

Recursos tecnológicos y digitales para la gestión del lenguaje científico en español[☆]

Ricardo Mairal-Usón^a y Pedro A. Fuertes-Olivera^b

^a *Catedrático de Lengua y Lingüística Inglesa, Departamento de Filologías Extranjeras de la UNED
Vicerrector de Profesorado y Vicerrector Primero de la UNED
Presidente de la Asociación Española de Lingüística Aplicada*

^b *Catedrático de Filología Inglesa, Departamento de Filología Inglesa, Facultad de Comercio, Universidad de Valladolid, Valladolid, España
Profesor Extraordinario, Department of Afrikaans and Dutch, University of Stellenbosh, Sudáfrica*

PALABRAS CLAVE

Tecnologías del lenguaje
Digitalización de contenidos
Lexicografía del Nuevo Milenio
Extractores terminológicos
Web semántica
Datos enlazados
Ontologías
Búsquedas conceptuales

Resumen En este artículo se hace un repaso de algunas de las tecnologías del lenguaje existentes en la actualidad, especialmente de las que están más relacionadas (o pueden estarlo) con el lenguaje científico en español. Además de señalar la relevancia de algunas tecnologías, por ejemplo las que hacen posible la existencia de recursos enlazados, nos centramos en los dos principales modelos teóricos y prácticos relacionados con la lexicografía de internet, especialmente la que ofrece una panorámica completa, ya que no solo reflexiona sobre la naturaleza de esta y las características principales de sus productos —es decir, de los diccionarios en línea o diccionarios de internet— sino que también ofrece innovaciones y modelos de negocio que explican la conexión clara y diáfana entre la gestión del lenguaje científico en español y la economía basada en el conocimiento.

© 2016 Elsevier España, S.L.U. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

KEYWORDS

Language technologies
Contents digitization
New Millennium
Lexicography
Terminology
extractors

Technological and digital resources for management of scientific language in Spanish

Abstract This article is a review of some of the existing language technologies today, especially those that are more related to (or can be) with scientific language in Spanish. In addition to point out the relevance of some technologies, for example those that make possible the existence of linked resources, we focus on two main models and skills related to the lexicography of internet, especially one that offers a complete panoramic view since it not only reflects on

Correo electrónico: rmairal@flog.uned.es; pedro@emp.uva.es

[☆] Este trabajo forma parte del trabajo realizado en el marco del proyecto financiado por el Ministerio de Economía y Competitividad FFI2011-29798-C02-01, FFI2014-53788-C3-1-P y FFI2014-52462-P.

Semantic web
Linked data
Ontologies
Concept searches

the nature of it and the main features of its products –i.e., dictionaries online or dictionaries of internet– but also offers innovations and business models that explain the clear and transparent connection between the management of the scientific language in Spanish and knowledge-based economy.

© 2016 Elsevier España, S.L.U. This is an open access item distributed under the Creative Commons CC License BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

Es un lugar común atribuir la importancia del español al número de sus hablantes, cercano hoy en día a los 500 millones de personas, repartidos en 21 países en los que esta lengua es el idioma oficial. Estas cifras sitúan al español como la segunda lengua del mundo, por encima del inglés, en cuanto a su número de hablantes nativos, quedando tan solo por debajo del chino mandarín. Y aunque es cierto que el inglés supera en estos momentos al español en influencia por su papel como lengua franca, el español es la tercera lengua más empleada en medios de comunicación y en internet, con un 7,6% del total. A esto se suma el vertiginoso aumento de su uso científico en los últimos 15 años, con un crecimiento superior al 130% del número de revistas científicas publicadas exclusivamente en español¹.

No obstante la innegable importancia de estos datos para situar al español en un puesto de prominencia mundial, no se debe soslayar el lugar del español como lengua de transmisión literaria, con comienzos ya notables en la poesía popular juglaresca de principios del siglo XII. La evolución posterior, de reconocido impacto en la literatura mundial, requiere poco comentario adicional: las composiciones del Mester de Clerecía y los trabajos de la Escuela de Traductores de Toledo, en el siglo XIII, la poesía de Garcilaso de la Vega, Fernando de Herrera y Fray Luis de León, la ascética y mística de Fray Luis de Granada, Santa Teresa de Jesús, San Juan de la Cruz y San Ignacio de Loyola, del Renacimiento en el siglo XVI; el impresionante Siglo de Oro, con insignes figuras cuyas obras se estudian una y otra vez a nivel internacional, como Miguel de Cervantes, Francisco de Quevedo, Luis de Góngora, Lope de Vega, Tirso de Molina y Calderón de la Barca; la literatura de la Ilustración, con Fray Benito Jerónimo Feijoo, Gaspar Melchor de Jovellanos y José Cadalso, en prosa, Juan Meléndez Valdés, Iriarte y Samaniego, en lírica, y Leandro Fernández de Moratín y Ramón de la Cruz, en teatro; las grandes figuras del romanticismo, en poesía, con José de Espronceda, primero, y posteriormente Gustavo Adolfo Bécquer, Augusto Ferrán y Rosalía de Castro, a lo que se suma la obra de José de Zorrilla y Bretón de los Herreros en teatro; los escritores del realismo, como Galdós, Clarín, Pardo Bazán y Blasco Ibáñez, y finalmente, el modernismo, la generación del 98, la del 27 y la novela de posguerra. Cuenta la literatura española con 5 premios Nobel españoles (José Echegaray, Jacinto

Benavente, Juan Ramón Jiménez, Vicente Aleixandre y Camilo José Cela) y otros varios de distintas naciones de habla hispana: Gabriela Mistral (Chile), Miguel Ángel Asturias (Guatemala), Pablo Neruda (Chile), Gabriel García Márquez (Colombia), Octavio Paz (México) y Mario Vargas Llosa (Perú).

Con semejante esplendor literario, fecundo y de largo recorrido, con tan amplia implantación a lo largo y ancho de la Tierra, con una gran tradición de desarrollo histórico, no debe extrañar que el español atraiga, de forma creciente, no solo el interés de estudiosos y eruditos, sino también el de quienes quieren disfrutar de una lengua de cultura y ciencia y de entendimiento mundial.

Por otro lado, estamos inmersos en la era de la digitalización, del auge y presencia ya consolidada de las nuevas tecnologías, que están cambiando los sistemas de comunicación (p. ej., la irrupción de las redes sociales, especialmente de twitter, como medio de comunicación instantánea de las principales noticias), que nos permiten acceder a una cantidad de información inimaginable en un tiempo menor, que están introduciendo una nueva cultura social (p. ej., el comercio electrónico, la administración electrónica, etc.). Así, en esta línea, no nos sorprende el lugar central que ocupan en la agenda digital europea los proyectos vinculados a la digitalización de los contenidos, los datos enlazados, la mejora de los sistemas de acceso y recuperación de la información, etc. (p. ej., la Digital Agenda for Europe (DAE), <https://ec.europa.eu/digital-agenda/digital-agenda-europe>, o el Plan de Impulso de las Tecnologías del Lenguaje promovido por el Ministerio de Industria, <http://www.agendadigital.gob.es/planes-actuaciones/Paginas/plan-impulso-tecnologias-lenguaje.aspx>).

Pues bien, en este contexto, donde la lengua española ocupa un lugar tan relevante en el mundo, junto a la necesidad incuestionable de impulsar una agenda digital competitiva, el objetivo de este trabajo es plantear de forma sumaria de qué forma las tecnologías del lenguaje pueden contribuir al desarrollo de recursos lingüísticos innovadores y competitivos, con algoritmos y técnicas rupturistas que saquen el máximo partido a los grandes avances y desarrollos que se han logrado en campos como la inteligencia artificial, el aprendizaje automático (*machine learning* o el *deep learning*) o el procesamiento del lenguaje natural y, en general, la interacción hombre-máquina, con el fin de que el creciente número de hispanohablantes, estudiosos e investigadores de la lengua puedan disponer de un elenco de herramientas analíticas que faciliten su estudio e investigación. En este sentido, entendemos que una de las líneas de investigación prioritarias debería ser el desarrollo de una arquitectura tecnológica –un entorno virtual en la nube–,

¹ Remitimos al lector a la siguiente página electrónica del Instituto Cervantes, donde se ofrecen cifras de la presencia del español en el mundo y de su previsión de crecimiento: http://cvc.cervantes.es/Lengua/anuario/anuario_14/informes/p01.htm

que pudiera proporcionar el acceso unificado a todos estos recursos y herramientas a investigadores, estudiantes, así como a cualquier otra persona interesada en realizar una consulta sobre la lengua, prestando especial atención a personas con diferentes discapacidades y así lograr que el español tenga una mayor proyección.

Reflexiones en torno a los recursos lingüísticos informatizados

En los últimos años, con la consolidación de internet como fuente de información, la proliferación de recursos lingüísticos ha sido abrumadora, desde herramientas sencillas (p. ej., conjugadores verbales o analizadores morfológicos) hasta aquellos más complejos que integran satisfactoriamente técnicas y metodología propia de otros ámbitos como la inteligencia artificial y el procesamiento del lenguaje natural^{II}. Sin embargo, es cierto que con frecuencia se afirma, y no sin cierta razón, que muchos de los recursos existentes en internet (sean gramáticas, diccionarios, glosarios, etc.) no dejan de ser copias de sus correspondientes trabajos en formato papel, es decir, que, en el mejor de los casos, el paso del papel al formato electrónico ha permitido introducir sistemas de búsqueda más refinados que serían impensables en un formato papel. Aun así, uno de los desafíos que se plantean, y a nuestro entender muy razonable, es si se podrían diseñar nuevos recursos que nos permitan sacar el máximo rendimiento a la tecnología y acceder a otro tipo de consultas: por ejemplo, con la ayuda de refinados algoritmos de búsqueda y de una tecnología apropiada, ¿podríamos disponer de un diccionario que, además de buscar las palabras por lemas, pudiera atender consultas conceptuales? Imaginemos que no logro recordar, ni siquiera en la lengua materna, el instrumento que utilizamos para colgar un cuadro o para desarrollar una tarea específica (p. ej., diseccionar un órgano específico durante una operación), es decir, solo recuerdo la función de este instrumento, ¿podríamos disponer de un recurso que, al introducir la función de este instrumento como único parámetro para realizar la búsqueda, nos permitiera recuperar la palabra (p. ej., *taladro* o *bisturí*) y, aún más, sus equivalentes en otras lenguas? Siguiendo con el ejemplo del diccionario, ¿podríamos diseñar sistemas de acceso al léxico que superaran la disposición alfabética y plasmaran las relaciones entre palabras no relacionadas alfabéticamente?

^{II} Como exponentes de esta línea de trabajo remitimos al lector a los siguientes proyectos: PANACEA (www.panacea-lr.eu), un repositorio de recursos lingüísticos orientados principalmente a la creación y desarrollo de laboratorios virtuales para el procesamiento del lenguaje natural; CLARIN-ES-LAB (clarin-es-lab.org), un catálogo de recursos y herramientas que permiten realizar diversas tareas de procesamiento y análisis de textos; DIRT (<http://dirtdirectory.org/>), que incluye un inventario de recursos para el análisis lingüístico así como en el ámbito de las humanidades digitales; ELRA (<http://www.elra.info/en/>), que tiene la misión de promocionar los recursos lingüísticos relacionados con el sector de las tecnologías lingüísticas y su evaluación (incluye un catálogo de recursos con herramientas gratuitas y de pago); DBPedia (<http://wiki.dbpedia.org>), que es una comunidad colaborativa dedicada a extraer datos estructurados procedentes de Wikipedia, con el objetivo de que estos se conviertan en información, por ejemplo, en información semántica, pragmática, discursiva, etc.

En esta línea, Bosque y Mairal^{1,2} señalan las limitaciones de una organización alfabética del léxico y argumentan, siguiendo a Simone³, que resulta extraño que el conocimiento que cualquier hispanohablante tiene sobre la relación existente entre las entradas *risa* y *ataque de risa* no esté reflejado en la mayoría de los diccionarios, pues parece razonable esperar que o bien en la entrada *risa* o en la de *ataque* se explicara la expresión *ataque de risa*. Igualmente, mientras que se establece una remisión desde el adjetivo *rural* al sustantivo *campo*, no sucede lo mismo desde el sustantivo *campo* al adjetivo *rural* cuando un hablante necesita esta relación en ambas direcciones^{III}. Así, concluyen que sería razonable que algunos diccionarios –precisamente esos “diccionarios que todavía no existen”, como reza el título de la conferencia de Simone (2012)³– recogieran el mayor número posible de estas relaciones y proporcionaran rutas conceptuales de acceso al léxico.

El orden alfabético es el código digital que nos permite encontrar cualquier palabra al instante, pero el consultor de un diccionario debería poder disponer de alguna otra forma de organización, aunque solo fuera porque el último recurso organizativo no debería constituir el único recurso organizativo. Las posibilidades que ofrece la organización conceptual del léxico son muy numerosas (...), y su mayor ventaja es que casi todas son compatibles².

Por consiguiente, el desarrollo de recursos, en este caso lexicográficos, que representen el conocimiento léxico que poseen los hablantes sobre las relaciones conceptuales entre las palabras es, sin duda, una de las líneas de trabajo que conforman la agenda de investigación de las tecnologías del lenguaje^{IV}.

En otras palabras, los nuevos recursos lingüísticos no solo deben sacar el máximo uso de los nuevos y más modernos algoritmos de la investigación tecnológica, sino que, además, deberían ser capaces de responder a las necesidades particulares de los usuarios. A esta línea de trabajo se la conoce, en el campo de la lexicografía, como la Lexicografía del Nuevo Milenio⁴ o, dicho de otro modo, diseñar recursos que, inspirados en los algoritmos tecnológicos más innovadores, permitan al usuario disponer de herramientas con las siguientes funcionalidades: a) determinar sus necesidades; b) extraer y organizar la información que estos demandan en tanto en cuanto esta pueda ser tratada lexicó-

^{III} En la misma línea argumentativa, como el propio Simone (2012)³ afirma, resulta extraño que en la entrada del sustantivo *agua* no se incluyan remisiones a los adjetivos *hidrico*, *acuoso*, *acuático* e incluso, como Bosque y Mairal (2013)² añaden, a los verbos *fluir*, *calar* o a los sustantivos *lluvia*, *tormenta* o *aguacero*, además de las expresiones *gota*, *hilo* (o *un hilo de agua*).

^{IV} Entre los proyectos lexicográficos que ofrecen rutas de acceso al léxico desde un punto de vista conceptual y trascienden el orden alfabético, podemos citar los siguientes (unos en desarrollo y otros en un estadio ya muy avanzado):

Wordnet: <http://wordnet.princeton.edu/>
 Multiwordnet: <http://multiwordnet.fbk.eu/english/home.php>
 Framenet: <https://framenet.icsi.berkeley.edu/fndrupal/>
 Spanish Framenet: <http://sfn.uab.es:8080/SFN/>
 Diccionario de Colocaciones del Español: <http://www.dicesp.com/paginas>
 Teoría Sentido-Texto: <http://olst.ling.umontreal.ca/>
 Visuwords: <http://www.visuwords.com/>
 Ecolexicon: http://ecolexicon.ugr.es/visual/index_es.html
 FunGramKB: www.fungramkb.com

gráficamente, y c) configurar la forma óptima de acceso a dicha información^v. Se debe evitar un tratamiento poco accesible de la información, puesto que esto contravendría el principio fundamental de atención a las necesidades del usuario.

En suma, el panorama lexicográfico descrito en la nota a pie de página VI se enfrenta a una serie de problemas, en especial que no se tienen plenamente en cuenta los perfiles de los posibles usuarios ni las necesidades de uso de estos, a lo que se añade el desfase en los sistemas de búsqueda.

Recursos adaptados

Una propiedad fundamental de los diccionarios electrónicos actuales es la de la gran capacidad de búsqueda que ofrecen. El volumen de datos almacenados es mucho mayor que en los diccionarios convencionales y la rapidez de recuperación de información es mucho más alta. En el caso de los diccionarios, la utilización de hiperenlaces permite navegar de un vocablo a otro con rapidez. A esto se une la funcionalidad que aportan los operadores booleanos y los comodines (*wild cards*). Sin embargo, se hace también necesario crear rutas precisas de búsqueda que permitan llegar de manera precisa a la información que se busca y no a más. Esto significa que un diccionario electrónico es más que un sistema de recuperación de información que arroja entradas léxicas; un diccionario electrónico se debe concebir como un sistema de extracción de la información que resulta pertinente para el usuario.

En la práctica, para que un recurso lingüístico se adecúe a las necesidades reales de los usuarios, debe contemplar las necesidades comunicativas específicas de estos⁶. Así, hay ciertos tipos de información cuya inclusión debe depender plenamente del perfil del usuario. En este sentido, y siguiendo con la argumentación de la sección precedente, donde formulamos ideas sobre posibles recursos, entendemos que un recurso innovador sería aquel que fuera capaz de gestionar y presentar la información según las necesidades del usuario, a saber, si se trata de un estudiante que tiene un nivel avanzado de español o, por el contrario, está dando sus primeros pasos en el aprendizaje de esta lengua, o si se trata de un científico que quiere consultar un término en concreto de su campo de especialidad, o si se trata de un usuario con una discapacidad visual o auditiva, etc.

Para establecer correctamente este perfil de usuario es también necesario adaptar las descripciones a la capacidad y nivel del usuario. Mientras que un principiante puede agradecer que el diccionario proporcione equivalencias o traducciones, un estudiante avanzado debería requerir la inclusión de información gramatical y pragmática detalla-

da o, ya en dominio profesional concreto, por ejemplo, un médico puede apreciar tener acceso al uso correcto de términos especializados o incluso a las normas de estilo con notas gramaticales sobre los rasgos discursivos más destacados del dominio de su especialidad^{vi}. Por este motivo, el objetivo fundamental de un recurso lingüístico innovador se debe definir como el de proporcionar, de manera eficaz, información lingüística completa y suficiente. Esta definición supone que, por ejemplo, los diccionarios deben perseguir el equilibrio entre codificación y descodificación, en una adaptación plena a las necesidades del usuario. Resulta paradójico que este equilibrio y la adaptación a perfiles de usuario no hayan guiado todavía la confección de diccionarios electrónicos; estos todavía no aportan toda la información que los usuarios necesitan para el uso lingüístico.

En suma, sería deseable contar con un recurso que, al registrarse el usuario e identificarse con un determinado perfil, mostrara la información que es pertinente a las necesidades de este usuario, evitando una sobreexposición indeseada de información que no responda a sus necesidades, es decir, estaríamos ante una máquina capaz de gestionar de forma inteligente la información que almacena y dotada de recursos hipermedia susceptibles de adaptarse, de manera personalizada, a las necesidades del usuario. Se trataría de una adaptación de la funcionalidad de los buscadores (como "Google") capaces de efectuar sugerencias "inteligentes" al usuario. Estos recursos, que están en pleno desarrollo, emplean técnicas de minado de datos y *clusters*: por ejemplo, algunos modelos de *diccionarios minados* son capaces de analizar entradas de Wikipedia en diversas lenguas y así generar de forma semiautomática entradas bilingües o multilingües⁸. Aun es más, este recurso electrónico debería aspirar a convertirse en un recurso dinámico que evolucione al compás del progreso en el proceso de aprendizaje del usuario. Esto solo será posible en el futuro mediante el diseño de mecanismos adaptativos multimedia (integrando texto, audio y vídeo)⁴ capaces de incorporar técnicas de procesamiento del lenguaje natural y de la inteligencia artificial adecuadas al contexto.

Recursos enlazados

No solo se deben adaptar los contenidos, sino que corresponde también proporcionar al usuario los medios de acceso adecuados a tal información. En este punto, parece lógico pensar que si un usuario está buscando un recurso determinado, por ejemplo un corpus o una gramática, etc., accede a la página electrónica donde está almacenado este recurso. Ahora bien, es evidente que nuestra búsqueda resultaría más eficiente si pudiéramos ingerir información no solo de un único recurso alojado en una página electrónica determinada, sino de un conjunto de recursos que están enlazados

^v Siguiendo a Fuertes-Olivera y Tarp (2014)⁵ se pueden distinguir cuatro tipos de diccionarios en línea: 1) *copias al carbón*, es decir, diccionarios tradicionales pero impresos en formato PDF, lo cual los hace accesibles a través de internet; 2) *caballos más rápidos*, cuya concepción es la misma que la del diccionario impreso, solo que están dotados de mayor versatilidad al contar con varias rutas de acceso; 3) *balas perdidas*, esto es, diccionarios que se basan en los aspectos más llamativos de la tecnología en cuanto al tratamiento de los datos, las rutas de acceso o las necesidades de los usuarios; 4) *modelo T Ford*, como, por ejemplo, el *Diccionario Español de Contabilidad*, o diccionarios dirigidos a tipos de usuario y contextos de uso diferenciados.

^{vi} En el ámbito de los lenguajes de especialidad, remitimos al lector a DEXTER (Discovering and Extracting Terminology), una plataforma online multilingüe formada por una serie de herramientas con diferentes funcionalidades, tales como la compilación y gestión de corpus, la indexación y recuperación documental, la elaboración de consultas basadas en expresiones regulares, la exploración textual y la extracción y validación terminológica (cf. Periñán-Pascual, 2015)⁷.

lexicográfica de un modo más eficiente, sino que, además, permite realizar búsquedas más facetadas e incluso poder llegar a enlazar los recursos con otros recursos, de tal forma que nuestros recursos, lejos de formar parte de un portal aislado, puedan formar parte de una nube de recursos en la que todos están enlazados.

Infraestructuras de recursos

Así, en estrecha relación con la línea de trabajo sobre aprendizaje automático (*machine learning*), entendemos que, a partir de la tecnología con la que contamos, podríamos diseñar infraestructuras con aplicaciones de las tecnologías del lenguaje que tienen un destacado impacto en el sector empresarial en los siguientes ámbitos:

- Los sistemas de minería de opiniones o análisis de sentimientos tienen especial relevancia para aquellos sectores que desean invertir en la inteligencia de los negocios, analizando los comentarios de consumidores en redes sociales con el fin de evaluar la aceptación de un producto.
- La elaboración automática de resúmenes, y más concretamente de resúmenes que estén formados por agrupaciones de realizaciones lingüísticas presentes en el texto original.
- La categorización de un único documento o de toda una colección de documentos entre los dominios especializados de la base de datos IATE.
- La recuperación de documentos, donde la búsqueda de datos se centraría en grupos de documentos cuya temática fuera similar, agilizando así el procesamiento de los recursos.
- Mientras que existe un buen número de analizadores morfológicos y sintácticos, el número de analizadores semánticos es mucho menor precisamente por la complejidad de la tarea. Supone desarrollar una aplicación que permita a una máquina analizar automáticamente la representación semántica de un texto, y así entender su significado, lo que supondría mejorar los sistemas de extracción y recuperación de la información^{viii}.
- En esta línea de trabajo se engloba todo lo relacionado con el aprendizaje automático (*machine learning*) o el aprendizaje profundo (*deep learning*).
- A diferencia de lo que sucedía hace unos años, en los que los lingüistas utilizaban corpus para ejemplificar sus teorías, hoy es cada vez más frecuente realizar estudios sobre una palabra o un patrón gramatical desde un punto de vista pragmático y sociolingüístico utilizando el inmenso caudal de información que circula por las redes sociales, especialmente *twitter*. A tal efecto se han desarrollado recursos sofisticados que permiten analizar el desarrollo de una palabra o concepto en millones de tuits atendiendo a factores sociolingüísticos, pragmáticos, etc. Como exponente de esta tecnología, les remitimos a la siguiente página electrónica: <http://gplsi.dlsi.ua.es/gplsi13/es/node/251>

^{viii} Como exponentes de esta línea de trabajo remitimos al lector a Voyant (<http://voyant-tools.org>), Babelfy (<http://babelfy.org>) y ARTEMIS (Automatically Representing TExt Meaning via an Interlingua-based System).

Recursos lexicográficos del español en la era de internet

Una de las infraestructuras más conocidas en el ámbito de la gestión de recursos del español científico está relacionada con la lexicografía de internet, que ha modificado de forma sustancial la teoría y práctica de la lexicografía; ha traído consigo cambios tan importantes que la lexicografía de internet (o *e-lexicografía*) está sujeta a un gran debate teórico y práctico que va desde la misma naturaleza de la lexicografía al análisis de los diversos productos que pueden desarrollarse. En este capítulo nos limitaremos a presentar los dos enfoques principales completos, es decir, con soporte teórico y aplicaciones prácticas, que pueden relacionarse con la gestión de recursos del español científico^{ix}.

Uno de los enfoques —es el enfoque mayoritario y dominante, quizás porque puede asociarse de forma clara y nítida a empresas e instituciones tan importantes y conocidas como Oxford, Cambridge, Real Academia Española, Real Academia Nacional de Medicina, Larousse, Planeta, etc.— mantiene una concepción tradicional del diccionario en línea como *producto lingüístico* con dos modelos de negocio fundamentales: *a*) diccionarios de acceso libre financiados con anuncios o patrocinio (p. ej., la edición vigesimotercera del DRAE); *b*) diccionarios gratuitos y diccionarios con suscripción accesibles en un único portal lexicográfico. La diferencia entre la consulta gratuita y la de pago está en el número y tipo de diccionarios al que podemos acceder. Por ejemplo, en el portal Diccionarios.com (<http://www.diccionarios.com/>), Larousse ofrece de forma gratuita versiones poco actualizadas de un diccionario general y de varios diccionarios bilingües, así como varios tipos de suscripciones a versiones más actualizadas de alguno de los diccionarios y a un buen número de diferentes tipos de diccionarios, entre ellos diccionarios especializados para semiexpertos (p. ej., traductores) o para expertos, como puede ser el *diccionario médico de consulta* (fig. 2).

Para los partidarios de este enfoque, internet es *simplemente* una tecnología gracias a la cual puede ofrecerse un mayor número de entradas lexicográficas y accesos más rápidos a estas (fig. 2). Es decir, en este modelo, la tecnología existente y sus diferentes posibilidades se orientan a incrementar los recursos ofrecidos con varios tipos de acceso. Por regla general, este modelo descansa en la agregación de diccionarios —normalmente diccionarios conceptualmente diferentes—, que obligan al usuario potencial de estos a un trabajo extra de interpretación de los datos lexicográficos, ya que es habitual que datos lexicográficos idénticos se ofrezcan al usuario de varias formas y con formatos diferentes. Por ejemplo, la búsqueda de *radiculitis* en Diccionarios.com (portal de Larousse) recupera dos definiciones diferentes (una para un hablante normal y otra para un lego interesado o semiexperto), marcas gramaticales en varias formas (por ejemplo, “f”, “s.f” y “nombre femenino” para la misma entrada, y marcadores de especialidad diferentes, por ejemplo, “medicina” y “Anat. Patol.”). Además, hay un

^{ix} Remitimos al lector a los trabajos ya publicados o en vías de publicación sobre la lexicografía de internet y sus diversos temas objeto de estudio e investigación: Durkin, 2016⁹; Fuertes-Olivera, 2016¹⁰ y 2017¹¹; Fuertes-Olivera y Bergenholtz, 2011¹²; Fuertes-Olivera y Tarp, 2014⁵; Granger y Paquot, 2012¹³; Hanks y De Schryver, 2017¹⁴; Jackson, 2013¹⁵.

The image shows a search for 'radiculitis' on the Diccionarios.com website. It displays several search results:

- Top banner:** 'Lengua española' with a search icon.
- Result 1:** 'radiculitis' (s.f. MEDICINA) with a definition: 'Inflamación de la raíz de un nervio raquídeo, a veces secundaria a una infección o por fenómenos compresivos.' Includes a note: 'NOTA: En plural: radiculitis' and '© Larousse Editorial, 2012'.
- Advertisement:** 'Inversión Energía Solar' with details: 'Desde 10.000€. Rentabilidad 7,5%. Ventajas Fiscales. ¡Infórmese Aquí!' and a green arrow button.
- Result 2:** 'radiculitis' (nombre femenino) with a definition: 'radiculitis.' Includes a note: '© Vox, marca registrada por Larousse Editorial'.
- Result 3:** 'Diccionario médico de consulta' with a search icon.
- Result 4:** 'radiculitis' (f. ANAT. PATOL.) with a definition: 'Inflamación de las raíces nerviosas craneales o raquídeas a causa de una intoxicación, una infección o un traumatismo, como la compresión de un disco intervertebral. Los síntomas suelen manifestarse en forma de trastornos sensitivos, motores o como una disminución de los reflejos.' Includes a note: '© Vox, marca registrada por Larousse Editorial'.
- Bottom banner:** '"radiculitis" no aparece en nuestros foros'.

Figura 2 Búsqueda de *radiculitis* en Diccionarios.com

equivalente inglés y un fraseo terminológico español y su traducción inglesa (*radiculitis braquial aguda* y *acute brachial radiculitis*) aislado y sin contexto de uso. En resumen, solo un usuario que ya sepa que todos estos datos presentados de forma diferente y fuera de contexto ofrecen la misma información puede consultar este tipo de diccionarios con una cierta garantía de éxito. Por ejemplo, este usuario debe saber (o perder el tiempo en descubrirlo) que “f” y “nombre femenino” hacen referencia a la misma realidad gramatical.

El enfoque que acabamos de señalar en los párrafos anteriores puede observarse con alguna variante en la mayoría de los recursos relacionados con la gestión del español científico, especialmente en portales de organizaciones supranacionales en las que el español es una de las lenguas oficiales (p. ej., IATE en la Unión Europea, <http://iate.europa.eu/switchLang.do?success=mainPage&lang=es>) y diversos glosarios y recursos terminológicos de las Naciones Unidas (ver la base de datos terminológica de la FAO, <http://www.fao.org/faoterm/terminology-at-fao/es/>).

Este enfoque también está presente en el desarrollo de prototipos, es decir, de herramientas de información con un número muy limitado de entradas finalizadas^x y de “lexicones”, es decir, repositorios léxicos integrados en bases de

conocimiento legibles por máquinas o por humanos —p. ej., el glosario de “MailChimp”, que es una herramienta que traduce automáticamente los formularios de suscripción; también el glosario de CRISOL, que es una base de conocimiento de opiniones (p. ej., opinión sobre hoteles) para el español—¹⁶. Quizás el proyecto con más futuro en esta línea de trabajo e investigación sea *DBpedia*, “que es un proyecto para la extracción de datos de Wikipedia para proponer una versión *web* semántica.” (<https://es.wikipedia.org/wiki/DBpedia>). En términos generales, podemos afirmar lo siguiente:

- Esta línea de trabajo es muy activa y goza de bastantes recursos económicos destinados a la construcción de este tipo de herramientas; la mayoría de los proyectos realizados en España sigue este enfoque, destacando los trabajos realizados en Granada (p. ej., *EcoLexicon*, <http://ecolexicon.ugr.es/es/>), Cataluña (p. ej., *Diccionario enciclopédico de medicina*, <http://www.grec.net/home/cel/mdicc.htm>) y Comunidad Valenciana (<http://www.iulma.es/index.php/diccionarios-iulma/>).
- Clasifica las herramientas que construye atendiendo a postulados tradicionales como los que defienden la existencia de tipologías que diferencian entre diccionarios, glosarios, enciclopedias, ontologías, bases de datos terminológicas, bases de datos del conocimiento, etc.
- Está muy conectada con la investigación lingüística, especialmente porque sigue defendiendo que los diccionarios y productos similares son ante todo y sobre todo productos lingüísticos^{xi}.

El segundo de los enfoques —es un enfoque minoritario relacionado con la teoría funcional de la lexicografía iniciada en la Universidad de Aarhus a finales de la década de los noventa del pasado siglo⁵— defiende que el diccionario es una herramienta de información que debe conceptualizarse como un servicio en vez de como un producto—. Hablar de servicio en vez de producto tiene varias implicaciones importantes, algunas de las cuales son importantes para este artículo.

- La primera de las implicaciones es que el término *diccionario* es inclusivo en tanto en cuanto no diferencia entre distintos tipos de diccionarios (especializado, general, base de datos terminológica, glosario, etc.); en otras palabras, para los partidarios de este enfoque, un diccionario es una herramienta de información diseñada, compilada y actualizada para satisfacer las necesidades puntuales de información de la forma más rápida y amigable posible. Son necesidades específicas que un tipo de usuario específico puede tener en una o varias situaciones de uso específicas. En términos prácticos, esta idea implica que el diccionario no es una agregación de varios diccionarios (como vimos en la figura 2), sino un todo unitario.

^x Por razones de espacio no nos vamos a ocupar en este trabajo del análisis de estos prototipos. Aunque puedan tener nombres diversos todos ellos están interesados en construir diccionarios o herramientas de información y consulta, como el *Multilingual Glossary of Technical and Popular Medical Terms in Nine European Languages* (<http://users.ugent.be/~rvdstich/eugloss/welcome.html>).

^{xi} Remitimos al lector interesado a <http://www.lexicool.com/>. En este sitio hay información sobre 8.000 diccionarios en 70 lenguas. Por ejemplo, la búsqueda “Spanish-English” recupera un buen número de recursos lingüísticos, entre ellos varios diccionarios, conjugadores, analizadores, gramáticas, etc. (ver <http://www.lexicool.com/spanish-dictionary-translation.asp>).

- La segunda implicación es que un diccionario inclusivo pensado para ser usado en diferentes situaciones de uso y por diferentes tipos de usuario necesita una base de datos lexicográfica idiosincrática, es decir, una base de datos diseñada exclusivamente para el proyecto y dotada de tecnologías que permitan construir herramientas de información dinámicas, es decir, herramientas con sistemas de búsquedas diferentes adaptados y pensados para usuarios diferentes y situaciones de uso diferentes. Las principales tecnologías utilizadas en la actualidad para construir diccionarios integrados conceptualmente con datos dinámicos son las siguientes: buscadores y navegadores (*searching and navigating*); creadores de perfil (*user profiling*); sistemas de filtrado de datos (*filtering*); hipermedia adaptativo (*adaptive hipermedia*); marcadores de metadatos (*metadata markup*); recursos enlazados (*linked open knowledge*); sistemas de recomendación (*recommender systems*); sistemas de búsquedas maximizadoras o minimizadoras (*maximising and minimizing searches*); anotadores (*annotation systems*).
- La tercera implicación es que el diccionario es un conjunto *integrado* de 3 componentes básicos: la base de datos lexicográfica, los motores de búsqueda y la interfaz (o interfaces). Gracias a esta idea, los defensores de este enfoque no agregan diccionarios, como los que vimos en la figura 2, sino que trabajan en un sistema integrado utilizando normalmente el modelo de centro (*hub* en inglés) y radios (*spoke* en inglés). Esto quiere decir que normalmente hay un diccionario base o diccionario nodal (p. ej., un diccionario de español general) y diccionarios conectados con el diccionario base. Con este sistema no solo se facilita la consulta, sino que también se ahorra dinero y tiempo al compilarlos (p. ej., los datos almacenados del diccionario “a” pueden reutilizarse para el diccionario “b”; p. ej., la información gramatical de un lema en un diccionario general puede trasladarse a un diccionario bilingüe). En resumen, en un modelo integrado no hay divergencias como las que vimos en la figura 2 relacionadas con la descripción gramatical de *radiculitis*.
- La cuarta implicación es que ser un servicio exige un tipo de acceso las 24 horas del día de cada día de cada año. Al plantearlo como servicio es necesario que el diccionario se actualice con facilidad y con rapidez, que haya un servicio de ayuda a disposición del usuario, que el sistema de computación en nube utilizado no se “caiga”, y que exista una comunicación fluida y rápida entre el usuario potencial y el lexicógrafo.
- Finalmente, el diccionario se plantea como una herramienta lexicográfica (es decir, no es un producto lingüístico) compuesta por tres elementos que tienen entre sí una relación de interdependencia mutua: usuarios; datos lexicográficamente relevantes; rutas de acceso. Esta idea de la interdependencia de los tres componentes lexicográficos es muy relevante en este enfoque y está en el origen de algunos de los desarrollos más importantes. Uno de estos desarrollos procede del uso de internet como corpus lexicográfico¹⁷. Este hecho tiene tres implicaciones lexicográficas relevantes: a) se está dejando a un lado la idea de que el lexicógrafo debe limitar las fuentes lexicográficas a un tipo concreto, por ejemplo a un corpus lingüístico; b) se hace cada vez más uso de los recursos enlazados, como puede ser una foto de un ani-

mal para describirlo, o un vídeo de “youtube”, o una entrada de un conjugador, o de Wikipedia para completar una entrada lexicográfica; c) el proceso de almacenamiento de los datos debe contemplarse como una tarea lexicográfica similar a otras tareas lexicográficas tradicionales. Es decir, es necesario responder a preguntas como estas: ¿quién puede buscar los datos?, ¿dónde pueden buscarse los datos?, ¿cómo pueden tratarse los datos encontrados?, ¿cuándo pueden buscarse los datos?, ¿cómo pueden entrelazarse los datos?, etc.^{xii}. Como hemos dicho antes, este enfoque es novedoso y minoritario. En español hay dos proyectos lexicográficos basados en el mismo: los *Diccionarios de Contabilidad*; los *diccionarios en línea de español “Valladolid-UVa”*.

Los Diccionarios de Contabilidad

Los Diccionarios de Contabilidad son un conjunto de 12 diccionarios especializados que describen los hechos contables y su lengua en inglés y en español^{xiii,21-23}. El diseño y construcción de estos diccionarios se llevó a cabo en la Universidad de Valladolid a partir del año 2008 tomando como modelo los *Accounting dictionaries* diseñados en el año 2003 en la Universidad de Aarhus, con el objetivo de describir las lenguas y los hechos contables en inglés y en danés. Tanto los *Diccionarios de Contabilidad* como los *Accounting dictionaries* los comercializa la empresa danesa Ordbogen A/S. Los diccionarios inglés-danés ya se ofrecen en Escandinavia en forma de suscripción^{xiv}, mientras que los diccionarios inglés-español son de consulta libre y gratuita en <http://lemma.com/>^{xv}.

Para los objetivos de este capítulo, lo más destacado de los Diccionarios de Contabilidad es que ofrecen al usuario 12 posibilidades de búsqueda diferente, cada una de ellas adaptada a una situación de uso específica. Esta búsqueda puede realizarse por lema completo, parte de lema, por concepto o por parte de un concepto. Por ejemplo, un usuario puede buscar *pasivo* y recuperar diversos artículos según el botón utilizado. Por ejemplo, si busca *pasivo* en el diccionario “Spanish-English” tiene cuatro posibilidades: con el botón “recepción” recupera solamente la definición, la clase gramatical del lema y un equivalente inglés (fig. 3); con el botón “traducción” recupera también la clase gramatical del equivalente, y colocaciones y ejemplos traducidos

^{xii} Remitimos al lector interesado al volumen 31 de *Lexicographica*, publicado en el año 2015, en el que algunos artículos hacen referencia a alguno de estos problemas¹⁸⁻²⁰.

^{xiii} Remitimos al lector interesado a Fuertes-Olivera yTarp (2014)⁵, que, entre otras cosas, describen los fundamentos teóricos y prácticos con los que trabajan los partidarios de este enfoque.

^{xiv} Los principales clientes de estos diccionarios son las empresas, los despachos de abogados y consultores, los traductores *freelance* y las instituciones, entre ellas las universidades del norte de Europa. Se ofrecen en paquetes y tienen una gran demanda, ya que parece que ahorran el tiempo de búsqueda en cerca de un 30% frente a los diccionarios tradicionales; además, encuestas entre usuarios aseguran que estos diccionarios ofrecen gran seguridad y fiabilidad.

^{xv} Está previsto un rediseño completo de este sitio *web* domiciliado en Alemania. En los próximos meses se decidirá si estos diccionarios se ofrecen en forma de suscripción o son de acceso libre (en este último caso no se procedería a su actualización constante).

The screenshot shows the 'Recepción' button highlighted in the 'Search methods' section of the dictionary interface. The main content area displays the definition and liability information for the word 'pasivo'.

Recepción

pasivo 1 nombre masculino

Definition
El pasivo es el conjunto de deudas actuales de una empresa que proceden de sucesos pasados y cuyo pago es anticipado con el objetivo de producir una salida de flujos de efectivo de la empresa que son beneficios económicos.

liability 1

pasivo 2 adjetivo

Definition
Ser pasivo es aceptar lo que ocurre sin intentar su control o cambiar los hechos o reaccionar a las cosas, o no participar en la toma de decisiones.

passive

Figura 3 Pasivo con el botón “recepción” (Diccionario Español-Inglés de Contabilidad).

The screenshot shows the 'Traducción' button highlighted in the 'Search methods' section. The main content area displays the definition, liability information, and collocations for the word 'pasivo'.

Traducción

pasivo 1 nombre masculino

Definition
El pasivo es el conjunto de deudas actuales de una empresa que proceden de sucesos pasados y cuyo pago es anticipado con el objetivo de producir una salida de flujos de efectivo de la empresa que son beneficios económicos.

liability 1 noun <a liability, the liability, liabilities>

Collocations

- un pasivo a largo plazo
a long-term liability
- un pasivo reconocido
a recognised liability
- eliminar un pasivo individual de un grupo clasificado como mantenido para la venta
remove an individual liability from a disposal group classified as held for sale
- asumir un pasivo

Figura 4 Pasivo con el botón “traducción” (Diccionario Español-Inglés de Contabilidad) (extracto).

(fig. 4); con el botón “conocimiento” recupera también sinónimos (fig. 5) y con el botón “frases y expresiones” recupera contextos de uso en los que aparece el término buscado (fig. 6); clicando en cada uno de ellos, el usuario recupera todos los artículos lexicográficos en los que aparece el contexto de uso (fig. 7):

Diccionarios en línea de español “Valladolid-UVa”

El proyecto lexicográfico de los *diccionarios en línea de español “Valladolid-UVa”*^{xvi} se inició en mayo de 2012 en una reunión en la sede central de la empresa danesa Ordbogen A/S y los editores del proyecto (Pedro A. Fuertes-Olivera y Henning Bergenholtz). A lo largo del año 2013 se diseñaron

^{xvi} Es posible que el nombre “Valladolid-UVa” pueda modificarse dependiendo de que encontremos uno o varios patrocinadores del proyecto que aporten cantidades económicas al mismo.

los conceptos lexicográficos del proyecto, se creó la base de datos lexicográfica, se seleccionaron las tecnologías a utilizar, así como a los lexicógrafos y personal IT que iba a trabajar en el proyecto. El trabajo propiamente lexicográfico de búsqueda de datos y creación de fichas lexicográficas se inició en marzo de 2014 y en la fecha en la que escribimos este trabajo (abril de 2016), la base de datos del proyecto tiene 120.000 cartas lexicográficas, 35.000 de las cuales ya están completas^{xvii}. El futuro del proyecto vendrá determinado *fundamentalmente* por 3 factores: a) disponibilidad económica; b) apoyo de Ordbogen A/S o empresa similar; c) posibles desarrollos tecnológicos e innovaciones.

Para los objetivos de este capítulo merecen destacarse tres aspectos.

^{xvii} Cada carta incluye 30 campos, 29 de ellos para incluir datos lexicográficos y uno para la comunicación entre todos los participantes en el proyecto. El sistema está alojado en la nube de Ordbogen A/S y se accede a ella desde cualquier lugar del mundo con conexión a internet.

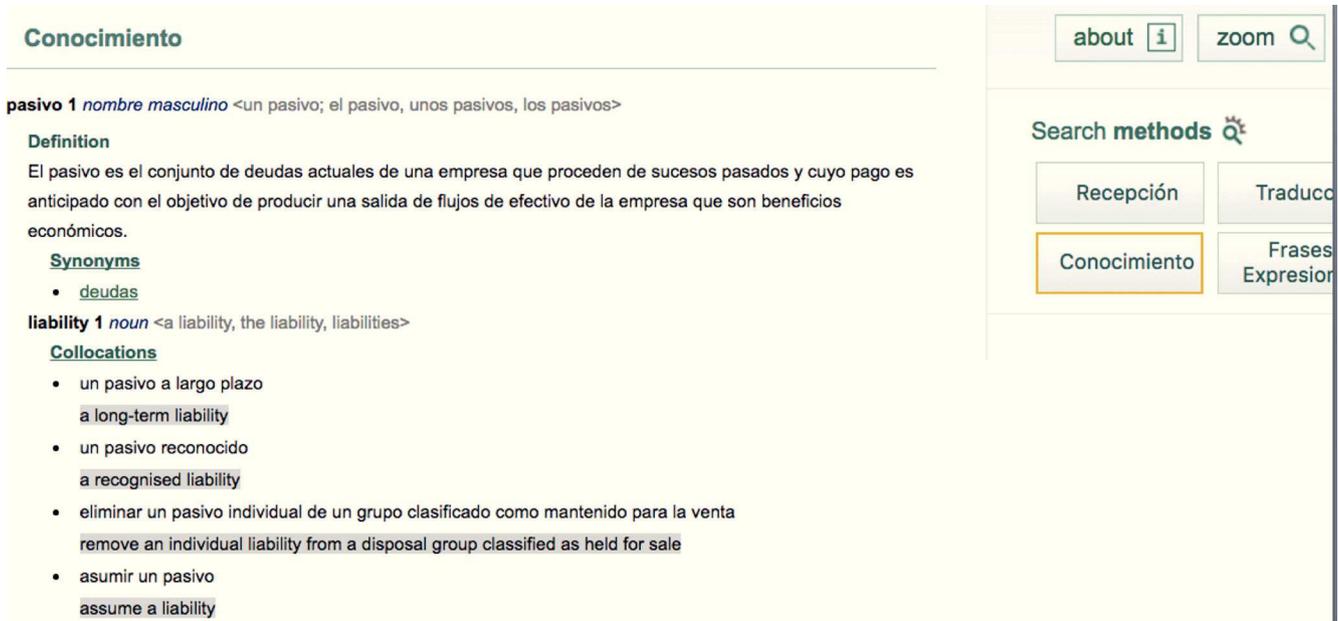


Figura 5 Pasivo con el botón “conocimiento” (Diccionario Español-Inglés de Contabilidad) (extracto).



Figura 6 Pasivo con el botón “frases y expresiones” (Diccionario Español-Inglés de Contabilidad) (extracto).

- El primero de ellos es que el proyecto comprende siete diccionarios relacionados con el español general: a) diccionario general del español; b) diccionario de ortografía y gramática; c) diccionario de expresiones idiomáticas y similares; d) diccionario de definiciones; e) diccionario para la producción de textos; f) diccionario de sinónimos y antónimos; g) diccionario conceptual, al que nos referiremos al final de esta sección.
- El segundo de los aspectos es que es un proyecto de diccionarios *dinámicos* e *integrados*, es decir, todos ellos extraen los datos de la misma base de datos lexicográfica, todos ellos utilizan las mismas tecnologías y metodologías y todos ellos ofrecen solamente los datos adaptados

- a la función prevista de antemano, por ejemplo, solo datos para comprender el significado de una palabra o para producir un texto, como veremos más adelante.
- El tercer aspecto es que en este proyecto el español es el sistema nodal (*hub*) del modelo nodo y radios antes mencionado. Esto quiere decir que cualquier ampliación a otras lenguas (está previsto incluir el inglés) se hará según el modelo diseñado para el español.

Aunque en estos momentos el proyecto no es público y solo puede ser consultado por sus editores, presentamos a continuación algunas pantallas de una versión beta, con la que queremos mostrar cuatro elementos esenciales de este

Frasas y Expresiones

actual *adjetivo*

Definition
Actual indica que algo ocurre o está referido al momento presente.

current *adjective*

Collocations

- gasto actual
current expenditure
- pagos actuales
current payments
- compromisos actuales
current commitments
- recibos actuales
current receipts
- ventas actuales
current sales
- gastos actuales
current expenses
- ingreso actual
current income
- supuestos actuariales actuales
current actuarial assumptions
- activos y pasivos por impuesto corriente y diferido
current and deferred tax assets and liabilities
- periodos actual y anterior

Figura 7 Entrada de *actual* al clicar en la frase “activos y pasivos por impuesto corriente y diferido” (Diccionario Español-Inglés de Contabilidad) (extracto).

proyecto que, en nuestra opinión, lo conectan con la gestión del español científico en la era de internet.

- La primera característica es que la base de datos lexicográfica está unida a un cargador de lemas, es decir, a un programa informático que permite al lexicógrafo responsable de la selección de los lemas (Pedro A. Fuertes-Olivera en este proyecto) introducir los lemas en la base de datos de forma sencilla y fácil. Normalmente, estos lemas se copian y pegan en el cargador de lemas como ficheros de Word. En nuestro caso, los lemas introducidos hasta la fecha en la base de datos (alrededor de 90.000) proceden de leuarios libres capturados con *crawlers*^{xviii} y de listados sistemáticos, la mayoría de los cuales se ha creado con datos extraídos de portales de acceso libre y gratuito, como este archivo que contiene los ríos más largos de la Tierra^{xix} (fig. 8).

^{xviii} Son programas que rastrean la red y capturan todo aquello para lo que han sido programados, por ejemplo para capturar y ordenar listados alfabéticos de palabras.

^{xix} La dirección es https://es.wikipedia.org/wiki/Anexo:R%C3%ADos_m%C3%A1s_largos_de_la_Tierra

Lemma loader

Paste a list of lemmas here!

Save lemmata

Figura 8 Pantalla del programa “Lemma loader” para asignar automáticamente un lema a una carta lexicográfica.

Den Spanske Grammatik- og Staveordbog

ababillarse verbo

Flexiones
 me ababillo, me ababilé, me ababilaba, me ababilaré, me ababilaría, me ababilie, ababilado, ababilandome ababillo, me ababilé, me ababilaba, me ababilaré, me ababilaría, me ababilie, ababilándome

Referencia gramatical
 → [conjugación del verbo](#)

Nota gramatical y/o ortográfica:
 El pronombre concuerda con la persona verbal: me ababillo, te ababilas, te ababilás, se ababila, nos ababilamos, os ababiláis, se ababilan.

Figura 9 Pantalla del verbo *ababillarse* en el diccionario de gramática y ortografía.

Este cargador ordena los lemas alfabéticamente de forma automática y elimina los lemas repetidos. Con esta fórmula no solo se ahorra una gran cantidad de tiempo, sino también de dinero, ya que, por ejemplo, no es necesario copiar cada lema en la base de datos, ni crear corpus, ni utilizar extractores de lemas y términos. Más aún, el programa garantiza la inclusión de cualquier lema en cualquier momento y su ordenación en el lugar adecuado sin desestabilizar el trabajo realizado hasta la fecha y eliminando la necesidad de publicar ediciones sucesivas de los diccionarios. En definitiva, el cargador de lemas hace posible la actualización constante y continua del lemario de la base de datos lexicográfica y, con ella, de los diccionarios.

- La segunda característica es que el diccionario hace un uso habitual de los recursos enlazados. Alrededor del 35% de las cartas lexicográficas utilizan recursos enlazados. Por ejemplo, en este diccionario todos los verbos van unidos a un conjugador de verbos, que permite al usuario ver su conjugación completa clicando en el hipervínculo “conjugación del verbo” (fig. 9). Igualmente, todos los animales y plantas llevan un hipervínculo a una imagen, que se despliega al clicar sobre ella; esta imagen facilita enormemente el trabajo de recepción de este tipo de lemas. La figura 10 muestra la imagen desplegada correspondiente a uno de los significados de *ababa* (amapola).
- La tercera característica es que los usuarios solo recuperan los datos que necesitan en una situación concreta. Por ejemplo, si buscan en el diccionario de ortografía y gramática (fig. 9) solo recuperan los datos ortográficos y gramaticales del verbo *ababillarse*. Si buscan en el dic-



Figura 10 Imagen de *ababa* (amapola) en los diccionarios en línea de español.

cionario de definiciones solo recuperan las definiciones, por ejemplo de *ababa* (fig. 11), las definiciones y sinónimos en el diccionario de sinónimos y antónimos (fig. 12) y todos los datos disponibles (incluyendo colocaciones, ejemplos, recursos enlazados) en el diccionario general (fig. 13).

- Finalmente, el sistema incluye un diccionario que permite al usuario buscar por conceptos. Es un diccionario para la producción y para extraer conocimiento, gracias al cual un usuario puede embarcarse en varias búsquedas hasta encontrar lo que busca. Por ejemplo, si alguien quiere saber la posible relación léxica entre “radioter-

Den Spanske Betydningsordbog

ababa nombre

1.
Definición
 tipo de planta herbácea de ciclo anual que suele crecer en campos sembrados caracterizada generalmente por su flor roja y semilla negra; en ocasiones se utiliza con fines medicinales por sus propiedades sedantes

2.
Definición
 flor de la planta herbácea de ciclo anual del mismo nombre que suele crecer en campos sembrados caracterizada generalmente por su flor roja y semilla negra; en ocasiones se utiliza con fines medicinales por sus propiedades sedantes

Figura 11 *Ababa* en el diccionario de definiciones.

Den Spanske Synonym og Antonym Ordbog

ababa *nombre*

1. Definición
 tipo de planta herbácea de ciclo anual que suele crecer en campos sembrados caracterizada generalmente por su flor roja y semilla negra; en ocasiones se utiliza con fines medicinales por sus propiedades sedantes

Sinónimo(s)
 amapola
 ababol

2. Definición
 flor de la planta herbácea de ciclo anual del mismo nombre que suele crecer en campos sembrados caracterizada generalmente por su flor roja y semilla negra; en ocasiones se utiliza con fines medicinales por sus propiedades sedantes

Sinónimo(s)
 amapola
 ababol

Figura 12 *Ababa* en el diccionario de sinónimos y antónimos.

Den Spanske Netordbog

ababa *nombre*

Flexiones
 una ababa, la ababa, unas ababas, las ababas

1.

Definición
 tipo de planta herbácea de ciclo anual que suele crecer en campos sembrados caracterizada generalmente por su flor roja y semilla negra; en ocasiones se utiliza con fines medicinales por sus propiedades sedantes

Hipervínculo
[ficha técnica](#)

Sinónimo(s)
 amapola
 ababol

Colocacion(es)
 roja el alba, roja la ababa
 el aroma de la ababa

Ejemplo(s)

- Me pregunto quién fue el primero en descubrir que la ababa es sedante.
- Y ahora deseaba de modo denodado tener una ababa en mi tabuco, tener una amapola en mi habitación.

Figura 13 *Ababa* en el diccionario general (extracto; significado 1).

“pia” y “cáncer” puede buscar “radioterapia + cáncer” y encontrará 19 entradas (fig. 14). Clicando en cada una de ellas puede ir precisando la posible relación conceptual entre ambos términos.

Conclusión

Además de su importancia como lengua global, con unos 500 millones de hablantes nativos y de su valor cultural, el español es también una de las lenguas dominantes en la era de internet. Este dominio está muy ligado a varios factores, dos de los cuales son relevantes para este artículo: *a)* el desarrollo de la digitalización, por ejemplo, la relacionada con las redes sociales y la minería de datos; *b)* la aparición de nichos de mercado que, además de contribuir

al desarrollo de la economía basada en el conocimiento, hacen posible la innovación constante, por ejemplo, en el desarrollo de recursos lingüísticos como conjugadores, analizadores morfológicos, extractores y diccionarios en línea. En este trabajo hacemos una reflexión más detallada sobre estos últimos desde una perspectiva lo más completa posible; va desde la base teórica que sustenta el proceso de conceptualización del diccionario hasta el modelo de negocio en el que se apoya. En concreto presentamos los dos enfoques principales, uno de orientación más tradicional y otro de orientación más innovadora. Ambos enfoques están modificando de forma sustancial la gestión del lenguaje científico del español y esperamos que contribuyan a la difusión del conocimiento experto y semiexperto en nuestra época, caracterizada por la rapidez, la inmediatez y la globalización.

The screenshot shows the Ordbogen.com website interface. At the top left, the logo reads 'DANMARKS ONLINE ORDBOG Ordbogen.com'. A search bar contains the text 'radioterapia + cancer' and a 'Search' button. To the right, there is a dropdown menu labeled 'Den Spanske Ordfinder'. Below the search bar, there are filters for '(19)' and 'Other searches'. The main content area displays three articles: 'radioactividad', 'radioterapia', and 'radioterápico', each with a brief description. On the right side, there is a sidebar with a '@' icon and the text 'SUGGEST TERMS/ADD COMMENTS', a magnifying glass icon labeled 'zoom', and a section titled 'Most recent searches' listing several related terms like 'radioterapia - cancer' and 'radioterapia + cancer'.

Figura 14 Búsqueda de “radioterapia + cáncer” en diccionario conceptual.

Bibliografía

- Bosque I, Mairal R. Definiciones Mínimas. En: Rodríguez González F, editor. Estudios de lingüística española. Homenaje a Manuel Seco. Alicante: Universidad de Alicante; 2012. p. 123-36.
- Bosque I, Mairal R. Hacia una organización conceptual del definiens. Capas nocionales del adverbio arriba. En: Corbella D, Dorta J, Fajardo Aguirre A, Izquierdo L, Medina López J, Nelsi Torres A, editores. Lexicografía Hispánica del Siglo XXI: Nuevos Proyectos y Perspectivas. Homenaje al Profesor Cristobal Corrales Zumbado. Madrid: Arco/Libros; 2013. p. 125-50.
- Simone R. Diccionarios que todavía no existen. V Congreso internacional de lexicografía. Madrid: Universidad Carlos III, junio de 2012.
- De Schryver GM. Lexicographers' dreams in the electronic dictionary age. Int J Lexicography. 2003;16:143-99.
- Fuertes-Olivera PA, Tarp S. Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography. Berlin/New York: De Gruyter; 2014.
- Periñán-Pascual C, Arcas F. Reusing computer-oriented lexica as foreign-language electronic dictionaries. Anglogermanica Online. 2006;69-93.
- Periñán-Pascual C. The underpinnings of a composite measure for automatic term extraction: the case of SRC. Terminology. 2015;21:151-79.
- Ye Z, Xiangji J, Ben H, Hongfei L. Mining a Multilingual Association Dictionary from Wikipedia for Cross-language Information Retrieval. J Am Soc Inf Sci Technol. 2012;63:2474-87.
- Durkin P, editor. The Oxford Handbook of Lexicography. Oxford: Oxford University Press; 2016.
- Fuertes-Olivera PA. A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. Int J Lexicography. 2016;29:226-47.
- Fuertes-Olivera PA, editor. The Routledge Handbook of Lexicography. London: Routledge; 2017.
- Fuertes-Olivera PA, Bergenholtz H, editors. e-Lexicography: The Internet, Digital Initiatives and Lexicography. London & New York: Continuum; 2011.
- Granger S, Paquot M, editors. Electronic Lexicography. Oxford: Oxford University Press; 2012.
- Hanks P, De Schryver GM, editors. International Handbook of Modern Lexis and Lexicography. Berlin and Heidelberg: Springer; 2017.
- Jackson H, editor. The Bloomsbury Companion to Lexicography. London: Bloomsbury; 2013.
- Molina González MD, Martínez Cámara E, Martín Valdivia MT. CRISOL: Base de conocimiento de opiniones para el español.

- Procesamiento del Lenguaje Natural. 2015;55:143-50 [consultado 27-4-2016]. Disponible en: http://rua.ua.es/dspace/bitstream/10045/49286/1/PLN_55_16.pdf
17. Fuertes-Olivera PA. Lexicography and the Internet as a (Re-) source. *Lexicographica*. 2012;28:49-70.
 18. Bergenholtz H, Agerbo H. Lexicographical Structuring: The Number and Types of Fields, Data Distribution, Searching and Data Presentation. *Lexicographica*. 2015;31:5-38.
 19. Fuertes-Olivera PA. Lexicographical storing: A key lexicographical task in the era of the Internet. *Lexicographica*. 2015;31:68-90.
 20. Tarp S. Structures in the communication between lexicographer and programmer: Database and interface. *Lexicographica*. 2015;31:217-45.
 21. Fuertes-Olivera PA, Bergenholtz H, Nielsen S, Gordo Gómez P, Mourier L, Niño Amo M, et al. Diccionario Inglés-Español de Contabilidad. Base de Datos y Diseño: Richard Almind y Jesper Skovgård Nielsen. Hamburgo: Lemma.com; 2012.
 22. Fuertes-Olivera PA, Bergenholtz H, Nielsen S, Gordo Gómez P, Niño Amo M, De los Ríos Rodicio A, et al. Diccionario Español de Contabilidad. Base de Datos y Diseño: Richard Almind y Jesper Skovgård Nielsen. Hamburgo: Lemma.com; 2013.
 23. Fuertes-Olivera PA, Bergenholtz H, Nielsen S, Gordo Gómez P, Mourier L, Niño Amo M, et al. Diccionario Español-Inglés de Contabilidad. Base de Datos y Diseño: Richard Almind y Jesper Skovgård Nielsen. Hamburgo: Lemma.com; 2014.