos con Garriga y Pardo (2014: 202) su opinión acerca de que se trata de un proyecto «sobrevalorado por los lexicógrafos» y en estas páginas pretendemos presentar otro proyecto que, gracias a la aplicación de unos estándares informáticos, revalide o no las motivaciones que posibilitaron su arranque, ya comentadas en los párrafos precedentes, la selección del léxico y los usos que se recogen o las fuentes utilizadas entre otros, aunque estos mismos autores declaran que «pese a ser un diccionario inacabado, es indispensable para entender la historia de la terminología del español durante el siglo XX» (2014: 210). Confiamos en que la profundización del estudio que queremos exponer contribuya a ratificar la admiración que, desde siempre, ha suscitado esta obra en los círculos lexicográficos.

3. EL ESTÁNDAR DE CODIFICACIÓN XML-TEI

La *Text Encoding Initiative* (TEI) es una sociedad formada por una comunidad internacional de investigadores cuyo objetivo es el de «*formulate and disseminate guidelines for the encoding and interchange of machine-readable texts intended for literary, linguistic, historical, or other textual research* (TEI Consortium, 1988)». Fundada en el año 1987 en un congreso celebrado en el *Vassar College*, situado en el estado de Nueva York (EE. UU), la historia y el origen de esta sociedad puede consultarse en uno de los apartados de la página del propio consorcio TEI (<https://tei-c.org/about/history/>). Esta organización internacional ha desarrollado un lenguaje estandarizado cuyo fin es el de permitir intercambiar información almacenada de manera electrónica (Burnard y Sperberg-McQueen, 2002; 2012).

El lenguaje de marcación que se ha desarrollado está basado en el metalenguaje XML (*eXtensible Markup Language*), que permite describir y procesar textos y se adapta a las necesidades de los datos que se están procesando (Cummings 2013: 21; Martín Aizpuru 2016: 139), y como añade Fradejas-Rueda (2010: 221), «de todos los lenguajes informáticos que posibilitan el marcado descriptivo el mejor de todos es el XLM», pues «*partire da un file XML possibile produrre diversi tipi di output, da una pagina web a un file PDF*» (Fiormonte, 2006: 2). Las posibilidades tan amplias que ofrece la utilización de este estándar son la causa de la fructífera relación que se produce entre el metalenguaje XML y la organización TEI, mediante su lenguaje de marcado, y construye un sistema estable en el que «uno proporciona las reglas, la sintaxis; la otra el vocabulario, la semántica (Rojas Castro, 2015: 25)». A lo largo de sus más de treinta años de existencia, este estándar ha generalizado su uso y ha sido adoptado por numerosos investigadores en sus proyectos, de tal manera que, hoy en día, iniciar un proyecto de digitalización textual implica comprometerse con las directrices de este estándar si el investigador desea que su proyecto cuente con unos mínimos de interoperabilidad y difusión.

Desde sus primeras aplicaciones hasta la actualidad, existen muchos proyectos construidos bajo el paradigma de las directrices TEI: puede consultarse una lista no exhaustiva en la propia página del consorcio (<https://tei-c.org/activities/projects/>). En esta lista predominan las bibliotecas digitales, los corpus, los manuscritos, los estudios de textos o la correspondencia postal de autores de renombre. Debido al tema tratado en este artículo, no vamos a prestar atención a este tipo de proyecto, pues hemos limitado nuestro estudio a los diccionarios. De esa lista publicada en la página del Consorcio TEI, el porcentaje de proyectos orientados a los repertorios lexicográficos es muy escaso, apenas se registran nueve proyectos relacionados con diccionarios (*African Languages Lexicon Project (ALLEX)*²; *Diccionario electrónico multilingüe de verbos amplios de movimiento (Dicemto)*³; *Digital Dictionary of Buddhism and*

² *African Languages Lexicon Project (ALLEX)*: Proyecto del Léxico de la Lengua Africana (ALLEX) que surgió junto con el corpus estándar de la lengua shona. El corpus se creó como fuente en el desarrollo del *Shona Dictionary*, que también se ha basado en diccionarios como el de *Hannan* y el de *Duramazwi*. Página de consulta: <http://www.edd.uio.no/allex/>

³ *Diccionario electrónico multilingüe de verbos amplios de movimiento (Dicemto)*: diccionario electrónico multilingüe que recrea de forma sistemática y homogénea las diferentes potencialidades semánticas que muestran los verbos de movimiento en varias lenguas como resultado de su combinación en diferentes contextos. Página de consulta: <http://hispanicasuum.es/upstairs/v31/index.htm>. Se puede encontrar información de este proyecto en Batiukova, O., & Miguel, E. d. (2020). Multilingual Electronic

CJKV-E Dictionary⁴; *Digital Dictionary of the German Language*⁵; *Early Irish Glossaries Database*⁶; *FreeDict*⁷; *Icelandic Online Dictionary and Readings*⁸; *Resianica Dictionary 2007*⁹), y la mayoría son proyectos realizados hace ya algunos años.

Ante este panorama se ha realizado una búsqueda que, finalmente, ha resultado ser poco satisfactoria y no exhaustiva, puesto que identificar proyectos de diccionarios codificados en XML-TEI es una tarea más compleja de lo previsto inicialmente y excede los objetivos propuestos en este trabajo. De los resultados obtenidos se podría establecer un patrón en el que se encuentran proyectos digitales que reproducen diccionarios en papel (*Diccionario Nacional Escocés Electrónico (e-SND)*¹⁰; *Du Cange*¹¹; *Petit*

Dictionary of Motion Verbs (DICEMTO): overall structure and the case of andar. In DOMÍNGUEZ VÁZQUEZ, María José, Mónica MIRAZO BALSAL, Carlos VALCÁRCEL RIVEIRO (Eds.), *Studies on Multilingual Lexicography* (Lexicographica. Series Maior ed., pp. 67-92). De Gruyter.

- ⁴ *Digital Dictionary of Buddhism and CJKV-E Dictionary*: el proyecto acoge dos diccionarios: un diccionario/enciclopedia del budismo, con términos en caracteres chinos como palabras de cabecera, y un diccionario/enciclopedia de la historia y la filosofía de Asia Oriental. Página de consulta: <http://buddhism-dict.net/ddb/index.html>.
- ⁵ *Digital Dictionary of the German Language*. En un sistema léxico digital que ofrece acceso a varios tipos de recursos diferentes: un diccionario de la lengua común, un diccionario etimológico y un tesoro, además de acceso a corpus y sus estadísticas. Página de consulta: <https://www.dwds.de/>.
- ⁶ *Early Irish Glossaries Database* es un proyecto para editar los glosarios del irlandés antiguo. Página de consulta: <https://www.asnc.cam.ac.uk/irishglossaries/>.
- ⁷ *FreeDict* ofrece diccionarios bilingües gratuitos para la traducción (140 diccionarios en unas 45 lenguas). Página de consulta: <https://freedict.org/>
- ⁸ *Icelandic Online Dictionary and Readings*. Página de consulta: <https://digidcoll.library.wisc.edu/cgi-bin/IcelOnline/IcelOnline.TEId-idx?id=IcelOnline.IEOrd>.
- ⁹ *Resianica Dictionary*: El diccionario se dirige tanto a los usuarios no profesionales como a los académicos. Su objetivo es proporcionar información ortográfica y morfológica sobre el dialecto esloveno de Resia (Friuli, noreste de Italia). Página de consulta: <http://147.162.119.1:8081/resianica/index.do;jsessionid=590CCFB258AA1E420ACB019F9C85EC24>.
- ¹⁰ *Dictionaries of the Scots Language*: El proyecto del *Diccionario Nacional Escocés Electrónico (e-SND)* tiene como objetivo producir una versión electrónica, publicada en Internet, de la versión en papel del *Diccionario Nacional Escocés (SND)*. Se puede consultar en la siguiente URL: <https://dsl.ac.uk/advanced-search/>. Se puede consultar información de este proyecto en el artículo de RENNIE, Susan. (2000) Encoding a Historical Dictionary with the TEI (With reference to the Electronic Scottish National Dictionary Project) In *Making historical dictionaries with the computer*. Proceedings of EURALEX 2000.
- ¹¹ *Du Cange*. *Du Cange* es un glosario de latín medieval redactado en latín moderno. La obra fue iniciada en el siglo XVII por el Señor de du Cange, pero continuó siendo reeditada y ampliada hasta finales del siglo XIX. <http://ducange.enc.sorbonne.fr/>. Se puede consultar información de este proyecto en el artículo de GLORIEUX KLINCKSIECK, Frédéric (2009) Pourquoi informatiser un vieux glossaire? Présentation du du Cange en ligne, in *Études de linguistique appliquée*, n° 156 (4). pp. 417-428.

*Larousse Illustré de 1905 en ligne*¹²; *Le Littré en ligne*¹³; *Dictionnaire Français-Wolof et Wolof-Français (1855)*¹⁴; *Edição Digital do Vocabulário Ortográfico da Língua Portuguesa (1940)*¹⁵; *Projecto Nenúfar*¹⁶, proyectos de concepción y confección de diccionarios contemporáneos *ad hoc* (*Dictionnaire hydrographique international (DHI)*¹⁷; *Diccionario del Mixteco de Mixtepec*¹⁸; *Diccionario Griego-Español (DGE)*¹⁹; *Diccionario da Língua*

¹² *Petit Larousse illustré de 1905 en ligne*: se trata de uno de los primeros diccionarios ilustrados de la historia de la lexicografía francesa <https://dictionnaire1905.u-cergy.fr/>. El proyecto ya no se actualiza, pero se puede consultar información de este proyecto en el artículo de Hélène Manuélian, Audrey Bruscard, Nicole Cholewka, Anne-Marie Hetzel Klincksieck (2009) *Le petit Larousse illustré de 1905 en ligne: secrets de fabrication et présentation*. In *Études de linguistique appliquée* ° 156 (4). 453-474.

¹³ *Le Littré en ligne*: Se trata de un diccionario publicado al final del siglo XIX cuyo atractivo reside en que está lleno de citas de todo tipo y de todas las épocas, y está adornado con consejos de uso, o reprimendas a autores famosos que se toman libertades con la lengua francesa. <https://www.littre.org/>

¹⁴ *Dictionnaire Français-Wolof et Wolof-Français (1855)*: forma parte de un proyecto más amplio denominado *Missionary linguistics in colonial Africa/ Corpus de travaux linguistiques missionnaires*, al que se puede acceder desde <http://corpus-missionnaire.bc.edu>. Se puede descargar la ODD en <http://tapasproject.org/sites/default/files/1459688094/tei/DictionaryA.xml>.

¹⁵ *Edição Digital do Vocabulário Ortográfico da Língua Portuguesa (1940)*: digitalización y codificación en el primer Diccionario académico de la lengua portuguesa. El proyecto está en curso, pero se puede consultar un adelanto en el sitio web: <https://volp-acl.pt/index.php/vocabulario-1940/projeto>

¹⁶ Proyecto nenúfar: este proyecto tiene como objetivo publicar las primeras ediciones del *Petit Larousse illustré*, con el fin de mejorar el conocimiento y promover el estudio de la historia y la evolución reciente de la lengua, la cultura y las técnicas en francés (<http://nenufar.huma-num.fr/presentation/>). Se puede consultar información de este proyecto en la siguiente publicación: BOHBOT, Hervé, Francesca FRONTINI, Giancarlo LUXARDO, Mohamed KHEMAKHEM, Laurent ROMARY. Presenting the Nénufar Project: a Diachronic Digital Edition of the Petit Larousse Illustré. GLOBALEX 2018 - Globalex workshop at LREC2018, May 2018, Miyazaki, Japan. pp.1-6. fihal-01728328.

¹⁷ *Dictionnaire hydrographique international (DHI)*: El *Diccionario Hidrográfico Internacional (S-32)* es un diccionario multilingüe que constituye la referencia autorizada para los términos y definiciones relacionados con la hidrografía utilizados en las normas, especificaciones y otras publicaciones informativas de la *International Hydrographic Organization (IHO)*. (<http://iho-ohi.net/S32/index.php>). Se puede consultar información de este proyecto en el artículo de ROMARY, Laurent, Patrice BONHOMME, Gilles BESSERO (1998) *L'information du Dictionnaire hydrographique international: normalisation et utilisation*. In, *Terminologie maritime: traduire et communiquer*, May 1998, Bruxelles, Belgique. Recuperado de <https://hal.inria.fr/inria-00487748>

¹⁸ *Diccionario del Mixteco de Mixtepec*: forma parte de un proyecto más amplio que incluye un inventario de rasgos fonéticos y fonológicos del mixteco, transcripciones de todas las expresiones habladas en mixteco, y un diccionario del mixteco. Se puede descargar la ODD en <http://tapasproject.org/sites/default/files/1553517301/tei/MIX-Lexicon-TEI-Dict.xml>

¹⁹ *Diccionario Griego-Español (DGE)* es un diccionario bilingüe de la lengua griega antigua a una lengua moderna. Contiene autoridades, es decir, todas las traducciones y acepciones están documentadas con citas de autores literarios y textos documentales (inscripciones y papiros) que abarcan desde la época micénica y Homero hasta el siglo VI d.C. Página de consulta: <http://dge.cchs.csic.es/>. Forma parte de este mismo proyecto *Léxico de magia y religión en los papiros mágicos griegos (LMPG)*.

Portuguesa Contemporânea (2001)²⁰), y bases de datos lexicográficas con objetivos diversos (*Euskararen Datu-Base Lexikala* (EDBL)²¹; *IMPACT-es diachronic corpus*²²; *Hupa Online Dictionary and Texts*²³).

Algunos de los proyectos son definitivos, otros están en curso y algunos de ellos responden a la iniciativa privada de un investigador. Algunos de estos productos codificados se presentan en un formato de página web, mientras que otros ofrecen su ODD alojada en plataformas digitales diseñadas a tal efecto, y algunos, además, han generado bibliografía académica al respecto. Aunque a primera vista puedan parecer muchas iniciativas lexicográficas comparadas con las aplicaciones de los otros módulos ofrecidos por las directrices TEI, la nómina de proyectos actualizados es más bien escasa. Sin embargo, parece que en los últimos años se ha producido un interés creciente en la digitalización de los diccionarios históricos «*moving past the mere publication of scanned texts to the conversion of paper dictionaries into easily exploitable lexical databases encoded using well established digital standards*» (Bohbot, Frontini, Luxardo, Khemakhem, Romary, 2018: 1), confiamos en que esta tendencia tenga continuidad.

La comunidad TEI recoge todas las posibilidades de marcado textual en un «manual de uso» publicado electrónicamente (<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>) al que denomina «*P5: Guidelines for Electronic Text Encoding and Interchange*», cuya última versión es la 4.4.0., actualizada en abril de 2022. Estas directrices están organizadas en una serie de módulos obligatorios y optativos, cada uno de los cuales está subdividido en

²⁰ *Dicionário da Língua Portuguesa Contemporânea* (2001): en proceso la codificación en XML-TEI de este diccionario. Se puede acceder al *schema* en: https://github.com/DARIAH-ERIC/lexicalresources/blob/master/Events/LexMC2018/Participants/Salgado,%20Ana/Test_ACL_TE_ILex0L.xml.

²¹ *Euskararen Datu-Base Lexikala* (EDBL): La base de datos léxica para el euskera se ha diseñado como un soporte léxico polivalente para el tratamiento automático del euskera. Se puede consultar información de este proyecto en el artículo de Aduriz I., Aldezabal I., Ansa O., Artola X., Díaz de Ilarraza A., Insausti J. M. (1998) EDBL: a Multi-Purposed Lexical Support for the Treatment of Basque. In *First International Conference on language resources & evaluation*: Granada, Spain, 28-30 May 1998 / coord. por Antonio José Rubio Ayuso, Natividad Gallardo San Salvador, Antonio Tejada; María Rosa Castro Prieto (aut.), Vol. 2. Pags., 821-826.

²² *IMPACT-es diachronic corpus*: proyecto europeo cuyo objetivo es el de digitalizar textos históricos. Entre otros, cuenta con un corpus diacrónico del español histórico (*IMPACT-es diachronic corpus*) que «recopila más de un centenar de libros -que contienen aproximadamente 8 millones de palabras- además de un léxico complementario que relaciona más de 10 mil lemas con atestaciones de las distintas variantes encontradas en los documentos». Página de consulta: <https://www.digitisation.eu/resources/language-resources/impact-es/>

²³ *Hupa Online Dictionary and Texts*: es una base de datos conformada por varios diccionarios de la lengua Hupa. Página de consulta: <http://nalc.ucdavis.edu/hupa/hupa-lexicon.php?mode=about>

etiquetas, que, a su vez, pueden enriquecerse con la asignación de un conjunto de atributos y valores. Esta estructura tan elástica, basada en la modularidad y escalabilidad, facilita al investigador elegir lo que más convenga a sus intereses y estructurarlo según su punto de vista particular; es decir, permite construir estructuras versátiles y flexibles, de ahí su popularidad. Además, se puede añadir otro beneficio adicional, como puede ser que no está subordinado comercialmente a ninguna empresa, ya que se trata de un estándar abierto. Esta evidente ventaja se complementa con la posibilidad que ofrece el estándar de generar diferentes formatos de salida de los datos a partir de un único documento XML. Esta interesante condición garantiza compatibilidades e intercambios entre las diversas plataformas y aplicaciones que el investigador tiene a su disposición para publicar sus datos en la actualidad. En resumen, los beneficios que ofrece el lenguaje de marcado semántico TEI (Fradejas 2010; Isasi et al. 2014; Isasi et al. 2020) lo convierten en el formato preferido por cualquier investigador que desee garantizar la permanencia e interoperabilidad de sus datos, y este es el motivo principal por el que lo hemos adoptado en el proyecto que describiremos más adelante.

Codificar un texto con el lenguaje XML-TEI impone al investigador la tarea de prepararlo semánticamente, es decir, hay que «procurar identificar con precisión las funciones o categorías de los textos» (Nicolás Vaughan, 2021: párrafo 13). El marcado semántico del texto nos puede proporcionar nuevo conocimiento acerca de la obra al ir más allá de la mera reproducción del material impreso, pues posibilita aprovechar el contenido textual al máximo (Allés, 2017). Este tipo de marcado cobra un mayor sentido en el caso de la explotación de diccionarios, pues defendemos que la visión diacrónica del diccionario debe proporcionar más información que la mera posibilidad de consultarlo. En efecto, cuando un diccionario pierde actualidad contextual, ya no interesa tanto su consulta para cubrir una laguna cognitiva, sino que lo que le interesa al investigador es poder rastrear el uso de las palabras y ver su recepción en los textos científicos o en cualquier otro tipo de manifestación escrita, sumirse en su macroestructura y microestructura y ver cómo se ha concebido, quiénes han sido sus impulsores, cuáles ha sido las autoridades utilizadas, qué selección de lemas se ha hecho, cómo se ha concebido su redacción, etc. Estas razones son las que imponen al editor digital del diccionario una meditada reflexión acerca del marcado que se va a llevar a cabo en la digitalización, pues de ese cálculo surgirá la riqueza de

posibilidades de consulta, usos diversos y publicación posterior de la obra.

Desde este punto de vista, someter un diccionario de épocas pasadas a la estructuración propuesta por la TEI enriquecerá nuestra percepción acerca de otro tiempo, pues no se codifica solo el propio diccionario como instrumento lexicográfico, sino la propia cosmovisión que aporta el diccionario. Suscribimos las palabras de Forgas Berdet que dicen:

Por eso nos resulta tan fascinante un diccionario antiguo: las palabras en él contenidas devienen referentes históricos, sus definiciones representan parcelas de un mundo pasado y sus ejemplos son pequeños flashes de luz proyectados sobre esa sociedad ya desaparecida» (2001: párrafo 2).

Penetrar en la representación del pensamiento especializado de hace casi un siglo ha sido una de las razones que nos ha empujado a elegir un diccionario que cumpla esas expectativas. El diseño del proyecto de codificación del *DTHA* que se propone en este artículo es el fruto de una labor de reflexión sobre todas estas premisas y pretende iniciar un camino que dé pie a que se pueda aplicar a otras obras lexicográficas. Por ese motivo se propone un modelo para editar diccionarios compuestos en otra época, a la que ahora le faltan referentes actuales, mediante una serie de etiquetas y atributos TEI que proporcionarán la visualización del texto original y de las relaciones profundas que se establecen en el propio diccionario. Como afirmaron hace más de diez años Husson, Viscogliosi, Romary, Descotte, y van Campenhoudt: «*Modern computing technology will result in a real revolution in the approach to dictionary making*» (2000: 11). La revolución ya está hecha, ahora falta que gane adeptos.

4. DISEÑO DEL PROYECTO

Las Humanidades Digitales nos proporcionan el marco adecuado para abordar tanto el análisis de datos como su preservación. Aprovechar esta doble faceta es lo que nos impulsa a emprender la tarea de modelar, codificar y publicar digitalmente el *DTHA*. Los recursos digitales proporcionan los mecanismos que posibilitan «*ver el conocimiento más allá del texto, moviéndonos del dato al texto o a la obra*» (Del Río, 2019: 53). Esta capacidad de las Humanidades Digitales nos proporciona evidencias tangibles donde antes había dudas o intuiciones y permite construir nuevos vínculos de

conocimiento, amplía nuestras percepciones y nos plantea otras preguntas, pues «ya no recibimos pasivamente el conocimiento de un especialista, sino que lo asumimos, a la vez, en un rol activo» (Del Río, 2019: 55).

Por tanto, las posibilidades de aplicación que ofrece la conversión de los diccionarios diacrónicos especializados en un recurso digital los convierte en un instrumento beneficioso para el estudio tanto de la lengua española como del pensamiento científico que transfiere. Estamos totalmente de acuerdo con el apunte de Gutiérrez Rodilla (2016) quien afirma que las obras lexicográficas son necesarias para que el estudio histórico de la lengua de la ciencia sea completo. Sin embargo, no es una tarea fácil realizar el viaje desde el formato papel al formato electrónico, pues no se trata de trasladar el diccionario a la pantalla como si se tratara de una obra analógica, sino que se trata de transformar digitalmente el diccionario en un producto que ofrezca versatilidad de consulta, y por tanto de análisis, que pueda ser interoperable con otros recursos digitales para que se pueda establecer vínculos con otros textos, en definitiva, se trata de convertirlo en un recurso eficaz para el estudio histórico de la lexicografía y de la lengua científica española.

4.1. Preparación del *DTHA* como recurso digital

La conversión digital de cualquier obra concebida en papel nos obliga a adoptar una «mentalidad digital», como propugna Fradejas-Rueda (2021: 229), que está ampliamente comentada en la literatura de las Humanidades Digitales (Thuillier, 2014; Allés, 2019b; Allés, 2020; Fradejas-Rueda, 2021; Fradejas-Rueda, 2022). Partir de esta mentalidad nos impone un flujo de trabajo acorde al soporte y las posibilidades que este ofrece. Uno de los más utilizados y conocidos es el que proponen Rehbein y Fritze (2012: párrafo 15) del que seguiremos sus fases de trabajo para organizar el nuestro. Para ello, utilizamos la traducción que Allés (2020: 84) hace de cada fase:

A grandes rasgos, podemos decir que el proceso ecdótico consiste en cinco grandes etapas: el análisis de la fuente (*analyzing*), la modelización (*modeling*), la transcripción (*transcribing*), la codificación (*encoding*) y la visualización (*visualizing*); a lo largo de las cuales entran en juego otros elementos y técnicas.

La adopción de este flujo de trabajo nos introduce en la edición digital cuyos beneficios hemos pormenorizado en las páginas precedentes.

Fase 1: Análisis de la fuente

Esta primera fase es en la que se selecciona el objeto de estudio, que, en este caso, como ya hemos comentado, es el *Diccionario Tecnológico Hispano Americano*, publicado en 1930. Es imprescindible analizar la fuente con esa mentalidad digital que se ha mencionado, es decir, hay que visualizar las diferentes estructuras en las que está organizado el diccionario, esto es, portada, advertencia (paginada y dividida en secciones, incluye una lista de abreviaturas dividida en dos columnas, otra lista de erratas dividida en seis columnas) y cuerpo del diccionario (texto paginado y dividido en dos columnas y numerosas ilustraciones). En otro nivel, sobre todo orientado al cuerpo del diccionario, también hay que fijarse en los aspectos ortotipográficos (uso de la negrita en los lemas, en las remisiones de artículo, en las palabras poliléxicas, en la numeración de las acepciones; uso de la cursiva en las marcas temáticas, en las etimologías, en los nombres científicos, en los extranjerismos) así como en los cambios de columnas, de página e incluso de línea.

	— 5 —	
<p>Lema</p> <p>Artículo</p> <p>Palabra derivada</p>	<p>ABALIZABLE (de <i>abalizar</i>): adj. <i>Mar.</i> Que puede o debe ser abalizado.</p> <p>ABALIZADIZO, ZA (de <i>abalizar</i>): adj. <i>Ing. Cam.</i> Que se abaliza fácilmente.</p> <p>ABALIZADO, DA (p. p. de <i>abalizar</i>): adj. <i>Ing. Cam. y Mar.</i> 1. Semejante a una baliza. 2. Marcado con balizas.</p> <p>ABALIZADOR, DORA (de <i>abalizar</i>): adj. <i>Ing. Cam.</i> Que abaliza; que se ocupa en poner balizas, o entiende en ello. U. t. c. s.</p>	<p>ABANCALAR (de <i>a</i>, part. de transform. y <i>banca</i>): tr. <i>Mil.</i> Fortificar con líneas o frentes abaluartados. Es procedimiento anticuado.</p> <p>ABALLADO, DA (p. p. de <i>aballar</i>): adj. <i>Arg.</i> Trábandolo de una bodega, lo mismo que <i>cebajada</i>.</p> <p>ABALLAR (etim. dudosa: según la Acad. proviene <i>a</i> y <i>valle</i>, y en tal caso procedería directamente del cat. <i>avallar</i>, bajar, descender, el cual proviene de <i>a</i>, hacia,</p>
		<p>Acepción</p> <p>Definición</p>
	<p style="text-align: center;">FRENTE ABALUARTADO DE UNA FORTALEZA CONSTRUIDA EN LA SEGUNDA MITAD DEL SIGLO XVIII</p> <p style="text-align: center;">Explicación: A-Baluartes BC-Caras CP-Flancos DD-Corneas</p>	<p>Legenda de la ilustración</p> <p>Explicación de la ilustración</p> <p>Ilustración</p> <p>Título de la ilustración</p>
<p>Categoría gramatical</p> <p>Etimología</p> <p>Uso impropio</p> <p>Tipología</p> <p>Definido en otro artículo</p>	<p>ABALIZAMIENTO (de <i>abalizar</i>): m. <i>Ing. Cam. y Mar.</i> Acción o efecto de abalizar.</p> <p>ABALIZANTE (de <i>abalizar</i>): adj. <i>Ing. Cam.</i> Que abaliza.</p> <p>ABALIZAR (de <i>a</i>, que indica provisión o abastecimiento y <i>baliza</i>): tr. <i>Ing. Cam. y Mar.</i> 1. Marcar con boyas o balizas cualquier sitio peligroso de un canal o un puerto. 2. Marcar un objeto sumergido. 3. refl. Marcarse, situarse, arrembarse.</p> <p>ABALORIO (del ár. <i>albalor</i>, vidrio) m. <i>Ing. Ind.</i> 1. Cuentas de vidrio perforadas que se utilizan como decoración. 2. Por ext., cada una de dichas cuentas.</p> <p>ABALGO: m. <i>Agr.</i> Ortografía viciosa, por AVALGO.</p> <p>ABALUARTADO, DA (p. p. de <i>abaluartar</i>): adj. <i>Mil.</i> 1. Se dice del frente de fortificación sujeto a trazado geométrico y determinado por cinco líneas cuya dirección siguen los parapetos y que se denominan <i>caras, flancos y cortinas</i>. Se llama <i>abaluartado</i> porque las caras y flancos de dos frentes contiguos constituyen el ba-</p>	<p>Remisión</p> <p>Autoridad</p> <p>Marca temática</p> <p>Lema flexionado</p>

Ilustración 1: Análisis de la fuente: *Diccionario Tecnológico Hispano Americano* (1930). Fuente: elaboración propia.

Aunque la casuística del diccionario es mayor que la mostrada en la Ilustración 1, sirva esta a modo de ejemplo de lo pormenorizado que debe ser el análisis microestructural del diccionario. Como se puede observar, se pretende codificar cada artículo, cada lema, cada etimología, cada marca, ya sea gramatical o temática, cada acepción, cada definición, y diferenciando entre tipos de lemas, tipos de marcas, tipos de definiciones, o señalando cuestiones que afectan a la ortografía o a la lengua española. Recordemos en este punto las preocupaciones de los redactores del diccionario señaladas en el apartado 2.2., así como otras cuestiones relativas a la variación en los lemas, las autoridades, los lugares o los nombres científicos, tal y como ya hemos señalado en apartados anteriores.

Hay que prestar especial atención a la microestructura del diccionario, ya que el resultado final depende, en realidad, de la granularidad del análisis que el investigador haga. Con la intención de lograr el máximo detalle en el análisis de las entradas del *DTHA*, y después de identificar los diferentes casos correspondientes a la microestructura, se procedió a jerarquizar la posible estructura de sus entradas para comprender los mecanismos lexicográficos de su elaboración que permitieran tomar las decisiones pertinentes en las siguientes fases del flujo de trabajo. Esta tarea es esencial, puesto que a partir de este análisis se debe obtener el modelo abstracto que permita habilitar el trabajo de codificación mediante el metalenguaje XML-TEI. En la ilustración 2 puede observarse el resultado de trasladar esta información desde la observación reflexiva hasta el papel. En ella se puede advertir que en el *DTHA* las entradas pueden estar conformadas por uno o más artículos. A su vez, cada artículo describe su contenido a partir de cinco elementos (lema; origen; categoría gramatical; marca y definición) que distribuyen ordenadamente la información en varias tipologías e, incluso, algunas de estas tipologías son susceptibles de subdividirse en otros niveles que contienen más casos no sistematizados. Estos últimos son opcionales, dependen de las condiciones de cada entrada y obedecen a una gran variabilidad que, en su editado digital, necesitan de otro tipo de tratamiento

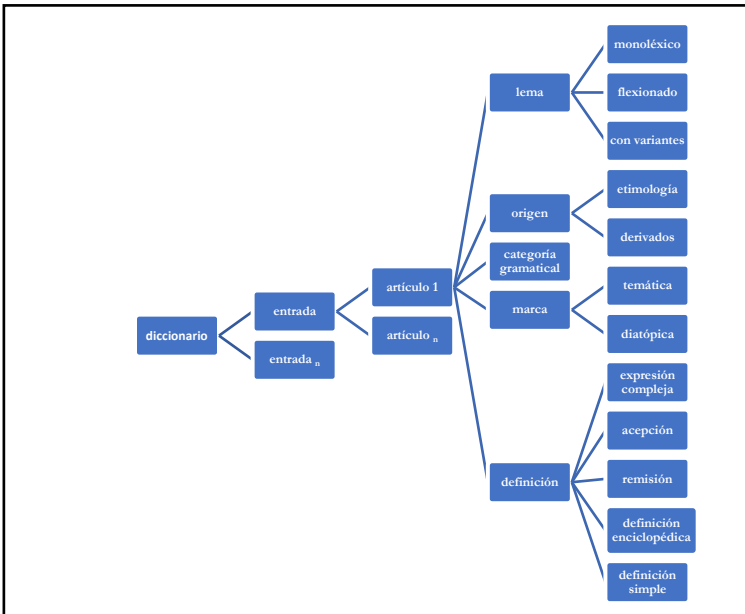


Ilustración 2: Jerarquía de la microestructura del *Diccionario Tecnológico Hispano Americano* (1930). Fuente: elaboración propia.

Queremos subrayar que intentar sistematizar la casuística microestructural no es una tarea sencilla, ya que requiere no solo la lectura profunda y consciente del diccionario, sino que demanda una organización muy detallada con la que se pueda trabajar en la siguiente fase.

Fase 2: Modelización

El segundo paso que hay que dar consiste en abstraer el análisis, realizado en la fase anterior, en un modelo de datos del que realizaremos un esquema (*schema*) de datos, tal y como expone Allés (2020: 85). Se trata de «*modeling the representation of data means analyzing the material with the project's objectives and deriving from both—by applying modeling techniques such as abstraction and generalization—the rules for transforming the material into the digital medium*», Rehbein y Fritze (2012: párrafo 24). Este momento es, quizá, el momento central del flujo de trabajo, pues el modelo de datos es «una especie de guía que nos indi[ca] de qué modo debemos proceder para codificar ese texto, de manera que, si interviene más de una persona, todas lleven a cabo un marcado homogéneo» (Allés, 2019a: L6, párrafo 2).

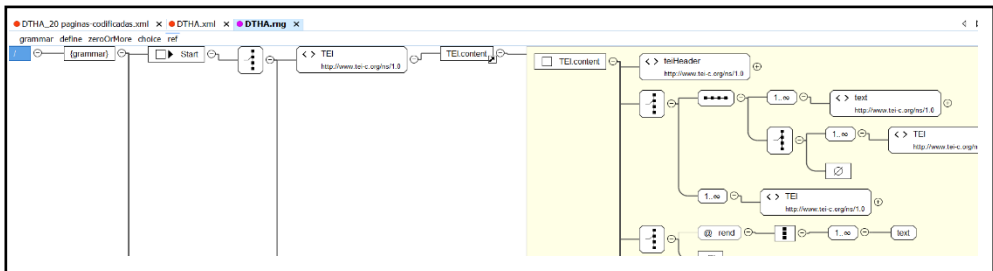


Ilustración 3. Fragmento del esquema .rng del *Diccionario Tecnológico Hispano Americano* (1930). Fuente: elaboración propia.

Este esquema se debe personalizar, adecuándolo según las «Guías directrices para la codificación y el intercambio de textos electrónicos» (<https://tei-c.org/Vault/P5/1.3.0/doc/tei-p5-doc/es/html/index-toc.html>), para que contenga los módulos, elementos y atributos que convengan al proyecto que se lleva a cabo y, finalmente, asociar al documento XML-TEI para que se pueda validar el documento de trabajo (Allés, 2019a: L6). Se ha generado el esquema, cuyo fragmento puede observarse en la ilustración 3, utilizando el estándar normalizado más habitual, es decir Relax NG

(*REgular LAnguage for XML Next Generation*), cuya norma ISO/IEC especifica un patrón para la estructura y el contenido de un documento XML (ISO/IEC 19757-2:2008).

En nuestro caso, y según el análisis previo que se ha realizado, hemos elegido varios de los módulos declarados por las guías directrices del consorcio TEI. Además de los módulos obligatorios de cualquier documento TEI, es decir, *core*, *tei*, *header* y *textstructure*, se han añadido otros tres módulos que responden a las necesidades derivadas del análisis, esto es, *dictionaries* (<https://tei-c.org/Vault/P5/1.3.0/doc/tei-p5-doc/es/html/DI.html>), *figures* (<https://tei-c.org/release/doc/tei-p5-doc/es/html/FT.html>), y *namesdates* (<https://tei-c.org/release/doc/tei-p5-doc/es/html/ND.html>), o lo que es lo mismo, y respectivamente, los módulos 9, 13 y 14 de las guías directrices. Se han elegido estos módulos porque son los más adecuados para el propósito investigador de este proyecto. En cada uno de estos módulos se han seleccionado aquellos elementos, atributos y se han generado los valores indispensables para que el esquema recoja toda la casuística que se quiere codificar.

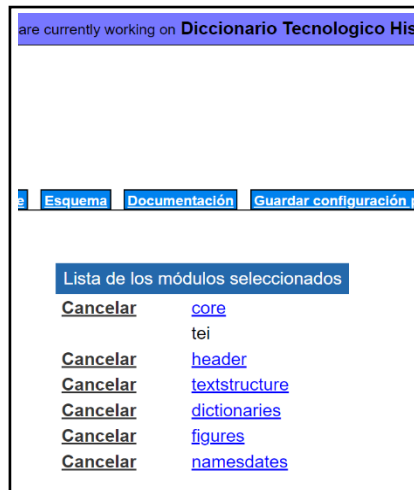


Ilustración 4: módulos seleccionados para el *Diccionario Tecnológico Hispano Americano* (1930). Fuente: elaboración propia mediante la herramienta ROMA.

En la ilustración 4 se muestran los módulos elegidos tal y como los proporciona la herramienta ROMA. El propio consorcio TEI ha elaborado una herramienta que genera el esquema personalizado denominada ROMA, descrita de la siguiente manera en la página

web de esta aplicación²⁴: «*Roma is an ODD Editor, using the TEI ODD (One Document Does-it-all) format for meta-schema documentation and local encoding guidelines as created by the Text Encoding Initiative*» (*Roma - ODD Customization. Version 0.3.1*). Compartimos con Allés (2015: 20) que «cada proyecto debe crear un esquema a su medida, adoptando los módulos que considere convenientes, eliminando y añadiendo elementos y atributos, y concretando valores. Así la labor de codificación será mucho más fácil y consistente».

Fase 3: Transcripción

Rehbein y Fritze (2012: párrafo 26) definen la transcripción como «*the process of transforming the texts carried in the letters as physical objects into machine-readable format, basically as a sequence of characters*». En la redacción de este artículo hemos respetado el orden sugerido por Rehbein y Fritze (2012: párrafo 15). Sin embargo, en la organización de nuestro proyecto hemos optado por empezar por la fase de transcripción. Esto se ha debido a que para considerar viable este proyecto se necesitaba que los datos estuvieran listos para su procesamiento, pues este diccionario se ofrecía bajo un soporte no editable. En este sentido, ha sido crucial para el arranque del proyecto, contar con que nuestra fuente de datos, en este caso el *Diccionario Tecnológico Hispano Americano*, pudiera convertirse mediante un proceso de reconocimiento óptico de caracteres (OCR) de un formato imagen a un formato legible para las máquinas y adecuado para su codificación.

Lamentablemente, el resultado que arrojan los OCR suele requerir una intervención humana que comporta labores de limpieza del texto, así como de comparación con el original. Entre las labores de limpieza acometidas se han eliminado códigos ocultos que se arrastran en la labor de escaneado, se han corregido algunos caracteres que no se han reconocido por el escaneado como palabras en otros alfabetos (griego, árabe), o algunos errores de lectura habituales tales como confundir la *I* por *1*, *i* por *l*, etc. También se ha corregido aquella tipografía que no ha reconocido la negrita que se utiliza para diferenciar acepciones dentro de un lema o para destacar la variante estándar del lema de un artículo, y se ha corregido también el tamaño del lema de los artículos, pues los redactores del *DTHA* consideraron que ciertos artículos debían llevar el lema en un tamaño menor porque son homógrafos del que se considera principal, al que siempre acompaña su etimología. Por

²⁴ En la actualidad, la herramienta ROMA mantiene dos versiones, una con una interfaz tradicional (<https://roma2.tei-c.org/>) y otra con una interfaz más moderna (<https://romabeta.tei-c.org/>). De ambas se obtiene el mismo resultado.

último, se han arreglado todas las erratas encontradas en el texto, pues no se trata de hacer una edición crítica o una edición que analice exclusivamente la tipografía del diccionario, sino que, como ya hemos comentado, se trata de hacer un estudio lexicográfico.

Fase 4: Codificación

Se trata de la fase en la que, gracias al esquema generado en la fase 2 de modelización, se aplican las guías directrices de la TEI al texto objeto de estudio. Este proceso de codificación al que se somete el texto consiste en marcar cada uno de los segmentos que consideramos que tiene un significado para nuestra investigación. Este último punto de vista es interesante, pues enfrentarse a la labor de codificación plantea el reto de tomar decisiones acerca de cómo los creadores del texto han considerado los datos, es decir, nos impone situarnos en una determinada perspectiva desde la cual, y mediante la etiqueta elegida, se va a dotar de valor el segmento que se esté codificando. Como se puede ver, no es un procedimiento sencillo, tampoco se puede decir que sea objetivo, puesto que los elementos, atributos y valores que se eligen en cada caso revelan la visión que tiene el investigador acerca del texto. Adoptamos estas palabras de Allés (2015: 20) en las que manifiesta que «cuando editamos un texto, lo interpretamos, y toda edición, sea cual sea su soporte, es un acto interpretativo susceptible de ser considerado más o menos adecuado según la naturaleza del texto y los fines del proyecto».

No hay que perder de vista que un diccionario ya es un documento estructurado en sí mismo que sigue unos cánones y principios establecidos desde la lexicografía. Incrustar esa estructura en una estructura de codificación no es una tarea tan simple como a priori pueda parecer, pues ambas estructuras no son equivalentes ni mucho menos. La tarea a la que se enfrenta el investigador consiste en diseñar este encaje para que cada una de las partes del diccionario se ajusten en la estructura de marcado que las directrices de la TEI han redactado para los diccionarios (TEI Consortium, 2021b). En su introducción al módulo de Diccionarios se pueden leer estas palabras que suscribimos totalmente:

Both typographically and structurally, print dictionaries are extremely complex. Such lexical resources are moreover of interest to many communities with different and sometimes conflicting goals. As a result, many general problems of text encoding are particularly pronounced here, and more compromises and alternatives within the encoding scheme may be required in the future. (TEI Consortium, 2021a: párrafo 2)

La herramienta más extendida para ejecutar una codificación robusta es una herramienta comercial denominada *<oXygen/>XML Editor*²⁵. Este editor de XML es muy popular debido a las herramientas de creación, edición y publicación de documentos que ofrece. Es muy versátil y está diseñado para ser compatible con los principales sistemas operativos y para multitud de usuarios, desde principiantes a expertos. Se ha hecho uso de esta herramienta para facilitar la labor de codificación del *DTHA*.

```

<superEntry xml:id="A1">
  <entry type="main">
    <form type="lemma">
      <orth> A </orth><pc>.</pc>
    </form>
    <sense>
      <def>Pref. suele emplearse en las ciencias en palabras de origen griego que empiezan con letra consonante.</def>
    </sense>
  </entry>
  <entry xml:id="A2" ana="#A1" type="hom">
    <form type="lemma">
      <orth> A </orth><pc>.</pc>
    </form>
    <sense n="1">
      <usg type="dom" expand="matematicas"> Matem.</usg>
      <def>Se emplea para designar cantidades conocidas o datos supuestos. Es muy general ver empleada esta letra para designar semiejes de curvas o abscisas de sus centros.</def>
    </sense>
    <sense n="2">
      <usg type="dom" expand="astronomia"> Astron.</usg>
      <def>Se emplea para designar la estrella 25 de una constelación, después de haber agotado las letras griegas. </def></sense>
    <sense n="3">
      <usg type="dom" expand="astronomia"> Astron. </usg><def>Abreviatura de Austral.</def>
    </sense>
    <sense n="4">
      <usg type="dom" expand="metrologia"> Metrol. </usg>
      <def>En su forma minúscula, con rasgo final que la circunda, es abreviatura de Arroba.</def>
    </sense>
  </entry>

```

Ilustración 5: Codificación de entradas del *Diccionario Tecnológico Hispano Americano* (1930). Fuente: elaboración propia.

El ejemplo de codificación que se presenta en la ilustración 5 se corresponde con las dos primeras entradas del *DTHA* (ver ilustración 6). Se han elegido estas dos primeras entradas en primer lugar por representar el comienzo de la obra y, en segundo lugar, porque ejemplifican una decisión lexicográfica tomada por los

²⁵ <https://www.oxygenxml.com>

editores, que consiste en que los homógrafos tienen entradas diferenciadas, pero, para mostrar la relación que mantienen, su lema de entrada contiene la particularidad tipográfica de estar escrito en un punto menos de letra. Esta, en apariencia, pequeña cuestión plantea un reto de codificación, ya que hay que relacionar las entradas de alguna manera. Estas entradas homógrafas están relacionadas de una manera semántica, pues comparten la misma etimología de la palabra representada, aunque constituyan bloques diferenciados de significado. Para diferenciar estos dos tipos de artículos consideramos que se tienen que etiquetar de manera diferente. Hemos elegido el elemento <superEntry>²⁶ del módulo 9 para identificar los artículos relacionados con una serie de artículos homógrafos. Para el resto de los artículos relacionados con el artículo principal el elemento que se elige es <entry>²⁷. Todos los artículos homógrafos están anidados dentro de la etiqueta <superEntry> como puede observarse en la ilustración 5.

Se añade también un atributo tipológico para señalar que se trata del artículo principal <entry type="main">. Para completar la identificación del artículo, se utiliza el elemento <form>²⁸ con el atributo @type y el valor ="lemma". Para conectar este artículo con el resto de artículos homógrafos se utilizan otros atributos y valores del elemento <form>, además de utilizar el elemento <entry> y no el <superEntry>. Se ha utilizado también el atributo @ana para relacionar las entradas homógrafas de una <superEntry> (ver ilustración 5).

26 <superEntry> (super entry) agrupa las entradas sucesivas para una serie de homógrafos. [9.1 Dictionary Body and Overall Structure].

27 <entry> (entry) contiene una entrada razonablemente bien estructurada. [9.1 Dictionary Body and Overall Structure 9.2 The Structure of Dictionary Entries]

28 <form> (grupo de información de forma) agrupa toda la información relativa a la forma oral y escrita de una palabra. [9.3.1 Information on Written and Spoken Forms]

A: Pref. inseparable que denota carencia o falta de algo, y que suele emplearse en las ciencias en palabras de origen griego que empiezan con letra consonante.

A: *Matem.* Se emplea para designar cantidades conocidas o datos supuestos. Es muy general ver empleada esta letra para designar semiejes de curvas o abscisas de sus centros. || *Astron.* Se emplea para designar la estrella 25 de una constelación, después de haber agotado las letras griegas. || *Astron.* Abreviatura de **Austral**. || *Metrol.* En su forma minúscula, con rasgo final que la circunda, es abreviatura de **Arroba**.

Ilustración 6: Primeras entradas del *Diccionario Tecnológico Hispano Americano* (1930). Fuente: *Diccionario Tecnológico Hispano Americano* (1930).

Para una mejor comprensión de la decisión tomada, en la ilustración 6 se pueden observar las diferencias tipográficas de las entradas, así como el resto de los apartados de los artículos que han sido codificados con las etiquetas habituales que se utilizarán para el resto de entradas, es decir, <entry>, <form>, <orth>, <sense>, <def> y <usg>. Algunas de ellas, <form>, <sense>, <def>, contienen atributos y valores que permitirán caracterizarlas enriqueciendo el contenido semántico de la codificación (ver ilustración 5).

Fase 5: Visualización

Esta fase supone la última del flujo de trabajo de la edición digital. Llegar a esta fase supone que nuestros datos están listos y pueden ser accesibles para el usuario final. Para facilitar el diseño de esta fase, Allés (2020: 85) propone utilizar papel y lápiz para «idear un modelo de presentación de datos» que sea satisfactorio y adecuado para el planteamiento de nuestro trabajo. Como ya hemos comentado en la fase anterior, la flexibilidad del estándar del consorcio TEI y el lenguaje XML lo hace adecuado «*for encoding different "views" of a dictionary simultaneously in the same document, specifically, a view which sees the dictionary in its textual format, and a view which sees the information in the dictionary without concern for its physical rendering*» (Véronis, Warwick-Armstrong y Calzolari 1992: 9). En efecto, la interoperabilidad del estándar proporciona un resultado susceptible de ser manipulado de manera ágil mediante su conversión a otros estándares lo suficientemente versátiles como es el caso del estándar XSLT (*eXtensible Stylesheet Language Transformations*). Allés (2015: 21) expone de manera muy

comprensible el uso de este estándar, las cuestiones técnicas que se necesitan conocer y las diferentes aplicaciones que se han creado para transformar los archivos de entrada XML en los archivos de salida que demande el proyecto. En nuestro proyecto esta fase está en esbozo y depende de una serie de cuestiones administrativas y técnicas que aún no han sido superadas.

4.2. Cronograma del proyecto *DTHA*

El flujo de trabajo propio de las ediciones digitales que hemos desgranado en el apartado anterior se completa con las etapas de trabajo que hemos diseñado para planificar este proyecto. Hemos dividido el proyecto en tres grandes fases con la intención de someterlo a prueba, desarrollarlo, y en un futuro ampliarlo a otras obras lexicográficas especializadas. A su vez, cada una de esas fases se subdivide con el propósito de establecer objetivos y tareas viables que permitan hacerlo progresar.

FASE I: PROYECTO PILOTO. Si bien un diccionario es un producto altamente estructurado en su origen, se necesita una labor de reflexión previa para someterlo al proceso de edición digital, es decir, se trata de detectar los patrones por los que se rige el diseño del diccionario, pues aunque los elementos esenciales de la información lexicográfica suelen ser más o menos homogéneos (información gramatical, etimología, definiciones, etc.), la diferencia estriba en el diseño de sus micro y macroestructuras (Ide y Véronis, 1995; 1998), debido a los diferentes necesidades de los usuarios a los que se dirija, o a la complejidad y variación de sus entradas (Véronis, Warwick-Armstrong y Calzolari, 1992). En esta primera etapa, y con la finalidad de probar la viabilidad de la propuesta, se ha analizado la fuente al completo, se ha transcrito, y se ha diseñado el modelo abstracto de trabajo. Sin embargo, para la fase de codificación, y debido a su enorme complejidad y a la escasez de los recursos humanos, se ha elegido una muestra del diccionario que ha consistido en codificar las 20 primeras páginas del cuerpo del diccionario, es decir, desde el lema A y hasta el lema ABIGA. En este primer espacio de codificación se han identificado los diferentes tipos de entradas, <entry> y <superEntry>, las marcas gramaticales <gram>, las marcas de uso <usg>, las diferentes acepciones <sense>, y las definiciones <def> (se puede consultar un ejemplo en la ilustración 5). Este paso manual ha permitido identificar varios retos estructurales a los que hay que dar una solución óptima que permita avanzar en las fases siguientes sin tener que enfrentarse a grandes dilemas. Para la fase de visualización

se pretende implementarlo en una página web, pero debido a ciertas cuestiones de tipo administrativo no ha sido posible por el momento. Es la fase en la que aún se encuentra el proyecto.

FASE II: DESARROLLO. Esta fase se centrará en completar la codificación del *DTHA* y en su posterior publicación y puesta a disposición del público. Siguiendo la idea de Manuelian, Bruscan, Cholewka y Hetzel (2014), se ha considerado escalar la codificación en varias etapas con el fin de poder ir publicando los avances que se hagan y, por ese motivo, se articularán diferentes estadios que permitan ir cumpliendo los objetivos propuestos. El primer estadio consistirá en automatizar, mediante lenguajes de programación como R, la identificación y la codificación de los elementos más usuales de los que ya hemos hablado en la fase anterior y en el subapartado 4: codificación del flujo de trabajo (ver apartado 4.1.). Después de una revisión manual para ajustar las pequeñas incongruencias que se puedan producir, se procederá a pasar al siguiente estadio, que consistirá en profundizar en la granularidad de los datos, por ejemplo, el elemento <usg>, además de la información de ámbito especializado, atributos relacionados con información diatópica (americanismo), diacrónica (desusado) o diastrática (vulgar) o usos semánticos, como indicaciones de hiperonimia, por ejemplo. Para abordar toda esta información, y para que la tarea se pueda ir publicando paulatinamente, se ha subdividido el diccionario en fragmentos de 100 entradas aproximadamente cada uno con el fin de facilitar el escalamiento del trabajo.

FASE III: AMPLIACIÓN. Por último, la fase de ampliación pretende aplicar el modelo a otros diccionarios especializados pertenecientes a esta época intersecular. Para ello, y a la finalización del proyecto, está previsto publicarlo en un sitio web en el que se pueda interrogar el diccionario de manera simple (pensando en el gran público) o avanzada (pensando en otros investigadores interesados por la lexicografía o la época) y difundir sus contenidos. El proyecto *Nenúfar*, orientado a la lexicografía francesa (ver nota al pie 16), supone una fuente de inspiración para esta iniciativa, que espera convertirse en una gran colección para la lengua y lexicografía española. Asimismo, la voluntad con la que nace este proyecto tiene que ver con la utilidad que pueda tener el modelo para otros diccionarios del mismo tipo, pues aspiramos a que el modelo de datos se pueda reutilizar en otros proyectos, y con esa finalidad se establecerá un espacio de trabajo en alguna plataforma del tipo GitHub, por ejemplo. Como deseo futuro, y para cumplir con una aspiración que se ha esbozado en las primeras páginas de

este artículo, ambicionamos recopilar el inventario del léxico científico-técnico que alcanza el final de la historia moderna y el inicio de la contemporánea con el objetivo de vincularlo, mediante los textos, a la sociedad que los elabora, los divulga y los consume. Esta pretensión solo puede verse cumplida si se consigue una recopilación completa que pueda ser procesada bajo el paraguas de las Humanidades Digitales y gracias a la interrelación entre lingüística computacional, tecnologías de la web semántica e inteligencia artificial.

5. A MODO DE CONCLUSIÓN

En las páginas precedentes se ha mostrado una iniciativa ambiciosa que está en curso y que aspira a seguir en crecimiento completando las fases que se han proyectado. La muestra que se ha expuesto evidencia que la propuesta es viable, pues se considera que el modelo puede ser útil a la comunidad científica, ya que podrá aplicarse a otros trabajos similares. No ocultamos que al proyecto le queda un largo camino por recorrer y que necesita sumar seguidores que lo ayuden a prosperar, pero mantenemos que la historia de la lexicografía y la de la lengua españolas necesitan profundizar su estudio ampliándolo a la esfera especializada.

Los diccionarios elaborados en otras épocas han perdido su carácter primigenio, esto es, el de servir como recurso de consulta y, en ese sentido, la pérdida de su función artefactual, que los condena al olvido, puede revertirse gracias a su codificación, pues es un punto de partida para preservarlos y para darles un uso renovado como puerta de entrada a la conexión con el conocimiento de la época a la que representan. En este caso concreto, el periodo seleccionado tiene mucha importancia para la historia de la lexicografía y lengua españolas, ya que proporciona las bases sobre las que se asienta buena parte del conocimiento especializado del siglo XX y, a la vez, sirve como precedente del siglo XXI. En ese sentido, la edición digital del *DTHA* permitirá establecer un punto de partida desde el que articular una amplia red de conexiones entre los textos y los términos, la producción científica y las personalidades más importantes, los redactores del diccionario y las autoridades científicas citadas, pues todos estos lazos pueden servir como fuente y conexión de saberes y conocimientos. En resumen, las tareas que se propone este proyecto pretenden identificar la relación que existe entre el pensamiento, la lengua y la sociedad españolas de un espacio y

tiempo determinados, vehiculada mediante las recopilaciones léxicas.

REFERENCIAS BIBLIOGRÁFICAS

- ADURIZ I., ALDEZABAL I., ANSA O., ARTOLA X., DÍAZ DE ILARRAZA A. y INSAUSTI J. M. (1998): EDBL: «A Multi-Purposed Lexical Support for the Treatment of Basque». In *First International Conference on language resources & evaluation*: Granada, Spain, 28-30 May 1998 / coord. por Antonio José Rubio Ayuso, Natividad Gallardo San Salvador, Antonio Tejada; María Rosa Castro Prieto (eds.), Vol. 2. Pags., 821-826
- AHUMADA LARA, I. (2007): *La metalexicografía del español: revisión y perspectivas*. Polimetrica International Scientific Publisher. Recuperado de <https://explore.openaire.eu/search/other?orpId=od_____4094:d804b0058627ecdcca9315d0908d32a0> [Consulta: 02/08/2021]
- (2016): *Metalexicografía del español: Clasificación orgánica y tipología de los diccionarios en el Diccionario bibliográfico de la metalexicografía del español (DBME)*. *Annuario de estudios filológicos*, 39, 5-24. Recuperado de <<https://dehesa.unex.es:8443/handle/10662/6955>> [Consulta: 02/08/2021]
- ALLÉS TORRENT, S. (2015): «Edición digital y algunas tecnologías aliadas». *Insula*, 18-21. Recuperado de <<https://academiccommons.columbia.edu/doi/10.7916/D89S1QFN>> [Consulta: 23/08/2021]
- (2017): «Notas para una edición digital de Alfonso de Palencia». *Atalaya (Paris, France)*, 17(17) Recuperado de <<https://journals.openedition.org/atalaya/2462>> [Consulta: 11/08/2021] <https://doi.org/10.4000/atalaya.2462>
- (2019a): *Esquemas RNG y personalización de TEI*. Zenodo. Recuperado de <<https://doi.org/10.5281/zenodo.4555346>> [Consulta: 23/07/2022]
- (2019b): «Sobre la complejidad de los datos en Humanidades, o cómo traducir las ideas a datos». *Revista Humanidades Digitales*, 4, 1-28. Recuperado de <<https://revistas.uned.es/index.php/RHD/article/view/24679>> [Consulta: 19/08/2021] <https://doi.org/10.5944/rhd.vol.4.2019.24679>

- (2020): «Crítica textual y edición digital o ¿dónde está la crítica en las ediciones digitales?» *Studia aurea: revista de literatura española y teoría literaria del Renacimiento y Siglo de Oro*, 63-98. Recuperado de <<https://studiaaurea.com/article/view/v14-alles>> [Consulta: 19/08/2021] <https://doi.org/10.5565/rev/studiaaurea.395>
- ALVAR EZQUERRA, M. (1993): *Lexicografía descriptiva*. Bibliograf.
- ÁLVAREZ DE MIRANDA, P. (2011): «Los diccionarios históricos», en *Los diccionarios del español moderno*. Somonte-Cenero, Gijón, Asturias: Trea. pp. 119-140.
- AZORÍN FERNÁNDEZ, M. D. (1990): «El Diccionario General de la Lengua frente a los vocabularios científicos y técnicos». Comunicación presentada en *Proceedings of the 4th EURALEX International Congress*, 445-453.
- (2001): *Los diccionarios del español en su perspectiva histórica*. San Vicente del Raspeig: Publicaciones de la Universidad de Alicante.
- BATTANER, P. (1996): «Terminología y diccionarios» en CABRÉ, M. T. (ed.). *Jornada Panllatina de Terminologia: perspectives i camps d'aplicació*, IULA (UPF).
- BATIUKOVA, O. y DE MIGUEL, E. (2020): «Multilingual Electronic Dictionary of Motion Verbs (DICEMTO): overall structure and the case of andar». In DOMÍNGUEZ VÁZQUEZ, M. J. et alii (eds.), *Studies on Multilingual Lexicography* (Lexicographica. Series Maior ed., pp. 67-92). De Gruyter.
- BOHBOT, H., FRONTINI F., LUXARDO, G., KHEMAKHEM, M. y ROMARY, L. (2018): «Presenting the Nénufar project: a diachronic digital edition of the Petit Larousse Illustré». *Globalex 2018 - Globalex Workshop* at LREC2018, may 2018, Miyazaki, Japan. pp.1-6. fihal-01728328.
- CASTRO-PRIEETO, M. R. (2018): «Producción del discurso de divulgación científica en lengua española durante el primer tercio del siglo XX contexto e ideología». In ÁLVAREZ PÉREZ, X. A. et alii (eds.), *Nuevas perspectivas en la diacronía de las lenguas de especialidad* (1ª ed., pp. 33-48). Alcalá de Henares: Universidad de Alcalá. Servicio de Publicaciones.
- CUMMINGS, J. (2013): «The text encoding initiative and the study of literature». In *A companion to digital literary studies*, 451-476. Recuperado de <https://companions.digitalhumanities.org/DLS/?chapter=content/9781405148641_chapter_25.html> [Consulta: 7/7/2022]

- DEL RÍO RIANDE, G. (2019): «La mirada humana, la mirada crítica». *Telos. HUMANIDADES EN UN MUNDO STEM*, (112), 50-55. Recuperado de <<https://telos.fundaciontelefonica.com/revista/telos-112/>> [Consulta: 30/07/2021]
- Diccionario Tecnológico Hispano Americano* (1930): Editorial Arte y Ciencia. Madrid.
- DUBOIS, J. y DUBOIS, C. (1971): *Introduction à la lexicographie*. Larousse.
- FIORMONTE, D., GERI L. y SILVI D. (2006): «XML e i testi contemporanei: problemi di codifica, rappresentazione e analisi». Comunicación presentada a *Clip 2006*, 1-22. Recuperado de <https://clip2006.cch.kcl.ac.uk/clip2006/redist/pub/Testo_3.pdf> [Consulta: 30/07/2021]
- FORGAS BERDET, E. (2001): «Diccionario e ideología: Tres décadas de la sociedad española a través de los ejemplos lexicográficos». *Espéculo (Madrid, Spain)*, 17(17) Recuperado de <<https://webs.ucm.es/info/especulo/numero17/3decadas.html>> [Consulta: 3/03/2022]
- (2006): «Diccionarios e ideología». *Interlingüística (Alacant, Spain)*, (17), 2-16.
- FRADEJAS RUEDA, J. M. (2010): «La codificación XML/TEI de textos medievales». *Memorabilia*, (12), 219-247. Recuperado de <<https://parnaseo.uv.es/Memorabilia/Memorabilia12/PDFs/Codificacion.pdf>> [Consulta: 23/05/2022]
- (2021): «Las Siete Partidas: del pergamino a la red». In Mechthild, A., Becker, U. y Schmidt E. (eds.), *Alfonso el Sabio y la conceptualización jurídica de la monarquía en las 'Siete Partidas'* (pp. 223-264). Gottingen: Vandenhoeck & Ruprecht. Recuperado de <<https://uvadoc.uva.es/handle/10324/53399>> [Consulta: 20/05/2022]
- (2022): «De editor analógico a editor digital». *Historias Fingidas, Numero Speciale 1: Digital Humanities e studi letterari ispanici*, 39-65. Recuperado de <<https://historiasfingidas.dlss.univr.it/article/view/1108>> [Consulta: 25/05/2022]
- GARRIGA ESCRIBANO, C. (2015): «Historia del léxico y lexicografía especializada: El Diccionario industrial: Artes y oficios de Europa y América (1888-1891) de Camps y Armet como fuente». *Etudes romanes de Brno*, 36(1), 61-84. Recuperado de <http://gateway.proquest.com/openurl?ctx_ver=Z39.88-2003&xri:pqil:res_ver=0.2&res_id=xri:ilcs-

- us&rft_id=xri:ilcs:rec:abell:R05447901> [Consulta: 23/05/2022]
- GARRIGA ESCRIBANO, C., y PARDO HERRERO, P. (2014): «El Diccionario Tecnológico Hispano-Americano, un nuevo intento en la institucionalización de la lengua de la ciencia y de la técnica en español». *International journal of lexicography*, 27(3), 201-240. <https://doi.org/10.1093/ijl/ecu006> [Consulta: 23/05/2022]
- GLORIEUX KLINCKSIECK, F. (2009): «Pourquoi informatiser un vieux glossaire? Présentation du du Cange en ligne», in *Études de linguistique appliquée*, n° 156 (4). pp. 417-428.
- GONZÁLEZ DE POSADA, F. (1992): *Leonardo Torres Quevedo* (5th ed.). Fundación del Banco Exterior de España.
- GUTIÉRREZ CUADRADO, J. (1978): «Sobre la historiografía de la lingüística española del siglo XX». *Llull: Revista de la Sociedad Española de Historia de las Ciencias y de las Técnicas*, (2), 66-80.
- (1989): «La lengua y las relaciones hispanoamericanas alrededor de 1900: ideología y trabajo lingüístico». In *Ciencia, vida y espacio en Iberoamérica: trabajos del Programa Movilizador del CSIC. «Relaciones científicas y culturales entre España y América»*. Consejo Superior de Investigaciones Científicas, CSIC.
- (2012): «Los diccionarios inacabados», en *Avances de lexicografía hispánica*, edición de Nomdedeu Rull, A. et alii. Tarragona: Publicacions URV.
- GUTIÉRREZ CUADRADO, J. y GARRIGA ESCRIBANO, C. (2019): «El vocabulario científico y técnico del español entre los siglos XIX y XX: planteamientos generales». *Revista De Lexicografía*, 25, pp. 193-218.
- GUTIÉRREZ RODILLA, B. (2003): «La historia del lenguaje científico como parte de la historia de la ciencia». *Asclepio*, 55(2), 7-26. [10.3989/asclepio.2003.v55.i2.101](https://doi.org/10.3989/asclepio.2003.v55.i2.101) [Consulta: 7/03/2022]
- (2016): «Reflexiones historiográficas sobre el léxico científico y los repertorios lexicográficos». In GARRIGA ESCRIBANO, C. y PÉREZ PASCUAL, J. I. (ed.), *Lengua de la ciencia e historiografía* (pp. 117-128). Universidade da Coruña, Servizo de Publicacións.
- (2017): «La preocupación por la lengua y su reflejo en la lexicografía: el caso de los vocabularios españoles de medicina en el siglo XIX y principios del XX». *Moenia: Revista Lucense De Lingüística & Literatura*, (23) Recuperado de <<https://revistas.usc.gal/index.php/moenia/article/view/3873>> [Consulta: 7/03/2022]
- HUSSON, J., VISCOGLIOSI, N., ROMARY, L., DESCOTTE, S. y VAN CAMPENHOUDT, M. (2000): «Specialized lexicography by

- means of a conceptual data base: establishing the format for a multilingual marine dictionary». Paper presented at the *Second Conference on Maritime Terminology 2000*, 1-18. Recuperado de <<https://hal.inria.fr/inria-00099351>> [Consulta: 7/07/2022]
- IDE, N. y VÉRONIS, J. (1995): «Encoding Dictionaries». *Computers and the Humanities*, 29(1-3), 167-179. Recuperado de <<https://www.cs.vassar.edu/~ide/papers/Chumdict.pdf>> [Consulta: 7/07/2021]
- IDE, N. y VÉRONIS, J. (1998): «A feature-based model for lexical databases». Paper presented at the *Actas de COLING-92*, 588-594. Recuperado de <<https://aclanthology.org/C92-2089.pdf>> [Consulta: 14/07/2021]
- ISASI MARTÍNEZ, C., MARTÍN AIZPURU, L., PÉREZ ISASI, S., PIERAZZO, E. y SPENCE, P. (2020): *Edición digital de documentos antiguos: marcación XML-TEI basada en los criterios CHARTA*. Editorial Universidad de Sevilla (Colección Lingüística, n.º 58). DOI: <https://dx.doi.org/10.12795/9788447222049>.
- LAFUENTE, A. y SARAIVA, T. (2002): *Los públicos de la ciencia en España, siglos XVIII a XX*. Madrid: FECYT.
- LE PESANT, D. (2020): «Élaboration d'un nouveau modèle de la langue et lexicographie», *Langue(s) & Parole*, Núm. 5, 23-42. Recuperado de <https://ddd.uab.cat/record/235891> [Consulta: 14/07/2021]
- MAINER, J. C. (1981): *La Edad de Plata de la Cultura española (1900-1931)*. *Ensayo de interpretación de un proceso cultural*. Cátedra.
- MANUÉLIAN, H., BRUSCAND, A., CHOLEWKA, N. y HETZEL KLINCKSIECK, A-M. (2009): «Le Petit Larousse illustré de 1905 en ligne: secrets de fabrication et présentation». In *Études de linguistique appliquée* 156 (4). 453-474.
- MARTÍN AIZPURU, L. (2016): «Algunos recursos informáticos al servicio de la edición de textos: la edición en XML-TEI». In Albertin y del Rey Quesada (Ed.), *Hispanica Patavina. Estudios de historiografía e historia de la lengua española en homenaje a José Luis Rivarola* (pp. 139-154.). CLEUP.
- MORENO VILLANUEVA, J. A. (2021): «Diccionarios científicos y técnicos del español contemporáneo (1875-1936): estudio y catalogación». In MORENO MORENO, Á. y TORRES MARTÍNEZ, M. (eds.), *Líneas de investigación del Seminario de Lexicografía Hispánica Investigaciones léxicas. Estados, temas y rudimentos* (pp. 334-344) Recuperado de <<https://dialnet.unirioja.es/servlet/articulo?codigo=8187810>> [Consulta: 13/07/2021]

- OLAGÜE DE ROS, G. (2006): «La colaboración hispano-latinoamericana en los orígenes y desarrollo del movimiento documental europeo contemporáneo (1900-1920)». *Encontros Bibli: Revista Eletrônica De Biblioteconomia E Ciência Da Informação*, (2), 82-95. 10.5007/1518-2924.2006v11nesp2p82 [Consulta: 13/07/2021]
- ORTIZ, E. L. (1988): «Las relaciones científicas entre Argentina y España a principios de este siglo. La Junta para la ampliación de estudios y la Institución cultural española». In J. M. Sánchez Ron (Ed.), *La Junta para Ampliación de Estudios e Investigaciones Científicas 80 años después, 1907-1987: Simposio internacional, Madrid, 15-17 de diciembre de 1987*. Consejo Superior de Investigaciones Científicas, CSIC.
- OTERO CARVAJAL, L. E. (2001): «La destrucción de la ciencia en España. Las consecuencias del triunfo militar de la España franquista». *Historia Y Comunicación Social*, (6), 149-186. Recuperado de <https://doaj.org/article/79d4c82fc7494ae1a28624310765cf86> [Consulta: 22/07/2021]
- PESET, J. L. (1989): *Ciencia, vida y espacio en Iberoamérica. trabajos del Programa Movilizador del C.S.I.C. "Relaciones científicas y culturales entre España y América"*. Consejo Superior de Investigaciones Científicas.
- RAE. (s.f.): Presentación del «Diccionario de la lengua española» y sus ediciones. Real Academia de la Lengua Española. Recuperado el 22 de julio de 2022, <https://www.rae.es/obras-academicas/diccionarios/presentacion-del-diccionario-de-la-lengua-espanola-y-sus-ediciones> [Consulta: 5/07/2022]
- REHBEIN, M., y FRITZE, C. (2012): «Hands-On Teaching Digital Humanities: A Didactic Analysis of a Summer School Course on Digital Editing». In Brett D. Hirsch (Ed.) *Digital Humanities Pedagogy*. Recuperado de <https://books.openedition.org/obp/1617?lang=es> [Consulta: 22/07/2022]
- RENNIE, S. (2000): «Encoding a Historical Dictionary with the TEI (With reference to the Electronic Scottish National Dictionary Project)» In *Making historical dictionaries with the computer*. Proceedings of EURALEX 2000.
- RODRÍGUEZ BARCIA, S. (2002): «Sociedad y diccionario: Notas para un acercamiento a la lexicografía española de la primera mitad del siglo XX». *Res diachronicae*, 1, 301-315. Recuperado de https://resdiachronicae.files.wordpress.com/2013/12/volumen-1-30_rodriguez_barcia.pdf [Consulta: 02/08/2021]

- ROJAS CASTRO, A. (2015): *Editar las Soledades de Luis de Góngora en la era digital: Texto crítico y propuesta de codificación XML/TEI*. Recuperado de < <http://hdl.handle.net/10803/323083> > [Consulta: 10/03/2022]
- ROMARY, L., BONHOMME, P. y BESSERO, G. (1998): «L'informatisation du Dictionnaire hydrographique international: normalisation et utilisation». In, *Terminologie maritime: traduire et communiquer*, May 1998, Bruxelles, Belgique. Recuperado de <<https://hal.inria.fr/inria-00487748>> [Consulta: 19/08/2021]
- SÁNCHEZ RON, J. M. (2020): *El país de los sueños perdidos* (Primera edición ed.). Taurus.
- TEI CONSORTIUM. (1990): *Design Principles for Text Encoding Guidelines: TEI ED P1*. Recuperado de <http://projects.ox.ac.uk/teiveb/Vault/ED/edp01.xml?ID=boddy.1_div1.3> [Consulta: 19/07/2022]
- (2021a). «Dictionaries,» in: *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.3.0. Last updated on 19st April 2022, revision ff9cc28b0 Recuperado de <<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html>> [Consulta: 19/07/2022]
- (2021b). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.3.0. Last updated on 19st April 2022, revision ff9cc28b0 Recuperado de <<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>> [Consulta: 19/07/2022]
- THUILLIER, S. (2014): «De una obra digitalizada a una obra digital: reflexiones en torno al Diccionario Griego-Español en línea», en *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, en LÓPEZ POZA, S. y PENA SUEIRO, N. (eds), *Janus*, 1, 459-471, publicado el 11/04/2014. Recuperado de URL: <https://www.janusdigital.es/anexos/contribucion.htm?id=40> [Consulta: 26/07/2022]
- UNIÓN INTERNACIONAL DE BIBLIOGRAFÍA Y TECNOLOGÍA CIENTÍFICAS. Discursos leídos ante S.M. el Rey en la solemne sesión celebrada por dicha Unión Internacional en la Real Academia Española, con motivo de la publicación del primer cuaderno del Diccionario Tecnológico Hispano-Americano. Madrid: Imprenta Clásica Española, 1926.
- VAUGHAN, N. (2021): «Introducción a la codificación de textos en TEI (parte 1)». *Programming Historian En Español*, 5 Recuperado de <https://doi.org/10.46430/phes0053> [19/08/2021]

VÉRONIS, J., WARWICK-ARMSTRONG, S., y CALZOLARI, N. (1992): «Principles for encoding machine readable dictionaries». Paper presented at the *Fifth Euralex International Congress*, 239-246. Recuperado de <<https://euralex.org/publications/principles-for-encoding-machine-readable-dictionaries/>> [Consulta: 29/07/2021]

