Tecnología con fines lexicográficos: su aplicación en los Diccionarios Valladolid-UVa¹

Pedro A. Fuertes-Olivera

Centro Internacional de Lexicografía y Departamento de Filología Inglesa
(Universidad de Valladolid) y Department of Afrikaans and Dutch
(Universidad de Stellenbosch)

pedro@emp.uva.es

pedro.a.fuertes@gmail.com

Marta Niño Amo

Centro Internacional de Lexicografía y Departamento de Economía Financiera
y Contabilidad (Universidad de Valladolid)

mna@emp.uva.es

Ángeles Sastre Ruano

Centro Internacional de Lexicografía y Departamento de Lengua Española
(Universidad de Valladolid)

mariaangeles.sastre@uva.es

1. Una versión más breve de este texto fue presentada como comunicación en el VIII Congreso de AELE, celebrado en Valencia en junio de 2018. El trabajo se enmarca en el Proyecto “La Teoría Funcional de la Lexicografía: Diseño y Construcción de diccionarios de Internet”, financiado por el Ministerio de Economía y Competitividad (Ref. FFI2014-52462-P). Nuestro agradecimiento también a los revisores del artículo por sus comentarios y sugerencias y a los responsables de la revista por su invitación a participar en este monográfico sobre lexicografía.

Resumen: Los *Diccionarios Valladolid-UVa* son un portal lexicográfico integrado que se está construyendo en la Universidad de Valladolid gracias a un acuerdo de colaboración y financiación entre la empresa danesa Ordbogen A/S y el Centro Internacional de Lexicografía de esta universidad. El proyecto se basa en el marco teórico de la Teoría Funcional de la Lexicografía y parte del supuesto de que los datos lexicográficos tienen valor económico en la era de internet. Para lograr maximizar el potencial económico de los mismos es necesario poner en valor lo siguiente: (a) la defensa del diccionario como herramienta de información; (b) un análisis económico del coste-beneficio de la actividad lexicográfica; (c) el empleo de una metodología de trabajo que haga posible un tratamiento de los datos que garantice la calidad de los mismos y su empleo para solucionar las necesidades de información puntual del usuario potencial de los datos; y (d) que la utilización de tecnologías que permitan poner en práctica el concepto de diccionario previamente diseñado mejoren la *productividad lexicográfica*, es decir la capacidad del lexicógrafo de preparar cada entrada lexicográfica de forma segura, fácil, y rápida. En este trabajo se analizan algunas de las tecnologías desarrolladas en el marco de este proyecto y aplicadas al diseño y construcción de los *Diccionarios Valladolid-UVa*.

Palabras clave: lexicografía de internet, teoría funcional de la lexicografía, portal lexicográfico integrado, innovación lexicográfica, tecnología con fines lexicográficos.

Technology for lexicographical purposes: Application in the *Diccionarios Valladolid-UVa*

Abstract: The *Diccionarios Valladolid-UVa* is an integrated dictionary portal which is being built at the University of Valladolid within the framework of a collaborative Project between the Danish company Ordbogen A/S and the International Centre for Lexicography of the Spanish University. Such a project is based on the tenets of the Function Theory of Lexicography and assumes that lexicographic data have economic value. This value can be maximized providing the following assumptions: (a) a dictionary is an information tool; (b) making dictionaries imply the analysis of their costs and benefits; (c) the lexicographic day-to-day work demands the use of adequate methodologies with the aim of preparing high-quality lexicographic data used for meeting users' punctual needs in different situations and through different possibilities; and (d) technologies can and should be used for putting into practice the lexicographic concept

previously designed and for improving *lexicographic productivity*, i.e. the ability of lexicographers for making every dictionary article in a very secure, easy and fast way. This article explains the working of several technologies with lexicographic applications and shows how they are being used in the design and construction of the *Diccionarios Valladolid-UVa*.

Keywords: e-lexicography, function theory of lexicography, integrated.dictionary portal, lexicographic innovation, technology for lexicographic purposes.

Recibido el 18/06/2018

Aceptado el 09/10/2018

1. Introducción

El desarrollo y consolidación de los datos como gran fuente de riqueza del siglo XXI no debe hacernos olvidar que estos datos pueden tener múltiples aplicaciones. Por ejemplo, pueden usarse para construir servicios y productos lexicográficos. Su diseño y construcción, sin embargo, permiten asegurar que, hoy por hoy, la lexicografía no ha cambiado al nivel más abstracto posible. La lexicografía sigue siendo la ciencia relacionada con la teoría y práctica de los diccionarios, es decir, de los diccionarios, enciclopedias, glosarios, lexicones, bases de datos terminológicas y cualesquiera otras herramientas de información diseñadas y creadas para solucionar los problemas de información que un usuario tipo pudiera tener en una situación tipo de la forma más rápida y fácil posible (Fuertes-Olivera, 2018a).

Antes de la consolidación de internet y su gran influencia en todos los aspectos de la vida actual, la lexicografía tenía por costumbre y tradición diferenciar entre herramientas de información centradas en una o varias lenguas, hechos y cosas y solía reservarle a cada una de ellas un nombre diferente: normalmente el de *diccionario* para las lenguas, *enciclopedia* para los hechos y *glosario* para las cosas. La consolidación de internet en la lexicografía ha puesto en duda la vigencia de este tipo de tipologías tanto desde un punto de vista teórico (por ejemplo, ¿cuál es la diferencia *para el usuario tipo* entre *palabra* y *término*?) como desde un punto de vista práctico y económico teniendo en cuenta que las posibilidades económicas de la lexicografía de internet estarán alejadas de la práctica extendida de colocar en internet un diccionario impreso con muy pocos, o incluso, ningún cambio conceptual. Por ejemplo, los diccionarios de español como el *Diccionario de la Lengua Española* (DLE), *Diccionarios.com* y *WordReference* siguen con la costumbre de utilizar abreviaturas como “f” y “s” y otras marcas similares para referirse a los datos gramaticales de un lema, como por ejemplo *leche* (Figura 1).

leche ◀

(Del lat. *lac*, *lactis* .)

1 s. f. FISIOLOGÍA Líquido blanco producido y segregado por los pechos de las hembras de los mamíferos para alimentar a sus crías, que se emplea también como alimento corriente.

2 BOTÁNICA Líquido de aspecto parecido al de la leche, segregado por algunas plantas .

EJEMPLO: *la leche de la higuera.*

Figura 1. Entrada de *leche* en *Diccionarios.com* (captura parcial)

En realidad, la práctica lexicográfica asociada a diccionarios como los que acabamos de mencionar puede interpretarse desde una perspectiva diferente: estos diccionarios indican que hay muchas posibilidades de hacer diccionarios diferentes a los que existen actualmente en la lexicografía española. Para llevarlos a buen puerto conviene saber qué se quiere hacer, cuánto dinero cuesta lo que se quiere hacer y cómo se puede hacer. Estas preguntas están relacionadas con análisis específicos que pueden abordarse teniendo en cuenta varias perspectivas:

a) Una perspectiva teórica que nos permita concluir sobre el concepto de diccionario que queremos hacer. Tarp (2018: 246), por ejemplo, afirma que un diccionario es

a utility tool, which is conceived for consultation with the genuine purpose of meeting punctual information needs experienced by specific types of potential users in specific types of extra-lexicographical context, and which is designed to assist its users by providing manual or automatic access to lexicographically prepared data, which can either be used directly by the users in order to retrieve the required information which they can subsequently use to solve specific problems in the context where the needs originally occurred, or by a digital tool in order to make automatic corrections in texts or translations produced by the users of this tool.

La conclusión de esta perspectiva es que debemos diseñar *herramientas de información* que estarán encaminadas a conseguir que el usuario de estas pueda convertir los datos lexicográficos en información de la forma más rápida y fácil posible. Por ejemplo, seguir utilizando metalenguajes lexicográficos como los que vimos en la Figura 1 es incompatible con el concepto de diccionario de internet.

b) Una perspectiva económica que nos permita asegurar que el proyecto lexicográfico podrá llegar a buen puerto en un plazo razonable de tiempo, con un coste asumible y con un modelo de negocio viable. Este dato es especialmente importante en España, donde hay un buen número de proyectos lexicográficos en marcha. Creemos que la viabilidad de la mayoría de estos proyectos es poco probable, especialmente porque son proyectos muy parciales (por ejemplo, los verbos y sus valencias) cuya financiación depende exclusivamente o fundamentalmente de convocatorias públicas y estas están cada vez más sujetas a diversos tipos de restricciones. Un análisis de los proyectos lexicográficos descritos en Bosque y Barrios Rodríguez (2018: 648-651) muestra que ninguno de estos proyectos va asociado a un análisis de su viabilidad económica. Fuertes-Olivera (2013) llevó a cabo un análisis de los proyectos lexicográficos financiados en España entre los

años 2000 y 2010 y llegó a la conclusión de que más del 90% de ellos seguían siendo proyectos, es decir, no habían logrado convertirse en una obra lexicográfica.

c) Una perspectiva relacionada con el plan metodológico y de trabajo gracias a la cual pueda asegurarse que el proyecto avanza y tiene un régimen de ejecución en tiempo constante y seguro. Sin una metodología de trabajo eficiente es prácticamente imposible completar un proyecto lexicográfico en un tiempo razonable. Al igual que con la perspectiva anterior, un proyecto 'estancado' es un proyecto fallido. No es posible ofrecer con garantías el estado en el que se encuentran los proyectos lexicográficos de internet mencionados por Bosque y Barrios Rodríguez. Por ejemplo, *VERBARIO* es un proyecto de análisis léxico del español iniciado en 2014. En la página web del proyecto pueden consultarse el análisis manual y automático de unos 215 verbos españoles. Aunque el análisis mostrado es muy interesante, el número de datos es muy pequeño, lo que nos lleva a pensar que el proyecto no ha podido avanzar por imponderables que desconocemos en este momento; quizás porque su plan metodológico deba ser revisado.

d) Una perspectiva tecnológica gracias a la cual podamos hacer uso de tecnologías y recursos con tres fines primordiales: (a) mejorar el contenido lexicográfico de cada uno de los artículos del diccionario; (b) aumentar la *productividad lexicográfica*, entendida como relación constante entre los datos lexicográficos y las necesidades del usuario tipo. Dicha relación garantizará, entre otras cosas, el empleo de buenas prácticas lexicográficas como las que nos permiten diferenciar entre cantidad y calidad lexicográfica. Por ejemplo, el aumento de los datos no mejora la productividad lexicográfica *per se*; en muchos casos este aumento de datos solo crea problemas añadidos al usuario del diccionario al que puede llevar a un proceso de sobrecarga informativa dañino para la consulta y utilidad del diccionario (Gouws y Tarp, 2018); (c) reducir los costes lexicográficos, especialmente evitando caer en la trampa que podemos llamar la *invención de la rueda lexicográfica*, que consiste en creer que un proyecto lexicográfico nuevo no puede hacer uso de otros trabajos ya publicados. Es obvio que un proyecto nuevo debe ser *innovador*, y por *innovación* se entiende normalmente o bien la mejora de algo ya existente o bien la creación de algo diferente a partir de la experiencia conocida. En términos prácticos, este proceso consiste en analizar y valorar los trabajos lexicográficos existentes sometiéndolos a un análisis crítico gracias al cual pueda concluirse la validez o no validez de estos (en todo o en parte) para el nuevo proyecto. Por ejemplo, algunos elementos de la descripción ofrecida en el proyecto *FrameNet* pueden y deben consultarse ya que ahorran el tiempo de análisis necesario para

comprender el comportamiento sintáctico, semántico, funcional, etc. de algunas unidades léxicas. Es necesario indicar, no obstante, que la posible incorporación de los datos descritos en *FrameNet* y en otros recursos de acceso libre debe ir precedida de un análisis crítico de estos con, al menos, una doble finalidad: (a) determinar qué datos pueden ser de utilidad para nuestro proyecto; (b) ofrecer una presentación de estos adecuada a nuestros usuarios potenciales. Por ejemplo, la presentación de la unidad léxica *calendric unit* (Figura 2) la hemos adaptado a las características de los usuarios de los diccionarios bilingües inglés-español/español-inglés que estamos construyendo en el marco del portal lexicográfico integrado *Diccionarios Valladolid-UVa*:

LU index

<p>Frame Index</p> <p>A B C D E F G H I J K L M N O P Q R S T U V W X Y Z</p> <p>Abandonment Abounding_with Absorb_heat Abundance Abusing Access_scenario Accompaniment Accomplishment Accoutrements Accuracy Achieving_first Active_substance Activity Activity_abandoned_state</p>	<p>Calendric_unit Lexical Unit Index</p> <p>Definition:</p> <p>Words in this frame name the different parts of the calendric cycle, both man-made and natural. The Unit (e.g. <i>Tuesday</i>) specifies some time period as part of a specific larger temporal Whole (<i>Tuesday of next week</i>), or may be resolved to an exact time span by a Relative_time (<i>next Tuesday</i>).</p> <p>Jo's mother will visit on this coming TUESDAY.</p> <p>Very frequently, the Whole or Relative_time are left out (i.e., DNI), and often should be interpreted as a time period overlapping with the communication, but more generally must be understood from context. So, for example, in the first sentence below, we can imagine that the Whole refers to a year (e.g., 1983) that was already mentioned or established in context. In the second example below, the most likely interpretation is the nearest April, whether in this year or the next, and this interpretation is best indicated by indicating the Relative_time is DNI.</p> <p>On APRIL 2, she finally found her mother. DNI.</p> <p>I'll come back in APRIL. DNI.</p>
---	---

Figura 2: La unidad léxica *Calendric_unit* en *FrameNet* (captura parcial)

Aunque las cuatro perspectivas que acabamos de mencionar están muy relacionadas entre sí, en este trabajo nos vamos a centrar en la perspectiva tecnológica: en concreto vamos a referirnos a algunos de los recursos y tecnologías usadas en el diseño y construcción de los *Diccionarios Valladolid-UVa* y su implicación para la teoría y práctica lexicográfica.

2. Portal Lexicográfico Integrado: Diccionarios Valladolid-UVa

Engelberg y Müller-Spitzer (2013: 1023) definen un portal lexicográfico (o “dictionary portal” en inglés) como una “data structure that is presented as a page or set of interlinked pages on a computer screen and provides access to a set of electronic dictionaries, and where these dictionaries can also be consulted as standalone products”. La definición anterior es demasiado genérica y por ello poco aconsejable para la lexicografía de internet, tal y como esta se ha ido desarrollando en los últimos años. Hay varios elementos esenciales que deben tenerse

en cuenta a la hora de plantear el diseño y construcción de un portal lexicográfico. Algunos de estos elementos son los siguientes:

- La estructura de datos debe crearse y organizarse con el fin de dotarla de *valor lexicográfico*, que es aquel que se consigue cuando los datos están tratados y presentados para ofrecer al usuario respuestas a sus necesidades de información; dichas respuestas deben ser lo más rápidas y fáciles de entender y manejar posible. Por ejemplo, un conjunto de concordancias que ilustran un lema concreto no puede considerarse que tenga valor lexicográfico *a priori* ya que solo algunos usuarios podrán convertir las concordancias en información. Por el contrario, una imagen, por ejemplo de un caballo, tiene valor lexicográfico en tanto en cuanto facilita la comprensión del significado del lema *caballo* a todos los usuarios potenciales. En resumen, la estructura de datos tiene que organizarse de forma que tenga *valor lexicográfico*.
- La presentación de la estructura de datos puede hacerse también en pantallas diferentes a las de un ordenador; por ejemplo, pueden presentarse en las pantallas de un teléfono móvil (Kwary, 2015). Este hecho es muy importante porque las pantallas de ordenador y las de los teléfonos móviles son muy diferentes, especialmente en lo que respecta a su tamaño, y por ello ofrecen posibilidades de presentación de los datos con valor lexicográfico muy diferentes. Una manera de solucionar esta diferencia y lograr que el tamaño del dispositivo afecte lo menos posible a la presentación de los datos es apostar por la construcción de diccionarios con valor monofuncional (Bergenholtz, 2011). Otra posibilidad, quizás más sencilla y más adecuada para el usuario tipo, sea ofrecer por defecto un número de datos reducido (por ejemplo, la definición en un diccionario monolingüe, y la definición y equivalente en un diccionario bilingüe) y dotar a la interfaz del diccionario de menús opcionales. Cada uno de estos menús contendrá indicación de las diferentes categorías con valor lexicográfico (por ejemplo, la gramática, los ejemplos, los sinónimos, la pronunciación, etc.) y solo se desplegará y se hará visible cuando el usuario así lo decida. En definitiva, el portal lexicográfico debe tener en cuenta la existencia de dispositivos digitales diferentes, cada uno con sus propias pantallas.

- * El acceso a los diccionarios no solo debe ser fácil sino que también debe ser *a la carta*, es decir, debe contemplar posibilidades diversas dependiendo de las necesidades específicas de cada usuario tipo en cada situación tipo. Tarp (2011), entre otros, se ha referido a este acceso a la carta con nombres diversos, aunque similares en lo básico: el sistema de acceso empleado debe apostar por lograr satisfacer las necesidades del usuario de la forma lo más individualizada posible. Para poder lograrlo es muy importante saber que el diccionario de internet es la suma de tres componentes individuales, cada uno de ellos con una misión específica que cumplir: (1) editor o base de datos lexicográfica: es un repositorio en el que se almacenan los datos con valor lexicográfico; es uno de los componentes fundamentales de la economía digital ya que es el lugar en el que se almacenan el mayor número de datos posibles cuyo valor económico podrá servir para financiar el proyecto e, incluso, para lograr un retorno económico a la inversión inicial; en otras palabras, en la economía digital los datos lexicográficos pueden tener valor económico más allá de su uso concreto en un diccionario específico; (2) motor de búsqueda: es la tecnología que traslada al editor la pregunta o preguntas que un usuario hace cuando quiere encontrar algo y gracias a la cual se recuperan del editor los datos objeto de búsqueda; (3) interfaz: es el componente físico que se despliega en una pantalla y que muestra los datos lexicográficos del editor recuperados por el motor de búsqueda.

No nos consta en la lexicografía del español ningún portal lexicográfico que haya superado la visión generalista de la definición de Engelberg y Müller-Spitzer. Este es el motivo por el que nos propusimos algo diferente al construir un *portal lexicográfico integrado*, que Fuertes-Olivera (2016b) ha definido como

a reference tool whose Dictionary Writing System is equipped with disruptive technologies. These allow lexicographers to store as much data as possible and users to retrieve only the data they need in specific use situations. Its articles are prepared by the same team with the basic aim of helping human and/or machine users to meet their needs in a quick and easy way. They contain both lexicographically prepared data and open linked data with lexicographic value. The lexicographic data is reusable, subject to a constant process of updating and can be used in conjunction with other tools, e.g. assistants.

De acuerdo con la definición anterior, el portal lexicográfico es una herramienta compuesta por varios dispositivos encaminados a ofrecer al usuario un diccionario de internet que, en nuestra opinión, no es ni más ni menos que un

repositorio de datos organizados y tratados de una manera determinada con el principal objetivo de ayudar a un usuario humano o máquina a solucionar sus necesidades de información de la forma más rápida y fácil posible (Fuertes-Olivera, 2018a). Todos los datos deben tener valor lexicográfico y por eso es tan importante llevar a cabo tanto un tratamiento lexicográfico adecuado a las necesidades del usuario tipo como disponer de la tecnología necesaria para hacer uso de los datos abiertos que pueden tener valor lexicográfico y para lograr una interacción fácil y rápida entre lexicógrafos y usuarios.

Los *Diccionarios Valladolid-UVa* constituyen un portal lexicográfico integrado compuesto por tres dispositivos, cada uno de ellos con tres componentes. Uno de los dispositivos del portal lexicográfico está destinado a la construcción de diccionarios de español en línea (Fuertes-Olivera y Bergenholtz 2015, 2018). La figura 3 muestra los datos almacenados en febrero de 2018 en el editor del dispositivo (los millares de la figura 3 se separan con una coma en vez de con un punto, como es habitual en España; es decir, hay en el editor 126.568 lemas, 9.204 adjetivos, etc.):

▼ Estadísticas	
Lemas	126,568
- <i>clase de palabra vacío</i>	58,370
- <i>abreviatura</i>	392
- <i>adjective</i>	9,204
- <i>adverbio</i>	1,427
- <i>artículo</i>	22
- <i>conjunción</i>	34
- <i>expresión</i>	7,643
- <i>interjección</i>	162
- <i>nombre</i>	42,244

- <i>numeral</i>	39
- <i>frase</i>	1
- <i>frase, expresión</i>	1
- <i>preposition</i>	47
- <i>pronombre</i>	98
- <i>nombre propio</i>	1,616
- <i>proverbio</i>	394
- <i>verbo</i>	4,874
Sentidos	41,248
Antónimo	5,278
Sinónimo	64,036
Colocación	161,937
Ejemplos	43,065
Gramática	6,213
Formación de palabras	46,986
Enlaces	12,528

Figura 3: Datos lexicográficos almacenados en el editor para los diccionarios de español en línea

Otro de los dispositivos está destinado a la construcción de diccionarios de contabilidad (monolingües y bilingües inglés-español/español-inglés); y el tercer-

ro a diccionarios generales bilingües inglés-español/español-inglés. El dispositivo de contabilidad está en proceso de actualización y mejora y está previsto que esté de nuevo listo para el trabajo lexicográfico en unos meses. Finalmente, el dispositivo de los diccionarios bilingües se puso a disposición de los lexicógrafos a finales de octubre de 2017 y ya cuenta con un buen número de datos (Figura 4):

Lemmas	6,629
- <i>abbreviation</i>	38
- <i>adjective</i>	893
- <i>adverb</i>	453
- <i>article</i>	29
- <i>conjunction</i>	19
- <i>expression</i>	464
- <i>interjection</i>	53
- <i>noun</i>	3,618
- <i>preposition</i>	63
- <i>pronoun</i>	15
- <i>proper noun</i>	122
- <i>symbol</i>	4
- <i>verb</i>	858
» Valencies	344
Meanings	1,319
» Antonyms	302
» Synonyms	1,969
Translations	1,427
Collocation	4,391
Examples	1,581
tions	648

Figura 4: Datos lexicográficos almacenados en el editor para los diccionarios bilingües inglés-español/español-inglés

En resumen, el portal lexicográfico integrado *Diccionarios-Valladolid UVa* es un proyecto lexicográfico en marcha diseñado parcialmente en el año 2012 por Henning Bergenholtz y Pedro Fuertes Olivera e iniciado en el año 2014 en la Universidad de Valladolid. Dicho proyecto espera salir a la luz a finales del año 2018 ó en 2019. Se construye en la Universidad de Valladolid como Art. 83 firmado por el Grupo de Investigación Reconocido “Centro Internacional de Lexicografía” de la Universidad de Valladolid y la empresa danesa Ordbogen A/S. Cuenta con los siguientes recursos:

- 500.000 euros para pagar a tres o cuatro lexicógrafos por un periodo de unos seis o siete años, tiempo estimado para llevar a cabo el trabajo lexicográfico necesario para poder poner los diccionarios a la venta y atender consultas, reclamaciones y mejoras de servicio según consultas de usuarios y análisis de *log files*.
- El trabajo lexicográfico de los integrantes del “Centro Internacional de Lexicografía”. El tiempo dedicado varía desde las veinticinco horas semanales del director del proyecto a las tres o cuatro horas semanales de algunos de los integrantes españoles del GIR (el GIR tiene 6 integrantes españoles). Este trabajo no tiene coste económico en estos momentos.
- El trabajo tecnológico de los ingenieros y programadores de la empresa Ordbogen A/S. No podemos precisar las horas de trabajo de este personal; tan solo podemos indicar que el diseño y construcción del dispositivo para almacenar los diccionarios bilingües se inició en enero de 2017 y finalizó a finales de septiembre del mismo año.
- Programas, tecnologías y recursos de la empresa Ordbogen puestos a disposición del equipo lexicográfico; también tecnologías y recursos de acceso libre como “Google”, que en este proyecto desempeña un papel fundamental. Nos referiremos brevemente a algunas de estas tecnologías en la sección siguiente.

3. Tecnología y recursos utilizados en los Diccionarios Valladolid-UVa

Una revisión crítica de algunos de los trabajos publicados recientemente –Boelhouwer y otros, 2018; Dziemianko, 2018; Faaß, 2018; Fuertes-Olivera, 2016a y

2018; Fuertes-Olivera y Bergenholtz, 2011; Fuertes-Olivera y Tarp 2014; Hargraves, 2018; Horák y Rambousek, 2018; Meyer y Abel, 2018; Müller-Spitzer y otros, 2018; y Ooi, 2018— ofrece una visión bastante completa de las tecnologías y recursos con fines lexicográficos disponibles en la actualidad. Algunas de ellas, como por ejemplo las tecnologías para el diseño y construcción de corpus con fines lexicográficos, gozan de gran aceptación y son bastante conocidas por los lexicógrafos. En este trabajo no nos vamos a referir a ellas tanto por razones de espacio como por el hecho de que en los *Diccionarios Valladolid-UVa* utilizamos el concepto de corpus lexicográfico en vez del de corpus lingüístico (Fuertes-Olivera 2012).

Para nosotros un *corpus lexicográfico* es cualquier fuente documental que pueda ser utilizada en la construcción de cada uno de los artículos lexicográficos del diccionario. Desde esta perspectiva, características clásicas del corpus lingüístico como tamaño, representatividad, autenticidad, equilibrio, etc. no son tenidas en cuenta en los *Diccionarios Valladolid-UVa*. En su lugar, usamos las búsquedas con Google y utilizamos los minitextos (es decir, las 3 o 4 líneas iniciales que nos ofrece Google al buscar una o varias palabras entre comillas), como base empírica inicial (Tarp y Fuertes-Olivera, 2016). Por ejemplo, la búsqueda “cachupina” en Google recupera más de 5.700 resultados, algunos de los cuales mostramos en la Figura 5:

Patrimonio de LAREDO – Turismo Sostenible Cantabria

turismososteniblecantabria.com/patrimonio-de-laredo/ ▼

... Escalante y Alvarado Mori. Desde el siglo XVIII perteneció a la familia Vélez Cachupín, primero a D. Tomás, que fue capitán General de Nuevo México y posteriormente a Modesta Vélez Cachupín "la Cachupina". Ya en el siglo XIX, familia que albergó al conquistador Simón Bolívar en sus estancias en la Villa de Laredo.

Images tagged with #cachupina on instagram - Imgrum

www.imgrum.org/tag/cachupina ▼

Images on instagram about cachupina. Images and videos in instagram about cachupina.

LOS AMADORES 16 latin american favorites LP 1969 Artone - Amazon

<https://www.amazon.de/AMADORES-latin.../B06XCLYMSS> ▼ Traducir esta página

LOS AMADORES 16 latin american favorites LP 1969 Artone - cachupina VG++. los amadores (Künstler) Format: Vinyl. Geben Sie die erste Bewertung für diesen Artikel ab ...

Katrina - Garras y Patas

www.garrasypatas.cl/project/katrina/ ▼

Katrina es una hermosa perrita de cuatro añitos, es una rica cachupina negrita con manchas blancas. Es de tamaño mediano por lo que es ideal para casa o departamento. Es muy hermosa, es muy tierna, le encanta compartir con niños, gatos y otros amigos perros, por lo que es ideal para una gran familia. Esta amorosa ...

Figura 5: Resultados de la búsqueda “cachupina” en Google (captura parcial)

El análisis de estos minitextos ha hecho posible documentar un número de significados de los lemas *cachupina* y *cachupín* muy superior a los que contienen los diccionarios españoles existentes en la actualidad. Por ejemplo, el *DLE* y el *Diccionario del español de México* solo registran un significado de *cachupín*, *ina* (referido como *gachupín*, *ina*); sin embargo, en los *Diccionarios Valladolid-UVa* el uso de los minitextos de Google como base empírica (que, como hemos señalado, constituye nuestro corpus lexicográfico) nos ha permitido documentar 10 significados, 2 para *cachupín* y 8 para *cachupina* (por ejemplo, los minitextos 1 y 4 de la Figura 5 se refieren a los significados 1 y 5 de *cachupina* reflejados a continuación (ejemplo 1)):

Ejemplo 1: Sentidos de *cachupín* y *cachupina* en los *Diccionarios Valladolid-UVa*

cachupín:

1. denominación que los nativos americanos daban a los españoles peninsulares que habían emigrado y se habían asentado en el continente americano
2. personaje literario estereotipado que representaba a un hidalgo caricaturizado como personaje bastante prepotente

cachupina

1. denominación que los nativos americanos daban a las españolas peninsulares que habían emigrado y se habían asentado en el continente americano
2. baile tradicional de Chile, de origen incierto; pertenece a los denominados “bailes de la tierra”
3. arbusto perenne de la familia de las Malváceas, destaca por su grandes y vistosas flores de variados colores, dependiendo de la especie, sus hojas son de un verde brillante, es originario de Asia central aunque se ha adaptado perfectamente a otras zonas del mundo
4. flor de tamaño medio y vistosos colores dependiendo de la especie que da el arbusto perenne de la familia de las Malváceas, de hojas verde brillante y originario de Asia central aunque se ha adaptado perfectamente a otras zonas del mundo

5. denominación cariñosa que se da a las perras
6. forma infantil o muy cursi de llamarle al órgano sexual femenino
7. prenda médica utilizada en centros de salud mental que consiste en una camisa abrochada a la espalda con mangas largas y cerradas que se anudan forzando que el paciente cruce los brazos sobre su pecho; se utiliza para inmovilizar a los pacientes e impedir que se hagan daño a sí mismos o a los demás
8. cinturón o peto normalmente realizado en cuero que tiene una especie de compartimentos para poder guardar los cartuchos cuando se va de caza

Todos los significados de todas las entradas documentados con los minitextos van acompañados de otros elementos textuales y/o pictóricos que demuestran la existencia del significado, incluso cuando dicho significado puede proceder de fuentes engañosas (por ejemplo, de “fake news”). Este hecho nos ha llevado a una reflexión sobre qué significados debe registrar un diccionario. La decisión tomada está basada en la naturaleza de la lexicografía antes explicada. Un diccionario es una herramienta de información y como tal debe registrar todos los elementos que puedan ayudar al usuario potencial sin tener en cuenta la veracidad factual, ideología, gustos o tradiciones sociales existentes en un momento dado. Desde este punto de vista nos parecen poco acertadas las discusiones y comentarios aparecidos en la prensa española sobre los significados de “mujer fatal” o de “gitano” registrados en el *DLE*. En nuestra opinión, un diccionario simplemente registra los significados que pueden documentarse y dejará de registrarlos cuando no pueda documentarlos. En otras palabras, los minitextos son la base empírica de los *Diccionarios Valladolid-UVa* y como tal sirven de base a la creación de definiciones y búsqueda de otros datos lexicográficos incluidos en cada artículo con el único propósito de ayudar al usuario a solucionar el problema que motivó la consulta. Por ejemplo, esta es parte de la información léxica y pictórica del significado 8 de *cachupina* (Ejemplo 2), que justifica el significado “cinturón o peto normalmente realizado en cuero que tiene una especie de compartimentos para poder guardar los cartuchos cuando se va de caza”.

Ejemplo 2: Información léxica y pictórica de cachupina (datos lexicográficos parciales)

- ✦ **Sinónimos:** canana y cartuchera

- **Ejemplos:** ajustarse la cachupina a la cintura; colocarse la cachupina; preparar la cachupina para el día de caza
- **Imagen:**

El uso de minitextos genera un volumen de datos para analizar y preparar el artículo lexicográfico tan elevado que solo se puede manejar de forma razonable incorporando la tarea “almacenaje” al proceso lexicográfico. Fuertes-Olivera (2015:73) define el concepto de *almacenaje lexicográfico* (*lexicographical storing* en inglés) como “a set of decisions made by lexicographers, IT programmers and users. These decisions are concerned with both creating one or more storage system(s), e.g. a lexicographical database, and working with them with the main aim of saving data in electronic form for lexicographic purposes”.

Fuertes-Olivera (2015) justifica el valor de esta tarea en la cantidad y tipología de los datos almacenados. Por regla general, hay en cada artículo lexicográfico de los *Diccionarios Valladolid-UVa* alrededor de 60 veces más palabras que en los diccionarios actuales. Además hay también datos digitales, normalmente hipervínculos a FLICKR (imágenes), YouTube (imágenes), *Wikipedia* (textos), conjugadores de verbos, etc. También afirma que el almacenaje está relacionado con respuestas a preguntas como estas:

- ¿dónde se pueden almacenar los datos?
- ¿quién puede almacenar los datos?

- ¿cómo pueden almacenarse los datos?
- ¿qué tipos de datos pueden considerarse datos lexicográficos?
- ¿cuándo se pueden almacenar los datos?

Las respuestas a las preguntas anteriores exigen la utilización de otras tres tecnologías imprescindibles: (a) computación en la nube; (b) diseño y construcción de editores propios; (c) sistemas de búsqueda “a la carta”. La computación en la nube permite simplificar las tareas lexicográficas, explotar adecuadamente diversos programas y agilizar el trabajo lexicográfico. Sin la computación en la nube no sería posible construir diccionarios en los que trabajan simultáneamente más de 10 personas, localizadas en lugares diferentes (algunos en España (Valladolid y Madrid); algunos en Dinamarca (Odense y Aarhus)) y con obligaciones concretas, como hacer un trabajo concreto en una fecha concreta y en un lugar específico de cada entrada lexicográfica. Gracias a la computación en la nube puede hacerse uso de servicios informáticos sin ser expertos en este tipo de servicio y desde cualquier dispositivo con acceso a internet. Además, se garantiza la inviolabilidad y seguridad de los datos lexicográficos y la posibilidad de estar siempre a la última. Para hacer los *Diccionarios Valladolid-UVa* la empresa Ordbogen A/S nos transfiere el software como servicio, la plataforma como servicio y la infraestructura como servicio. Todos estos servicios se materializan en un editor propio diseñado por lexicógrafos y especialistas en TIC, bases de datos, internet, etc. gracias al cual el trabajo lexicógrafo avanza con bastante rapidez minimizando, además, los posibles errores (tipográficos; descuidos; olvidos, etc.) que pudiera haber. Vemos, por ejemplo, en la Figura 6 una imagen del editor lexicográfico en el que se almacenan los datos del español general. Este editor tiene varios colores y varias partes. En la parte “Lema” se almacena la parte más técnica de cada lema, por ejemplo, su categoría gramatical, flexión, conjugación del verbo, notas gramaticales, etc. Esta parte es responsabilidad directa del editor del diccionario, que decide, entre otras cosas, los lemas que incluye el diccionario, su categoría gramatical y la posibilidad de incluirlo o no en los diversos “diccionarios” o soluciones monofuncionales ofrecidas a los usuarios. En la parte “Sentido” trabaja el resto del equipo; los lexicógrafos contratados completan cada sentido de cada entrada y los miembros del Centro Internacional de Lexicografía revisan y corrigen el trabajo realizado por los lexicógrafos y llevan a cabo otras tareas específicas, como por ejemplo, preparar definiciones que tienen que ver con campos de conocimiento específico.

Overstigt

Generelt

Værktøjer

Sprogcenter

Mail

Español / Español

Buscar lema / Cr...

Lema

0. echar verb

+ Nuevo lema

Sentido

1. hacer, de manera consciente, que algo vaya a parar y termine en a...

2. emitir algo olor por sí mismo, por tener algo un olor intrínsec...

3. hacer, de manera consciente, que algo caiga, se deposite o acabe ..

4. hacer salir a alguien de un lugar, de manera violenta o contunden...

5. despedir a alguien de un trabajo, quitándole los derechos que di..

6. privarle a alguien de los privilegios de ...

Lema

Prioridad 0

Lema echar

Clase de palabra verbo

☐ Publicado

Referencia

Conjugación del verbo http://www.wordreference.com/conj/EsVerbs.aspx?v=echar

Nota de la gramática

Inflexiones

echo, eché, echaba, echaré, echaría, eche, echado, echando

no inflexión

echo eché echaba echaré

echaría eche echado echando

Inflexiones no recomendables

no inflexión

Gramática +

Gramática

Borrar

Estado Terminado

Asignar usuario Pedro A. Fuertes Olivera

☐ Historial

Sentido

Número 1

Definición hacer, de manera consciente, que algo vaya a parar y termine en algún sitio o lugar, dándole impulso a la acción

Tipo

Estilo neutral

Nota del sinónimo

Nota de uso

Nota lexical

Nota

Borrar

Antónimo +

Sinónimo +

Colocaciones +

Ejemplos +

Formación de palabras +

Enlaces +

Antónimo

Estilo neutral

Borrar

Figura 6: Pantallas del editor para español general (captura parcial)

Además de los diversos campos, el editor va equipado con tecnologías para el control del trabajo. Por ejemplo, el estado “Terminado”, que indica que el trabajo ya está finalizado y el lema está listo para ser mostrado; el botón “Publicado”, que permite al editor del diccionario publicar o no una entrada concreta (por ejemplo, puede eliminarse un significado normal en un diccionario para niños); el botón “Historial”, que registra la historia de cada lema desde su fecha de creación hasta la fecha de finalización, así como el nombre de las personas que han trabajado en cada fase del trabajo lexicográfico. Igualmente ofrece registros mensuales del trabajo finalizado por cada lexicógrafo, algo que facilita, por ejemplo, el control salarial de cada uno de los lexicógrafos con contrato laboral que trabajan en el proyecto (Figura 7). Por último, el editor “avisa” de cualquier modificación producida señalando el lugar exacto en donde se ha producido y no permitiendo guardarla hasta que la modificación ha sido aceptada o rechazada. Por ejemplo, la expresión en rojo “Has cambiado algo – acordar de guardar” y el color marrón en el Texto del Enlace señalan la modificación exacta que debe controlarse antes de dar por finalizado el trabajo de esta entrada (Figura 8).

Figura 7: Estadísticas del trabajo finalizado entre el 2 y el 6 de febrero de 2018

Guardar

Has cambiado algo - acordar de guardar

Idioma del editor
Español

Volver a la lista de ediciones

Estadísticas

Gramática

Borrar

Sentido

Número	Definición	Tipo	Estilo
1	mujer que se dedica al arte de la pintura de forma artística		neutral

Nota del sinónimo

Nota de uso

Nota lexical

Nota

Borrar

Antónimo +

Sinónimo +

Colocaciones +

Ejemplos +

Formación de palabras +

Enlaces +

Enlace	Texto	Tipo
https://es.wikipedia.org/wiki/Pintor	Wikipedia	Wikipedia

Borrar

→

Figura 8: Señalización del trabajo a realizar antes de finalizar (captura parcial)

Como hemos señalado, el editor va asociado a otros dos componentes tecnológicos, que son los motores de búsqueda y la interfaz. Los lexicógrafos y los ingenieros de Ordbogen A/S han desarrollado una tecnología para conectar los datos del editor con la interfaz que hace posible búsquedas internas del sistema que tienen como misión lograr resultados basados en lo que podríamos denominar *individualización tipológica*, concepto derivado de los trabajos de Bergenholtz (2011), Almind y Nielsen (2011) y de Tarp (2011) que, en síntesis, mantiene que podemos decidir los datos específicos del editor que deben mostrarse con cada acción desarrollada por el motor de búsqueda. Por ejemplo, se han desarrollado tecnologías de búsqueda denominadas *búsqueda maximizadora* y *búsqueda minimizadora*, que van unidas a otras tecnologías necesarias para poder filtrar datos, hacer búsquedas exactas, aproximadas o booleanas, y hacer uso de hipermedia adaptativo, por ejemplo para mostrar enlaces a datos abiertos y sin derechos de autor (Bothma, 2011). Vamos a ilustrar estas tecnologías con referencia a un tipo de diccionario que no existe en la lexicografía española (al menos no conocemos ninguno). El diccionario en cuestión utiliza búsquedas maximizadoras, es decir, la búsqueda no se detiene hasta que acaba el proceso completo (en la búsqueda minimizadora, el proceso de búsqueda se detiene en cuanto el motor de búsqueda encuentra el primer ejemplo de lo que busca), que tratan de encontrar

trar en el campo “definición” conceptos iguales o similares a los buscados. Para ello, el motor de búsqueda analiza en el campo “definición” el concepto colocado en el motor con marcadores booleanos como “+”, “-”, con o sin espacio en blanco. Esta búsqueda ofrece todos los artículos del diccionario en los que aparece el concepto según este haya sido descrito para ser más preciso (normalmente con el símbolo “+” sin espacio) o menos preciso (normalmente con el símbolo “-” sin espacio). Así, si alguien quiere buscar en el editor los lemas que hacen referencia, por ejemplo, a “mujer”, “dinero”, “hombre” y “sexo” puede buscar en este diccionario (en inglés se conoce como WordFinder) colocando en el motor de búsqueda *mujer + sexo + hombre + dinero* . Esta búsqueda da como resultado un número de posibles entradas, que pueden consultarse fácilmente. La Figura 9 muestra que esta búsqueda ofrece 6 lemas potenciales, uno de los cuales es **zorrupia**:

Figura 9: Búsqueda maximizadora con marcadores booleanos (captura parcial)

Si quisiéramos buscar ‘mujeres sin ninguna relación con el sexo’ (por ejemplo, profesora o ibera) podríamos hacerlo con la búsqueda *mujer – sexo*, lo que daría como resultado un número elevado de soluciones (más de 100 en la Figura 10). Esto quiere decir que la búsqueda ha sido muy genérica, quedando en manos del usuario la posibilidad de acotarla (por ejemplo con la secuencia *mujer + sexo + hombre + dinero* que mostramos antes o cualquier otra que se estime oportuna). Esta tecnología es especialmente útil en una situación de producción, cuando el

usuario tiene alguna idea sobre lo que quiere usar pero no recuerda la palabra o no sabe cómo se escribe, etc. (Fuertes-Olivera y Leroyer 2014 ofrecen una descripción más detallada del funcionamiento de esta tecnología).

Example: woman - sex

The screenshot shows a search interface for 'Den Danske Ordfinder'. At the top, there is a search bar with the text 'mujer - sexo' and a 'Søg' button. Below the search bar, there are two filters: 'SPOF (100)' and 'Andre søgninger'. The search results section is titled 'Den Spanske Ordfinder' and shows a result for 'ibera' with a link 'Se artikel'. The description below the result states: 'mujer perteneciente a uno de los pueblos denominados iberos por los antiguos griegos y que ocuparon gran parte del territorio de la Península Ibérica, principalmente el sur, este y nordeste, entre los siglos VII y II antes de Cristo'.

Figura 10: Búsqueda maximizadora con marcadores booleanos (captura parcial)

Además de las tecnologías analizadas, los *Diccionarios Valladolid-UVa* hacen uso de tecnologías de libre acceso como Ngram Viewer, que, por ejemplo, resulta muy útil para encontrar nombres compuestos que pueden lematizarse, encontrar expresiones idiomáticas y ejemplos de uso adaptados a cada uno de los significados de cada uno de los lemas. Por razones de espacio dejaremos esta y otras tecnologías para trabajos posteriores.

4. Conclusión

Internet no ha cambiado la naturaleza de la lexicografía al nivel más abstracto posible; sigue siendo una disciplina centrada en la descripción lexicográfica de “palabras”, “hechos” y “cosas” orientada a satisfacer las necesidades puntuales de información de un usuario potencial en una situación tipo de la forma más rápida y fácil posible. Esta descripción lexicográfica suele hacerse con datos organizados y estructurados de una manera específica, normalmente para satisfacer las necesidades de información del usuario potencial del diccionario. Sin embargo, internet sí ha cambiado profundamente la manera de trabajar y ha dado valor económico a los datos, incluyendo los datos lexicográficos. Desde esta perspectiva, el portal

lexicográfico *Diccionarios Valladolid-UVa* es una respuesta lexicográfica novedosa a las posibilidades que ofrece internet para convertir los datos en información asumiendo que, si dicha conversión es fácil, segura, sencilla y adecuada, habrá usuarios potenciales, en diversas situaciones de uso, dispuestos a pagar por ella. Para llevar a cabo este supuesto son necesarios tanto recursos humanos, económicos y tecnológicos como planteamientos teóricos y metodológicos que guíen el proceso de diseño y construcción de la herramienta. Para diseñar y construir los *Diccionarios Valladolid-UVa* hemos utilizado los fundamentos de la *Teoría Funcional de la Lexicografía* (Fuertes-Olivera y Tarp, 2014; Tarp, 2008). Además, hemos trasladado las ideas a la práctica uniendo nuestras fuerzas con una de las empresas tecnológicas más avanzadas del mundo (Ordbogen A/S se dedica a las Tecnologías del Lenguaje y ha sido galardonada 6 veces seguidas con el premio *Gazzelle* a las empresas más innovadoras de Escandinavia). El resultado es un proyecto ya avanzado en su construcción que ofrece una gran cantidad de novedades, algunas de las cuales son tecnológicas y están siendo usadas para mejorar la productividad lexicográfica incrementando tanto el número de lemas y significados como reduciendo los costes de producción de los mismos y mejorando el tiempo destinado al trabajo lexicográfico. Algunas de estas tecnologías son de uso libre; otras son propias y muy útiles, como acabamos de mostrar en este trabajo.

Referencias bibliográficas

- BERGENHOLTZ, H. (2011). "Access to and Presentation of Needs-Adapted Data". En Pedro A. FUERTES-OLIVERA y H. BERGENHOLTZ (eds.) (2011), 30-53.
- BOELHOUWER, B., DYKSTRA, A. y SIJENS, H. (2018). "Dictionary Portals". En P. A. FUERTES-OLIVERA (ed.) (2018), 754-766.
- BOSQUE, I. y Barrios Rodríguez, M. A. (2018). "Spanish Lexicography in the Internet Era". En Pedro A. FUERTES-OLIVERA (ed.) (2018), 636-660.
- BOTHMA, TJD. (2011). "Filtering and Adapting Data and Information in an Online Environment in Response to User Needs". En Pedro A. FUERTES-OLIVERA y H. BERGENHOLTZ (eds.) (2011), 71-102.
- DICCIONARIOS.COM: Versión en Línea: <https://www.diccionarios.com> (consulta: 7 de mayo de 2018)
- DLE = *Diccionario de la Lengua Española*. Real Academia Española. Madrid. Versión en línea: <http://www.dle.rae.es/> (consulta: 7 de mayo de 2018).
- DZIEMIANKO, A. (2018). "Electronic Dictionaries". En P. A. FUERTES-OLIVERA (ed.), 2018: 663-683.
- ENGELBERG, J. y Müller-Spitzer, C. (2013). "Dictionary Portals". En R. H. Gouws, U. Heid, W. Schweickard and H. E. Wiegand (eds.). *Dictionaries*.

- An International Encyclopedia of Lexicography*. Berlin: De Gruyter Mouton: 1023-1035.
- FRAMENET. Versión en Línea: <https://framenet.icsi.berkeley.edu/fndrupal/> (consulta: 7 de mayo de 2018).
- FUERTES-OLIVERA, Pedro A. (2012): "Lexicography and the Internet as a Resource". *Lexicographica* 28: 49-70.
- FUERTES-OLIVERA, Pedro A. (2013): "E-lexicography: The Continuing Challenge of Applying New Technology to Dictionary Making". En H. JACKSON (ed.), *Bloomsbury Companion to Lexicography*. Londres/Nueva York/Nueva Delhi y Sidney: Bloomsbury: 323-340.
- FUERTES-OLIVERA, Pedro A. (2015). "Lexicographical Storing: A Key Lexicographical Task in the Era of the Internet". *Lexicographica* 31: 68-90.
- FUERTES-OLIVERA, Pedro A. (2016a). "A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries". *International Journal of Lexicography* 29 (2): 226-247.
- FUERTES-OLIVERA, Pedro A. (2016b). "European Lexicography in the Era of the Internet: Present Situations and Future Trends", Conferencia plenaria. Pekín, 2 de diciembre de 2016. Invitación de la editorial Commercial Press y la Chinese Association of Lexicography.
- FUERTES-OLIVERA, Pedro A. (ed.) (2018). *The Routledge Handbook of Lexicography*. Londres y Nueva York: Routledge.
- FUERTES-OLIVERA, Pedro A. (2018a). "Introduction: Lexicography in the Internet Era". En Pedro A. Fuertes-Olivera (2018): 1-15.
- FUERTES-OLIVERA, Pedro A. y BERGENHOLTZ, H. (eds.) (2011). *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. Londres y Nueva York: Continuum.
- FUERTES-OLIVERA, Pedro A. y BERGENHOLTZ, H. (2015). "Los diccionarios en línea de español" Universidad de Valladolid". *Estudios de Lexicografía* 4: 71-98.
- FUERTES-OLIVERA, Pedro A. y BERGENHOLTZ, H. (2018). "Dictionaries for Text Production", en P. A. FUERTES-OLIVERA (ed.) (2018), pp. 267-283.
- FUERTES-OLIVERA, Pedro A. y LEROYER, P. (2014). "User-generated Exploratory Search Routes: ENCONTRAR UN TÉRMINO in the Accounting Dictionaries". En J. RUPPENHOFFER y G. FAASS (eds.), *Proceedings of the 12th Edition of the Konvens Conference*, 86-95. Hildesheim: University of Hildesheim ISBN: 978-3-934105-46-1. Acceso: http://opus.bsz-bw.de/ubhi/volltexte/2014/289/pdf/konvens_proceedings.pdf

- FUERTES-OLIVERA, Pedro A. y TARP, S. (2014). *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Berlin y Boston: de Gruyter.
- GOUWS, Rufus H. y TARP, S. (2018). "Information Overload and Data Overload in Lexicography". *International Journal of Lexicography* 30 (4): 389-415.
- HARGRAVES, O. (2018). "Information Retrieval for Lexicographic Purposes". En P. A. Fuertes-Olivera (ed.) (2018), pp. 701-714.
- HORÁK, A. y RAMBOUSEK, A. (2018). "Lexicography and Natural Language Processing". En P. A. Fuertes-Olivera (ed.) (2018), pp. 179-196.
- KWARY, D. (2015). "Challenges of Online Learners's Dictionaries". *Estudios de Lexicografía* 4: 199-201.
- MEYER, C.M. y ABEL, A. (2018). "User Participation in the Internet Era". En P. A. FUERTES-OLIVERA (ed.) (2018), pp. 735-753.
- MÜLLER-SPITZER, C., KOPLINIG, A. y WOLFER, S. (2018). "Dictionary Usage Research in the Internet Era". En P. A. Fuertes-Olivera (ed.) (2018), pp. 715-734.
- NIELSEN S. y ALMIND, R. (2011). "From Data to Dictionary". En Pedro A. FUERTES-OLIVERA y H. BERGENHOLTZ (eds) (2011), pp. 141-167.
- OOI, V. (2018). "Using the Web for Lexicographic Purposes". En P. A. FUERTES-OLIVERA (ed.) (2018), pp. 684-700.
- TARP, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer Verlag.
- TARP, S. (2011). "Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction". En Pedro A. Fuertes-Olivera y H. Bergenholtz (eds) (2011), pp. 54-70.
- TARP, S. (2018). "The Concept of Dictionaey". En Pedro A. Fuertes-Olivera (ed.) (2018), pp. 237-249.
- TARP, S. y FUERTES-OLIVERA, Pedro A. (2016). "Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-UVa". *Lexikos* 26: 273-295.
- VERBARIO. SEMÁNTICA DE LOS VERBOS EN CONTEXTO. Versión en Línea: <http://www.tecling.com/verbario/> (Consulta: 7 de mayo de 2018).
- WORDREFERENCE. Versión en Línea: <http://www.wordreference.com/es/> (consulta: 7 de mayo de 2018).