LA LEXICOGRAFÍA DE INTERNET: TRABAJANDO EN LA NUBE CON PRODUCTOS LEXICOGRÁFICOS DE LA ECONOMÍA BASADA EN EL CONOCIMIENTO

PEDRO A. FUERTES-OLIVERA
International Centre for Lexicography, Universidad de Valladolid
PABLO GORDO GÓMEZ
International Centre for Lexicography, Universidad de Valladolid
MARTA NIÑO AMO
International Centre for Lexicography, Universidad de Valladolid
ÁNGEL DE LOS RÍOS RODICIO
International Centre for Lexicography, Universidad de Valladolid
ÁNGELES SASTRE RUANO
International Centre for Lexicography, Universidad de Valladolid
MARISOL VELASCO SACRISTÁN
International Centre for Lexicography, Universidad de Valladolid

RESUMEN

La teoría funcional de la lexicografía ofrece el marco teórico y el método adecuado para construir diccionarios de Internet, que son aquellos cuyos conceptos tienen en cuenta la existencia de tecnologías de Internet pensadas para favorecer la individualización o customización. En esta comunicación presentamos el carácter holístico del trabajo lexicográfico, es decir su dependencia de un buen número de saberes y destrezas. Debido a que es prácticamente imposible esperar que una sola persona pueda tener todos los saberes y destrezas necesarios, conviene idear un sistema de trabajo con el que se pueda diseñar, compilar y actualizar el diccionario de Internet del Siglo XXI. En nuestra opinión ese sistema pasa, entre otras cosas, por el trabajo en la nube, que hace posible la participación de muchos expertos, cada uno aportando algo al producto final, el diccionario de Internet. En esta comunicación ilustramos este trabajo con datos procedentes de los accounting dictionaries *o* diccionarios de contabilidad.

Palabras clave: teoría funcional, lexicografía especializada, Internet, contabilidad, nube

ABSTRACT

The function theory of lexicography offers the theoretical approach and method for compiling online dictionaries that use Internet technologies and resources for favouring customization. This type of work demands different skills and knowledge. As it is very difficult to find a single person who has these skills and knowledge, proponents of the function theory favours using a working methodology for designing, compiling, and updating online dictionaries in which experts in different fields, sometimes living thousands of kilometres away, participate. The methodology rests on using resources placed in the cloud. This paper shows how we have being using these technologies and resources in the construction of the accounting dictionaries, a lexicographic project that started in the year 2008 and has already resulted in 27 usage-based online dictionaries.

Keywords: function theory, specialised lexicography, Internet, accounting, cloud

1. INTRODUCCIÓN

En el marco de la *teoría funcional de la lexicografía* (cf. Fuertes-Olivera y Tarp, 2014), la lexicografía de Internet es la teoría y práctica del diccionario de Internet, es decir, la ciencia que se ocupa del estudio teórico y su aplicación práctica en forma de diseño, compilación y actualización de una herramienta de consulta accesible a través de Internet que ha sido creada para satisfacer las necesidades puntuales de información que pudiera tener un usuario específico en una situación de uso extra-lexicográfica concreto.

Con esta definición indicamos que el trabajo relacionado con la lexicografía de Internet exige un cúmulo de saberes y destrezas que encuadramos en dos marcos, uno más general y abstracto y otro, más específico y concreto.

El marco general se aplica a cualquier obra lexicográfica. Defiende que la lexicografía es una ciencia y actividad académica independiente que formula preguntas y respuestas sobre los tres elementos esenciales de cualquier obra de consulta: (i) las necesidades de los usuarios potenciales; (ii) el tratamiento lexicográfico de los datos con el fin de satisfacer las necesidades específicas detectadas en los usuarios potenciales; (iii) el empleo de estructuras o rutas de acceso que faciliten la recuperación de los datos necesarios para solucionar las necesidades de información que dieron lugar a la consulta con el menor coste posible de tiempo y recursos.

El marco específico refleja la gran vocación multidisciplinar de la lexicografía. Además de los condicionantes generales, todos los diccionarios tienen también características específicas y exigen por ello saberes y destrezas específicos. Por ejemplo, un diccionario de Internet solo puede diseñarse, construirse y actualizarse haciendo uso de las tecnologías de Internet por lo que no podrá abordarse un proyecto de esta naturaleza sin la participación en el mismo de expertos en el funcionamiento de las tecnologías de Internet. Igualmente, el diseño, compilación y actualización de un diccionario especializado, por ejemplo, un diccionario de contabilidad, no podrá abordarse sin la participación activa de expertos contables en el mismo.

Una revisión de la bibliografía sobre diccionarios especializados de Internet pone de manifiesto la existencia de un gran número de proyectos que solo prestan atención a algunos saberes y destrezas, por ejemplo, a las características lingüísticas de los datos. No es este el lugar adecuado para refutar el planteamiento anterior al considerarlo restrictivo (remitimos al lector interesado a Fuertes-Olivera y Tarp, 2014). Solo queremos llamar la atención de los lectores de este trabajo que nosotros hacemos una distinción importante entre el *diccionario en Internet,* es decir, el diccionario en Internet diseñado, compilado y actualizado con conceptos similares al diccionario impreso y el *diccionario de Internet*, es decir, el diccionario en Internet diseñado, compilado y actualizado con conceptos *ex novo* que hacen uso de las tecnologías de Internet existentes (ver Fuertes-Olivera y Bergenholtz, 2011 [2013]; Granger y Paquot, 2012; Fuertes-Olivera y Tarp, 2014).

En esta comunicación nos centraremos en los diccionarios de Internet, especialmente en los diccionarios denominados *Model T Ford* en la tipología de Tarp (2011).

Los diccionarios *Model T Ford* son aquellos que además de tener en cuenta los elementos generales y básicos de todos los diccionarios utilizan las tecnologías de Internet para favorecer la individualización de las búsquedas, normalmente permitiendo el empleo de búsquedas maximizadoras o minimizadoras, la recuperación de artículos de diccionarios dinámicos con datos dinámicos (es decir, el usuario puede recuperar datos diferentes según el botón de búsqueda que emplee, que normalmente está pensado para ofrecer los datos que los lexicógrafos consideran adecuados para una situación de uso concreta) y el uso de motores de búsqueda alternativos. El trabajo lexicográfico con este tipo de diccionarios exige el dominio de ciertas destrezas y saberes, como son los que nos han permitido construir los *accounting dictionaries* o *diccionarios de contabilidad*. Por razones de espacio, nos centraremos únicamente en un aspecto relacionado con las industrias de la lengua y el cambio social, que es la temática de este Congreso: el trabajo en la nube.

2. LOS DICCIONARIOS DE CONTABILIDAD

Fuertes-Olivera y Nielsen (2012) y Tarp (2012) expresan una opinión poco satisfactoria con el desarrollo experimentado por la lexicografía especializada en los últimos años (Alonso y otros, 2011). Fuertes-Olivera y Nielsen (2012) y Fuertes-Olivera y Tarp (2014) describen la lexicografía especializada como la rama de la lexicografía que se ocupa de la teoría y práctica de los diccionarios, enciclopedias, bases de datos terminológicas, bases de conocimiento especializado, glosarios y cualesquiera otras herramientas de información que cubren áreas alejadas de la cultura general, el conocimiento general y la lengua general. Por ello señalan que no es objeto de estudio de la lexicografía especializada todos aquellos temas – por ejemplo, la multidimensionalidad de los términos, su naturaleza metafórica y su variación – que no aportan conocimientos necesarios para poder construir un diccionario especializado en el sentido comentado anteriormente.

El enfoque que defendemos en el marco de la Teoría Funcional de la Lexicografía nos ha llevado a construir el *Diccionario de Contabilidad* pensando en las funciones que debe tener para satisfacer las necesidades de tres tipos preferentes de usuarios con necesidades comunicativas y cognitivas en el campo de la contabilidad en danés, español e inglés. La confección de este diccionario ha tenido dos consecuencias importantes. La primera de ellas es la puesta en práctica de una herramienta de información especializada que adopta una orientación transformativa de la lexicografía; o dicho de otra forma, una orientación en la que los conceptos lexicográficos se van adaptando continuamente a las necesidades detectadas en los usuarios de los diccionarios (ver Fuertes-Olivera y otros, 2013; Tarp, 2008).

La segunda consecuencia es el desarrollo y puesta en práctica de ideas que hacen referencia al concepto *holístico* del trabajo lexicográfico, es decir, el diccionario es el resultado de un conjunto de saberes y destrezas y no solo de una parte de ellas; por ejemplo, no es muy adecuado analizar solo cómo deben ser los datos lexicográficos sin prestar atención a las rutas de acceso, a la tecnología disponible o a la mejor forma de presentar los datos. Es frecuente asistir a presentaciones de proyectos lexicográficos especializados que mencionan con frecuencia la existencia de uno o varios corpus para extraer los datos lexicográficos (Alonso y otros, 2011). Sin poner en

duda la utilidad del corpus para el trabajo lexicográfico debemos preguntarnos, por ejemplo, si es posible fiarnos de un corpus sin contar con la opinión de un experto, o si es necesario un corpus o para qué es necesario un corpus, etc. En resumen, el trabajo lexicográfico exige una visión global del mismo y una tecnología adecuada para llevarla a cabo, como por ejemplo, el trabajo en la nube que nos ha permitido trabajar en los diccionarios de contabilidad en los últimos cinco años. Por razones de espacio reflejaremos a continuación solamente una parte de ese trabajo relacionado con la compilación y actualización de uno de los diccionarios de contabilidad, el *diccionario español de contabilidad: Recepción* (Fuertes-Olivera y otros, 2013), es decir, el diccionario al que puede acceder un usuario que solo necesita datos lexicográficos cuando está leyendo un texto (remitimos al lector interesado a Fuertes-Olivera y Tarp, 2014 y a Nielsen y Fuertes-Olivera, 2013 en donde se presentan los 27 diccionarios de contabilidad diseñados y compilados hasta la fecha por los equipos del Centre for Lexicography de la Universidad de Aarhus y el International Centre for Lexicography de la Universidad de Valladolid).

3. EL TRABAJO EN LA NUBE

Los *diccionarios de contabilidad* se construyen en una plataforma con estructura *centro* y *radios* (*hub* and *wheels* en inglés) (Figura 1). Esta estructura contempla la existencia de un nodo, la lengua inglesa, y ramificaciones, por ejemplo, para el español, el danés, etc. Esto quiere decir que los conceptos lexicográficos básicos del diccionario son aquellos que se han diseñado para la lengua inglesa, que es la que marca el camino lexicográfico de todos los diccionarios que se construyen o pudieran construirse en esta plataforma.

La estructura de *centro y radios* va asociada a una plataforma de trabajo con tres componentes individualizados y relacionados entre sí: una base de datos lexicográfica, un motor de búsqueda y una página de Internet. El diccionario es el conjunto de los tres componentes y en cada uno de ellos trabajan expertos con saberes y destrezas adecuados e integrados entre sí. Por ejemplo, la base de datos lexicográfica está pensada para almacenar la mayor cantidad de datos lexicográficos

posibles; sin embargo, el usuario solo accede a una parte de estos datos, aquellos que permiten los filtros y rutas de acceso presentes en el motor de búsqueda que se muestran mediante botones en la página web. Los filtros y las rutas de acceso han sido diseñados e implementados por expertos en Tecnologías de la Información de acuerdo con instrucciones lexicográficas.

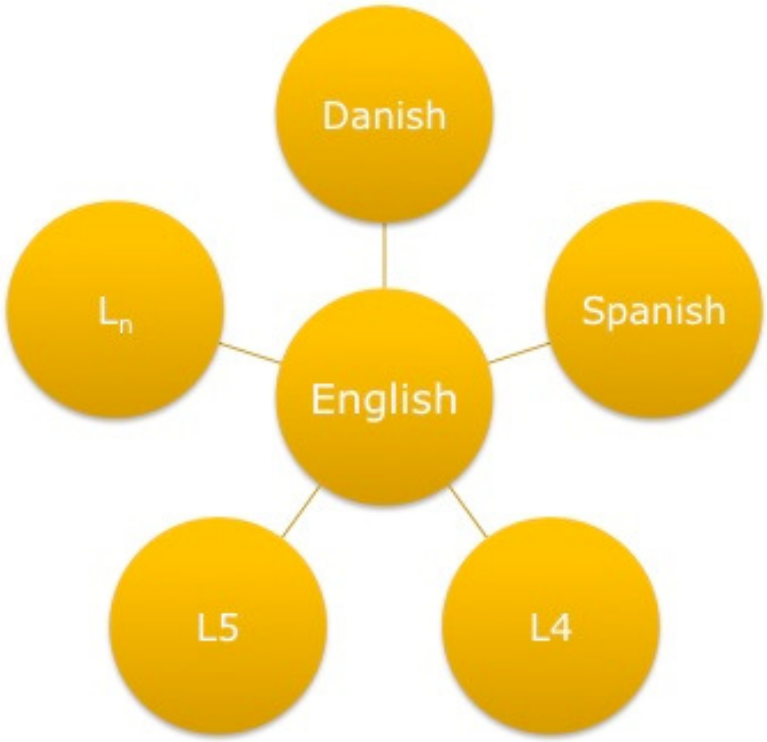

Figura 1. 1. Estructura *centro y radios* de la plataforma de los *diccionarios de contabilidad*

Como hemos dicho antes, la base de datos lexicográfica sirve para almacenar el mayor número posible de datos. Está en la nube y a ella se accede desde cualquier punto del planeta en el que haya conexión a Internet. Tiene varias pantallas o "layouts", cada una de las cuales se utiliza para almacenar datos específicos; por ejemplo, hay una pantalla para definiciones y gramática, otra para enlaces externos

e internos, etc. La extensión es enorme y no se podría hacer el trabajo que hacemos sin una herramienta de este tipo colocada en la nube. Algunos datos sobre el español corroboran esta afirmación (Tabla 1). Contiene más de 7.000 lemas españoles, 18.000 colocaciones, 4.000 ejemplos y un sinfín de datos lexicográficos como sinónimos, notas de uso, datos gramaticales, flexiones, etc.

Número de Lemas	Número de colocaciones	Número de Ejemplos	Número de referencias cruzadas	Número de notas lexicográficas
7.300	18.000	4.000	6.000	650

Tabla 1: Algunos datos numéricos en la base de datos del español relacionados con la los *diccionarios de contabilidad*

Los datos anteriores tienen todavía una mayor relevancia sabiendo que estos datos van con los correspondientes datos para el inglés y el danés; por ejemplo, las 18.000 colocaciones españolas tienen una traducción al inglés y una traducción al danés, etc. Es decir, la base de datos es un todo integrado con un buen número de pantallas o "layouts" para el trabajo específico. La Figura 2 muestra que hay tres tipos de pantallas básicas para cada lengua; por ejemplo, para la lengua nodal, el inglés, tenemos las pantallas siguientes:

gb lemma: pantalla para describir el lema inglés, su definición, gramática y aspectos flexivos;

gb_definition_A: pantalla para incluir referencias cruzadas y notas léxicas del lema inglés;

gb_definition_B: pantallas para sinónimos, antónimos, ejemplos y colocaciones del lema inglés.

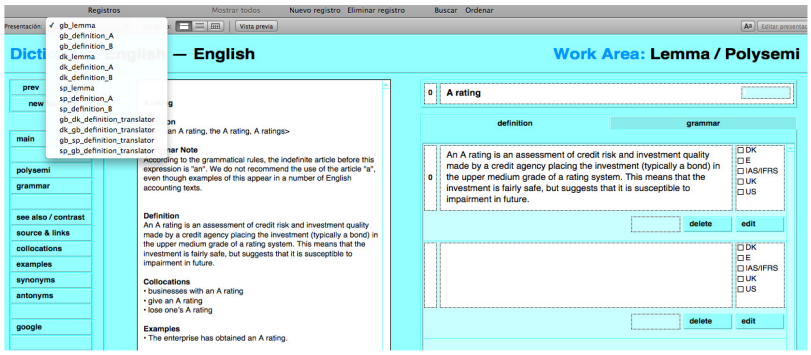

Figura 2: Layouts o Pantallas para la inclusión de datos lexicográficos.

Además, hay pantallas "translator", es decir, pantallas en las que se incluye el equivalente y se realizan las traducciones de las colocaciones y ejemplos entre las lenguas. Todo ello da una idea del tamaño enorme de la base de datos.

Finalmente, la base de datos tiene otras tres tecnologías para facilitar el trabajo en la nube: (i) una conexión directa al buscador de Google, con el que podemos obtener datos que pudieran ser adecuados para su inclusión; es decir, es una tecnología que mejora y optimiza el tiempo de búsqueda; (ii) un botón "edit" para restringir los datos a una lengua, por ejemplo, un lema es solo español y no existe en inglés o en danés; (iii) un fichero "memo" que se utiliza para la comunicación interna entre los investigadores que trabajan en el Proyecto.

4. CONCLUSIÓN

En el marco de la *Teoría Funcional de la Lexicografía,* el trabajo lexicográfico es un todo que exige saberes y destrezas generales y específicas. En esta comunicación hemos presentado brevemente el trabajo en la nube como la mejor posibilidad que tenemos para construir los diccionarios del siglo XXI. Tanto el volumen de datos necesarios para su construcción y actualización como la multidisciplinariedad de los saberes y destrezas necesarios para llevar a cabo este trabajo demandan el empleo de un sistema de trabajo flexible y actualizado; por ejemplo, un sistema que emplee recursos y tecnologías que mejoren la productividad, faciliten el contacto entre lexicógrafos que viven y trabajan a miles de kilómetros y hagan

posible el desarrollo de diccionarios que vayan en pos de la customización o individualización, que es a lo que aspiran los verdaderos diccionarios de Internet. En resumen, el trabajo que hemos señalado en esta comunicación solo puede llevarse a cabo trabajando en la nube.

AGRADECIMIENTOS

Nuestro agradecimiento al Ministerio de Economía y Competitividad y la Junta de Castilla y León por la financiación del Proyecto (Ref. FFI2011-22885 y VA067A12-1, respectivamente).

REFERENCIAS BIBLIOGRÁFICAS

Alonso, A., Millon, C., Williams, G. 2011. "Collocational Networks and their Application to an e-advanced learner's dictionary of verbs in science (DicSci)". En I. Kosem and K. Kosem (Eds.), 12-22.

Fuertes-Olivera, P. A. 2012. "On the Usability of Free Internet Dictionaries for Teaching and Learning Business English". En S. Granger and M. Paquot (Eds.), 399-424.

Fuertes-Olivera P. A., Bergenholtz, H. (Eds.) 2011. [2013]: *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London and New York: Continuum.

Fuertes-Olivera, P. A., Nielsen, S. 2012. "Online Dictionaries for Assisting Translators of LSP Texts: The *Accounting Dictionaries*", *International Journal of Lexicography*, 25(2):191-215.

Fuertes-Olivera, P. A., Bergenholtz, H., Gordo-Gómez, P., Nielsen, S., Niño-Amo, M., de los Ríos, A., Sastre-Ruano, A., Velasco-Sacristán, M. 2013. "From Theory to Practice: The Selection of Spanish Lemmas in the Accounting Dictionaries", *Fachsprache*, XXXV (1-2): 25-42.

Fuertes-Olivera, P.A., H. Bergenholtz, S. Nielsen, P. Gordo Gómez, M. Niño Amo, Á. de los Ríos Rodicio, Á. Sastre Ruano, S. Tarp, M. Velasco Sacristán. 2013. *Diccionario Español de Contabilidad. Recepción*. Base de Datos y Diseño: R. Almind and J. Skovgård Nielsen. Hamburg: Lemma.com.

Fuertes-Olivera, P. A., Tarp, S. 2014: *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*. Berlin, New York: De Gruyter.

Granger, S., Paquot, M. (Eds.) 2012. *Electronic Lexicography*. Oxford: Oxford University Press.

Kosem, I., Kosem K. (eds.) 2011. *Electronic Lexicography in the 21st Century: New Applications for New Users*. Bled: Institute for Applied Slovene Studies.

Nielsen, S., Fuertes-Olivera, P. A. 2013. "Development in Lexicography: From Polyfunctional to Monofunctional Accounting Dictionaries", *Lexikos*, 23: 323-347.

Tarp, S. 2008: *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Niemeyer.

Tarp, S. 2011. "Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction". En P. A. Fuertes-Olivera and H. Bergenholtz (Eds), 54-70.

Tarp, S. 2012. "Specialised Lexicography: 20 Years in Slow Motion", *Iberica*, 24: 117-128.