Los corpus especializados en la lingüística aplicada: traducción y enseñanza

Edición de
Elisa Sartor

Pliegos Hispánicos

Serie «Intersecciones»

Pliegos Hispánicos
12

Colección
Pliegos Hispánicos

Director
Matteo De Beni (Università degli Studi di Verona)

Coordinador de la serie «Intersecciones»
Antoni Nomdedeu Rull (Universitat Rovira i Virgili)

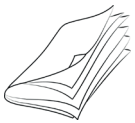
Comité científico

Carlos Alvar (Université de Genève)
Pedro Álvarez de Miranda (Universidad Autónoma de Madrid - Real Academia Española)
Paola Ambrosi (Università degli Studi di Verona)
Lisa Rose Bradford (Universidad Nacional de Mar del Plata)
Ivo Buzek (Universidad Masaryk de Brno)
Maximiano Cortés Moreno (Universidad Fujen de Taiwán)
Don W. Cruickshank (University College Dublin)
César Domínguez Prieto (Universidade de Santiago de Compostela)
Leonardo Funes (Universidad de Buenos Aires)
Natividad Gallardo San Salvador (Universidad de Granada)
Idalia García (Universidad Nacional Autónoma de México)
Cecilio Garriga Escribano (Universitat Autònoma de Barcelona)
Luigi Giuliani (Università degli Studi di Perugia)
Ana María Hernández (City University of New York)
Ilse Logie (Universiteit Gent)
César Manrique (Universidad Nacional Autónoma de México)
Alfredo Martínez Expósito (University of Queensland)
Carmen Navarro (Università degli Studi di Verona)
Veronica Orazi (Università degli Studi di Torino)
Elisabetta Paltrinieri (Università degli Studi di Torino)
Maria Grazia Profeti (Università degli Studi di Firenze)
Jesús Rubio Jiménez (Universidad de Zaragoza)
José María Santos Rovira (Universidade de Lisboa)
Mariano Siskind (University of Harvard)
María Mercedes Suárez de la Torre (Universidad Autónoma de Manizales)
Lía Schwartz † (City University of New York)
Sven Tarp (Aarhus University)
Barry Taylor (British Library)
Alexandre Veiga Rodríguez (Universidade de Santiago de Compostela)
Klaus Zimmermann (Universität Bremen)

Los corpus especializados en la lingüística aplicada: traducción y enseñanza

Edición de
Elisa Sartor

Pliegos Hispánicos



Serie «Intersecciones»
12

UNIVERSITAS STUDIORUM
EDITRICE

Volume pubblicato con il contributo
dell'Università degli Studi di Verona
Dipartimento di Lingue e Letterature Straniere

© 2021, Universitas Studiorum S.r.l. - Casa Editrice
via Sottoriva, 9 - 46100 Mantova (MN), Italy
P. IVA 02346110204
http://www.universitas-studiorum.it
info@universitas-studiorum.it

Redazione e impaginazione:
Luigi Diego Di Donna

I volumi e i saggi pubblicati nella collana scientifica *Pliegos Hispánicos*
sono sottoposti a un processo di *peer review*

Prima edizione dicembre 2021 nella collana *Pliegos Hispánicos*
Finito di stampare nel dicembre 2021

ISBN 978-88-3369-128-2

Índice

Lingüística aplicada a la traducción y la enseñanza: los corpus especializados en la investigación actual 7
Elisa Sartor

Parte 1. Corpus y traducción

Compilación semiautomática con Sketch Engine de un corpus *ad hoc* comparable bilingüe (inglés-español) de contratos de distribución exclusiva (DISTRIBUCOR) 15
María Cristina Toledo Báez – Ángeles Camero García

Extracción fraseológica basada en un corpus de fichas técnicas de impresoras 3D en español: colocaciones y locuciones 43
Ángela Luque-Giráldez – Miriam Seghiri

La terminología comercial a través de la fraseología: estudio de caso a partir de un corpus de especialidad 67
Dunia Hourani-Martín

Los corpus como herramientas para la traducción español-inglés en el sector chacinero en el siglo xxi 89
María Teresa Ortego Antón

La traducción médica (inglés-español): una propuesta de clasificación y evaluación de recursos lexicográficos en línea y en papel 113
Lucía Escribano Meseguer – Miriam Seghiri

El corpus paralelo especializado como herramienta de extracción terminológica: ejemplo de uso y aplicación en la Traducción e Interpretación en los Servicios Públicos (TISP) 159
Rebeca Busto Acedo – María del Mar Sánchez Ramos

Parte 2. Corpus y enseñanza

El uso del corpus DIACOM para la enseñanza de terminología comercial: aplicaciones y posibilidades 185
Dunia Hourani-Martín

Corpus paralelos y enseñanza del español LE/L2 a sinohablantes. Estado de la cuestión y propuestas didácticas 205
Weiqian Wei – Mar Cruz Piñol – Kris Buyse

La obtención de datos culturales mediante el uso de la web como corpus para la elaboración de productos lexicográficos de ELE 233
Antoni Nomdedeu-Rull

Lingüística aplicada a la traducción y la enseñanza: los corpus especializados en la investigación actual

Elisa Sartor
Università degli Studi di Verona

El volumen que tenemos el placer de presentar ha surgido a raíz del interés en la docencia a través de corpus, ámbito que se enmarca en el proyecto «Le Digital Humanities applicate alle Lingue e Letterature Straniere» que se está desarrollando en el Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona, gracias a la financiación otorgada por parte del Ministerio italiano de Educación y Universidades para el quinquenio 2018-2022. En el seno de este proyecto departamental, los investigadores del Área de Lengua Española trabajan en el proyecto «Patrimonio linguistico-culturale e Digital Humanities», destinado a la compilación de dos corpus de corte diacrónico, a saber, DIACOM —corpus especializado en el sector del comercio— y OccOr (cfr. De Beni y Hourani-Martín 2020, De Beni y Sartor 2021 en prensa, respectivamente). La aplicación de los resultados de nuestro trabajo al diseño y la elaboración de los currículos de lengua española, tanto en el grado como en el máster dedicados a las lenguas extranjeras para el comercio internacional, supone una gran oportunidad para la innovación docente en nuestro Departamento.

Desde luego, la explotación de los corpus en la enseñanza de las lenguas extranjeras ha conocido un notable crecimiento en los últimos años, como se ha documentado en numerosos estudios (Sinclair 2004, Cruz Piñol 2012, entre otros), en los que se da constancia del desarrollo y perfeccionamiento de las herramientas hoy en día al alcance del profesorado. Por otro lado, la lingüística de corpus ha permitido avances reseñables no solo en la formación de traductores en contextos académicos sino también en la misma práctica traductora (al respecto, véase los trabajos pioneros de Baker 1993 y Laviosa 1998, o

bien algunas entregas más recientes, por ejemplo, Gallego Hernández 2012 y Calzada Péres y Laviosa 2021).

En este volumen, por ende, se pretende ofrecer una muestra de las que, en la actualidad, son las líneas de investigación más destacadas y avanzadas acerca de la compilación o la explotación de corpus especializados en distintos ámbitos, como el comercio, el sector agroalimentario o la medicina, con un enfoque aplicado dirigido hacia la traducción y la didáctica del español como lengua extranjera. Por ello, se ha dado voz a investigadores que trabajan en este marco aplicando una pluralidad de enfoques metodológicos y herramientas de gestión y análisis de corpus, como veremos en esta breve presentación. Los capítulos que componen este libro se vertebran bajo un eje necesariamente sincrónico y se reparten en dos secciones, la primera dedicada a la compilación y explotación de los corpus en el contexto de la traducción, mientras que la segunda hace hincapié en su aprovechamiento en la enseñanza del español como lengua extranjera.

Abre la primera sección el capítulo «Compilación semiautomática con Sketch Engine de un corpus *ad hoc* comparable bilingüe (inglés-español) de contratos de distribución exclusiva (DISTRIBUCOR)» de María Cristina Toledo Báez y Ángeles Camero García. En su contribución, las autoras no solo ofrecen un acercamiento teórico a la compilación automática y semiautomática de corpus, sino que explican detenidamente el proceso de compilación haciendo hincapié en el caso concreto de un corpus *ad hoc* comparable bilingüe de textos redactados en inglés de Estados Unidos y español de la variedad peninsular procedentes de Internet. Para ello, se sirven de contratos de distribución exclusiva recopilados a través de la herramienta Sketch Engine, añadiendo, de esta manera, una fase más al protocolo de compilación establecido por Seghiri (2011 y 2017) y que consiste en las tres fases de diseño, compilación manual y análisis de la representatividad cuantitativa. Así, la innovación propuesta por María Cristina Toledo Báez y Ángeles Camero García radica justamente en ampliar

el proceso de compilación manual incorporando una fase de compilación semiautomática con Sketch Engine.

En el segundo capítulo de este volumen «Extracción fraseológica basada en un corpus de fichas técnicas de impresoras 3D en español: colocaciones y locuciones», firmado por Ángela Luque-Giráldez y Miriam Seghiri, se realiza una investigación fraseológica, con especial atención a las colocaciones y las locuciones, en un corpus virtual monolingüe de fichas técnicas de impresoras 3D, denominado ES3D-COR. En primer lugar, las dos estudiosas establecen el marco teórico a partir del cual llevarán a cabo su análisis fraseológico (Corpas Pastor 1996, Koike 2001 y Penadés Martínez 2012, entre otros); en segundo lugar, siguiendo el protocolo de compilación propuesto por Seghiri (2015 y 2017), compilan el corpus objeto de estudio, comprueban su representatividad tanto cualitativa como cuantitativa y proceden al etiquetado con TagAnt. Finalmente, gracias al programa AntConc se ejecuta la extracción de la fraseología contenida en el corpus para su posterior análisis.

La tercera contribución también está dedicada al estudio de la fraseología, esta vez con atención a las construcciones verbonominales descritas por Tabares Plasencia (2012). A partir de este marco teórico, en el capítulo titulado «La terminología comercial a través de la fraseología: estudio de caso a partir de un corpus de especialidad», Dunia Hourani-Martín presenta un análisis basado en el verbo fraseológico *comercializar* y sus derivados morfológicos; el corpus de especialidad empleado por la estudiosa es un subcorpus del DIACOM (cfr. De Beni y Hourani-Martín 2020), formado por textos académicos del ámbito del comercio redactados en Argentina, España y México entre 1990 y 2018.

El cuarto trabajo que conforma este volumen es «Los corpus como herramientas para la traducción español-inglés en el sector chacinero en el siglo xxi» de María Teresa Ortego Antón. En él, su autora se centra en las aplicaciones de los corpus como herramientas en la práctica profesional de la traducción en la combinación lingüística es-

pañol-inglés y presenta un caso de estudio sobre la traducción de las denominaciones que el chorizo recibe en relación con su forma (chorizo sarta, chorizo cular, chorizo vela, etc.). A partir del análisis de un corpus virtual paralelo bilingüe (P-GEFEM) (Ortego Antón 2019) con la ayuda de ParaConc, la autora registra las opciones de equivalentes que se emplean en la traducción. A continuación, contrasta los resultados en un corpus comparable (C-GEFEM) (Ortego Antón y Fernández Nistal 2020) empleando AntConc 3.5.7. para comprobar si en los textos originariamente redactados en lengua inglesa se utilizan los mismos equivalentes de traducción propuestos por los traductores profesionales.

Los dos capítulos que cierran la primera parte del volumen están enfocados a la traducción en el ámbito médico-sanitario. El primero de ellos, «La traducción médica (inglés-español): una propuesta de clasificación y evaluación de recursos lexicográficos en línea y en papel» de Lucía Escribano Meseguer y Miriam Seghiri, se centra en la fase documental previa a la tarea de compilación de un corpus. En él se ofrece una reseña de los recursos lexicográficos disponibles, tanto en línea como en papel, para la traducción de diferentes especialidades médicas en el ámbito sanitario español; para ello, se toman en consideración las especialidades de obstetricia y ginecología, cirugía ortopédica y traumatología, cardiología, oncología, cirugía, farmacología, botánica y fitoterapia, puesto que concentran el mayor volumen de trabajo de los traductores. Las herramientas lexicográficas, previamente seleccionadas y clasificadas, se evalúan a través de una plantilla de análisis diseñada por Escribano Meseguer (2021 en prensa) cuyo objetivo es poner a disposición del usuario, es decir, el traductor, una lista de recursos debidamente valorados según un índice de calidad (baja, media y alta).

El último capítulo, elaborado por Rebeca Busto Acedo y María del Mar Sánchez Ramos, se titula «El corpus paralelo especializado como herramienta de extracción terminológica: ejemplo de uso y aplicación en la Traducción e Interpretación en los Servicios Públicos (TISP)» y describe la compilación de un corpus paralelo bilingüe (inglés-es-

pañol) de documentos que tratan las cardiopatías congénitas, según la metodología ilustrada por Seghiri (2011), Vigier Moreno (2016) y Sánchez Ramos (2019). A partir de este corpus, las autoras realizan una extracción terminológica con el programa SynchroTerm y, empleando también la herramienta AntConc, elaboran un glosario especializado para la Traducción e Interpretación en los Servicios Públicos en el ámbito médico-sanitario.

La segunda parte de este volumen está dedicada a la explotación de los corpus para la enseñanza del español como lengua extranjera y se compone de tres contribuciones.

En la primera de ellas, «El uso del corpus DIACOM para la enseñanza de terminología comercial: aplicaciones y posibilidades», Dunia Hourani-Martín vuelve a aprovechar el subcorpus DIACOM para presentar unas actividades didácticas desarrolladas para estudiantes de grado de terminología comercial en español con un nivel B2/C1, siguiendo en el camino inaugurado en la enseñanza del francés comercial por Bonadonna (2020). Gracias a la herramienta de gestión de corpus Sketch Engine, Hourani-Martín elabora unas aplicaciones prácticas que se pueden llevar a cabo en el aula a través de las funciones que posibilitan la creación de listas de palabras, la consulta de las concordancias y las colocaciones. Su explotación, combinada con el uso de los recursos lexicográficos de la Real Academia Española, permite acotar algunos aspectos lingüísticos como las relaciones de hiperonimia e hiponimia, la estructura morfológica de los términos pluriverbales, la variación diatópica y la neología.

El segundo capítulo, firmado por Weiqian Wei, Mar Cruz Piñol y Kris Buyse y titulado «Corpus paralelos y enseñanza del español LE/L2 a sinohablantes. Estado de la cuestión y propuestas didácticas», parte de la noción de «corpuscionario» (Alonso Ramos 2009) para explorar los corpus paralelos (chino-español) actualmente disponibles, centrándose en cuatro de ellos: *Glosbe*, *Eshelper*, *ReversoContext* y *CPEIC* (*Corpus Paralelo de Español, Inglés y Chino*). Tras su valoración cualitativa, los autores describen potenciales actividades didácti-

cas para la enseñanza del español a estudiantes sinohablantes basadas en estos corpus y dirigidas al aprendizaje de locuciones y marcadores del discurso.

Y, finalmente, «La obtención de datos culturales mediante el uso de la web como corpus para la elaboración de productos lexicográficos de ELE» de Antoni Nomdedeu-Rull cierra la sección dedicada a la explotación de corpus para la enseñanza. En su contribución, el autor explora la posibilidad de seleccionar datos culturales finalizados a la creación de diccionarios para aprendices de español L2 directamente de Internet, obviando la creación de corpus *ad hoc*. Para ello, ofrece un repaso de las distintas tipologías de corpus para fines lexicográficos y, apoyándose en el marco de la *Teoría funcional de la lexicografía*, elaborada por Tarp (1992) y difundida por Bergenholtz y Tarp (1995), propone e ilustra un método de trabajo para la selección de elementos culturales por lema en diccionarios para aprendices no nativos de español.

Para ir concluyendo esta breve presentación, esperamos que este volumen pueda resultar útil a la comunidad de investigadores involucrados en el desarrollo de la lingüística de corpus y que pueda abrir caminos en la aplicación de los corpus a la traducción (y su enseñanza) y a la docencia del español como lengua extranjera.

Referencias bibliográficas

Alonso Ramos, Margarita (2009): «Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionario?», en *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus*, Murcia, AELINCO, pp. 1191-1207. En línea, <http://www.dicesp.com/app/webroot/files/file/CILC%2009.pdf> [Consulta 29/05/2021].

Baker, Mona (1993): «Corpus Linguistics and Translation Studies. Implications and Applications», en Mona Baker, Gill Francis, Elena Tognini-Bonelli (eds.), *Text and Technology: in Honour of John Sinclair*, Amsterdam, John Benjamins, pp. 233-250.

Bergenholtz, Henning; Tarp, Sven (eds.) (1995): *Manual of Specialised Lexicography*, Amsterdam, John Benjamins.

Bonadonna, Maria Francesca (2020): «L'enseignement des notions métaterminologiques en français L2: un parcours fondé sur les corpus spécialisés», *Cahiers de Lexicologie*, n. 117, pp. 11-34.

Calzada Pérez, María; Laviosa, Sara (eds.) (2021): *MonTI*, n. 13, número monográfico *Reflexión crítica en los estudios de traducción basados en corpus / CTS Spring-Cleaning: A Critical Reflection*.

Corpas Pastor, Gloria (1996): *Manual de fraseología española*, Madrid, Gredos.

Cruz Piñol, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*, Madrid, Arco/Libros.

De Beni, Matteo; Hourani-Martín, Dunia (2020): «El discurso especializado del comercio en español: diseño de un corpus para su estudio diacrónico», *Lexis*, vol. 44, n. 1, pp. 75-112.

De Beni, Matteo; Sartor, Elisa (2021, en prensa): «El corpus *OccOr*: una herramienta para el estudio léxico de textos en español sobre China (1850-1939)», *RILEX. Revista sobre investigaciones léxicas*, vol. 4, n. 2 (*Español excéntrico. Márgenes y periferias del léxico hispánico*, ed. de Matteo De Beni).

Escribano Meseguer, Lucía (2021, en prensa): «Implementación de una plantilla electrónica y puntuable para el análisis y la evaluación de glosarios, léxicos y diccionarios de Botánica», en *Actas de las IV Jornadas Doctorales organizadas por el Programa de Doctorado en Lingüística, Literatura y Traducción*, Málaga, Universidad de Málaga.

Gallego Hernández, Daniel (2012): *Traducción económica y corpus: del concepto a la concordancia*, Alicante, Universidad de Alicante.

Koike, Kazumi (2001): *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*, Madrid, Universidad de Alcalá.

Laviosa, Sara (1998): «The Corpus-Based Approach: A New Paradigm in Translation Studies», *Meta*, Special Issue, vol. 43, n. 4, pp. 474-479.

Ortego Antón, María Teresa (2019): *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*, Berlin, Peter Lang.

Ortego Antón, María Teresa; Fernández Nistal, Purificación (2020): «Estudio contrastivo de la terminología de embutidos en inglés y en español con ParaConc y tlCorpus a partir del corpus paralelo P-GEFEM y del comparable C-GEFEM», en Miriam Seghiri (ed.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*, Bern, Peter Lang, pp. 23-48.

Penadés Martínez, Inmaculada (2012): «La fraseología y su objeto de estudio», *Linred*, n. 10, pp. 1-17.

Sánchez Ramos, María del Mar (2019): «Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica», *Lebende Sprachen,* vol. 64, n. 2, pp. 269-285.

Seghiri, Miriam (2011): «Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad», *Revista de Lingüística Teórica y Aplicada*, vol. 49, n. 2, pp. 13-30.

Seghiri, Miriam (2015): «Determinación de la representatividad cuantitativa de un corpus *ad hoc* bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos / Establishing the quantitative representativiness of an E-Reader User's Guide *ad hoc* corpus (English-Spanish)», en María Teresa Sánchez Nieto (ed.), *Corpus-based Translation and Interpreting Studies: From Description to Application*, Berlin, Frank & Timme, pp. 125-146.

Seghiri, Miriam (2017): «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores», *Babel*, vol. 63 n. 1, pp. 43-64.

Sinclair, John (ed.) (2004): *How to Use Corpora in Language Teaching*, Amsterdam/Philadelphia, John Benjamins Publishing Company.

Tabares Plasencia, Encarnación (2012): «Analyse und Abgrenzung juristischer phraseologischer Einheiten im Spanischen und Deutschen und ihre Bedeutung für die Übersetzung», *Lebende Sprachen*, vol. 57, n. 2, pp. 314-328.

Tarp, Sven (1992): *Prolegomena til teknisk ordbog*, Århus, Handelshøjskolen i Århus, Spansk Institut, Tesis doctoral.

Vigier Moreno, Francisco Javier (2016): «Teaching the use of *ad hoc* corpora in the translation of legal texts into the second language», *Language and Law/Linguagem e Direito*, vol. 3, n. 1, pp. 100-113.

Compilación semiautomática con Sketch Engine de un corpus *ad hoc* comparable bilingüe (inglés-español) de contratos de distribución exclusiva (DISTRIBUCOR)*

María Cristina Toledo Báez
Universidad de Málaga

Ángeles Camero García
Universidad de Córdoba

1. Introducción

En un mundo cada vez más globalizado, el comercio internacional se convierte en un elemento clave en el desarrollo de la economía. Esta actividad ha ido ganando peso a lo largo de los años y ha ido escalando puestos dentro del PIB de cada país. De entre los datos más recientes, en 2019 destacan las cifras de Europa, con un PIB en exportaciones[1] del 15,6 %, China con un 17,43 % o Estados Unidos con un 7,68 %. Dentro de nuestro continente, la importancia de las exporta-

* El presente trabajo ha sido realizado (parcialmente) en el seno de los proyectos «NEUROTRAD: Traducción automática neuronal y paridad humano-máquina: aspectos de evaluación y posedición» (ref. B1-2020_07), «Adaptación multilingüe y multi-dominio para la optimización del sistema VIP (VIP II)» (ref. PID2020-112818GB-I00), «Sistemas de diálogo multilingües de base neuronal para el desarrollo de apps en el ámbito sanitario: el triaje (español → inglés/árabe)» (ref. UMA18-FEDERJA-067, Junta de Andalucía), «MI4ALL – Interpretación Automática Para Todos Mediante una API Basada en Aprendizaje Profundo» (CEI, Andalucía Tech), «El entorno PROFEsional del TrAductor: fiscalidad y herramientas informáticas para el ejercicio de la profesión (PROFETA)» (PIE 19-33), «POSTrad III: La virtualización de la formación de traductores en traducción automática y posedición» (Universidad de Valladolid) y de la red docente de excelencia «TACTRAD: aplicación de nuevas herramientas TAC para la enseñanza-aprendizaje de la traducción especializada» (ref. 719/2018).

1. Datos extraídos de las páginas web DatosMacro y Unión Europea: <https://datosmacro.expansion.com/comercio/exportaciones>; <https://europa.eu/european-union/about-eu/figures/economy_es>.

ciones en el PIB es especialmente relevante en Alemania (38,72 %), Italia (26,62 %), España (23,94 %), Francia (20,98 %) o Reino Unido (16,61 %). En lo que respecta a España, de acuerdo con los datos de la Secretaría de Estado de Comercio española,[2] los países a los que más exportaciones se realizaron fueron, por orden, Reino Unido, Francia, Alemania y Estados Unidos. En concreto, en 2019 las exportaciones españolas de servicios a Estados Unidos fueron de 6369 millones de euros, un 7,9 % superior a las del año anterior, siendo uno de los países cuyas cifras de comercio exterior mostraron un mayor crecimiento.

Los intercambios comerciales entre España y Estados Unidos llevan pareja la creación de contratos internacionales, los cuales han de traducirse a o desde español, inglés u otras lenguas como puedan ser catalán o vasco, por ejemplo. Dado que, además de la situación bilingüe o multilingüe, están implicados sistemas jurídicos radicalmente distintos, para las relaciones comerciales internacionales se han creado contratos atípicos que, por paradójico que parezca, no se recogen en ningún estatuto o ley (Vintimilla Saldaña 2016). Uno de estos tipos de contrato es el contrato de distribución exclusiva, del cual nos ocuparemos en el presente capítulo.

A pesar de carecer de legislación propia, el contrato de distribución exclusiva ha ido adquiriendo con el paso de los años una mayor relevancia no solo en las relaciones hispano-estadounidenses, sino también en los intercambios comerciales entre países de todo el mundo. De hecho, es tal su importancia que se ha convertido en la alternativa más frecuente a los contratos de colaboración o de agencia (Ramos Ibos 2006), los contratos mercantiles que tradicionalmente se emplean en comercio exterior.

En vista de la importancia para la economía española que supone el contrato de distribución exclusiva, en particular en sus exportaciones a Estados Unidos, consideramos que es de especial interés estudiarlo para

2. Datos disponibles en <https://www.icex.es/icex/es/navegacion-principal/todos-nuestros-servicios/informacion-de-mercados/paises/navegacion-principal/el-pais/relaciones-bilaterales/index.html?idPais=US#:~:text=Las%20exportaciones%20espa%C3%B1olas%20de%20servicios,Reino%20Unido%2C%20Francia%20y%20Alemania>.

observar sus principales características lingüísticas y traductológicas, así como para establecer paralelismos entre los contratos estadounidenses y los contratos españoles. Para ello, resulta de enorme utilidad hacer uso de un corpus virtual *ad hoc*, es decir, una colección de contratos distribución exclusiva extraída de Internet y compilada con el objetivo de proporcionar información temática y lingüística (Sánchez Gijón 2009).

Son numerosos los autores que informan de los beneficios del uso del corpus virtual (Bowker 1998; Zanettin 1998; Corpas Pastor 2008; Seghiri 2011 y 2017; Toledo Báez 2020) tanto si se trata de corpus comparables como paralelos (Corpas Pastor 2001). A pesar de las claras ventajas que ofrecen, compilar corpus virtuales resultaba una tarea tediosa años atrás cuando la compilación tenía que ser exclusivamente manual. Entre sus múltiples beneficios, tanto Internet como la imbricación de Lingüística de Corpus y de Lingüística Computacional han traído consigo nuevas formas de compilar y acceder a corpus. Son dos los principales cambios que han propiciado: por un lado, el uso de Internet y su relación con los corpus, empleando incluso la red como corpus (Gatto 2013) o como supermercado de corpus (Bernardini *et al.* 2006) y, por otro, la automatización o semiautomatización de la compilación de corpus (Esplà-Gomis y Forcada 2010) y de sus correspondientes fases, a saber, búsqueda y acceso de la documentación, descarga de datos, formato y almacenamiento (Seghiri 2011 y 2017).

En el presente estudio, continuando con el trabajo iniciado en Camero García (2019), combinaremos ambos avances al hacer uso, de una parte, de la web como supermercado de corpus y, de otra, de la compilación semiautomática de corpus gracias a la herramienta de análisis de corpus en línea Sketch Engine (Kilgarriff *et al.* 2014). El objetivo principal de nuestro trabajo es detallar la compilación semiautomática con Sketch Engine del corpus DISTRIBUCOR, un corpus de contratos de distribución exclusiva *ad hoc* comparable bilingüe con las variedades diatópicas de inglés de Estados Unidos y de español de España. En aras de alcanzar este objetivo principal, nos hemos planteado los siguientes cuatro objetivos secundarios:

1. Definir el contrato de distribución exclusiva y determinar su legislación en los dos países de las dos variedades diatópicas objeto de estudio: Estados Unidos y España.
2. Aproximarnos desde el punto de vista teórico a la compilación automática y semiautomática de corpus.
3. Presentar la herramienta Sketch Engine y detallar sus funcionalidades automática y semiautomática para la compilación de corpus.
4. Establecer un protocolo de compilación semiautomática de corpus añadiendo una fase más al protocolo propuesto por Seghiri (2011 y 2017).

2. El contrato de distribución exclusiva: definición y legislación

Suirantea Pérez (2006: 209) define el contrato de distribución exclusiva de la siguiente manera:

> El contrato de distribución es aquel contrato mercantil en virtud del cual una parte denominada «distribuidor» se compromete, de forma continuada, a vender en una zona geográfica y/o para un grupo humano determinados y en nombre e interés propios los productos que previamente ha adquirido de la otra parte, denominada «principal», asumiendo frente al cliente determinados compromisos, entre ellos, habitualmente, el de proporcionarle asistencia técnica.

En este tipo de contratos, el proveedor o fabricante pone a disposición exclusiva de un distribuidor sus productos para que pueda venderlos a terceros en un territorio y durante un tiempo determinado. Por consiguiente, el distribuidor obtiene sus beneficios de la reventa de dichos productos.

Los contratos de distribución exclusiva se caracterizan por ser contratos de tipo oneroso, en los que ambas partes se benefician económicamente; de tipo conmutativo, pues las partes conocen los riesgos y las obligaciones que asumen; *intuitu personae*, ya que se basan en la confianza mutua, y en estos las capacidades y las virtudes del distribuidor desempeñan un papel importante; de tipo consensual y autónomo (Vintimilla Saldaña 2016). Aunque, probablemente, su característica más destacable sea la bidireccionalidad (Ortega Giménez 2013), en

la que el concedente o proveedor tiene la obligación de suministrar sus productos al distribuidor, dentro de un territorio concreto, y el distribuidor se compromete a vender los productos de dicho contrato. Es importante mencionar que el distribuidor asume todos los riesgos y costes que supone la comercialización del producto, mientras que el fabricante se encarga de la fabricación del mismo.

El contrato de distribución exclusiva, también conocido como «contrato de concesión» o «contrato de suministro en exclusiva», es un contrato atípico, pues no se encuentra regulado por ninguna normativa específica. En lo que atañe a su legislación en España, a pesar de los varios intentos por crear un proyecto de ley que recogiera los aspectos fundamentales del contrato de distribución, sigue sin haber una ley que respalde completamente este tipo de contratos. Aun así, tal y como afirma Suirantea Pérez (2006), se rigen por el artículo 51 del Código de Comercio y 1255 del Código Civil. Cruz Rivero (2015: 1270) defiende que los contratos de distribución «son una categoría en la que se integran distintas modalidades contractuales atípicas, aunque parcialmente reguladas. De este modo, el contenido del contrato y las obligaciones que asumen los contratantes varían en función de lo pactado por las partes». Es decir, las relaciones entre la empresa y el distribuidor se ajustan a las condiciones estipuladas en el contrato.

En cuanto a Estados Unidos, los contratos de distribución exclusiva se rigen por las leyes antimonopolio (*Sherman Antitrust Law*). Estas leyes impiden todas aquellas actividades que limiten el comercio internacional e interestatal, aunque, para ello, es necesario que sean interpretadas a través de la *rule of reason,* donde los tribunales decidirán si las conductas llevadas a cabo por alguna de las partes son legales o no, como afirma Gajin (2011: 111) «if a practice is judged under the rule of reason, the court is supposed to examine procompetitive and anticompetitive aspects of the conduct and decide on the legality based on which of the two aspects prevails». Por lo tanto, la parte demandante deberá alegar y probar que la parte demandada ha llevado a cabo conductas anticompetitivas (Hanni 2019).

3. La compilación automática y semiautomática de corpus

Gracias a la gran ayuda que suponen los corpus tanto paralelos como comparables, como detallan Costa *et al*. (2015), desde el ámbito del Procesamiento del Lenguaje Natural y de la Lingüística Computacional se han desarrollado en los últimos años herramientas de diverso tipo que permiten compilar corpus de forma automática y aprovechar así la fuente inagotable que es la red. En lo que atañe a corpus paralelos, el objetivo principal es recabar de forma automática *bitextos* o textos paralelos (textos originales y sus correspondientes traducciones a una o varias lenguas). La herramienta pionera creada con dicho propósito fue STRAND[3] (*Structural Translation Recognition Acquiring Natural Data*), a la cual sucedieron otras como WeBiText (Désilets *et al*. 2008), Bitextor[4] (Esplà-Gomis y Forcada 2010) o ILSP-FC (Papavassiliou *et al.* 2013), entre otras.

Respecto a los corpus comparables, cabe apuntar que la mayoría de las herramientas existentes ofrecen compilación semiautomática en la que existe intervención humana en el proceso de compilación, intervención que suele implicar la filtración de dominios y la eliminación de páginas web (Gutiérrez Florido *et al.* 2013). De esta forma, la compilación semiautomática supone una clara ventaja con respecto a la compilación automática, ya que se evita el ruido documental, esto es, la presencia de información irrelevante o de escasa calidad en el corpus (Costa *et al.* 2016), de ahí que hayamos optado por esta opción en nuestro trabajo. A pesar de que se han desarrollado herramientas relevantes como Terminus,[5] Corpógrafo,[6] Babouk (De Groc 2011) e iCorpora (Costa *et al.* 2014), resulta indudable que las dos herramientas más conocidas y de uso más extendido son BootCat[7] (Baroni y Bernardini 2004) y WebBootCat (Baroni *et al.* 2006). La diferencia entre ambas es que Web-

3. Se puede acceder a la herramienta desde <http://www.umiacs.umd.edu/~resnik/strand/>.

4. Disponible en <https://sourceforge.net/p/bitextor/wiki/Home/>.

5. Accesible desde <http://terminus.iula.upf.edu//cgi-bin/terminus2.0/terminus.pl>.

6. La URL de Corpógrafo es <https://www.linguateca.pt/corpografo/>.

7. Es el acrónimo de *bootstrapping corpora and terms from the web* y se encuentra di-

BootCat se usa en línea a través de Sketch Engine y BootCat se emplea mediante la instalación de un programa; sin embargo, ambas son prácticamente idénticas en su funcionamiento, ya que se basan en la herramienta Corpus Architect.[8] Dado que en nuestro estudio emplearemos Sketch Engine, nos centraremos, por tanto, en explicar WebBootCat.

Tal y como se detalla en su manual,[9] Sketch Engine[10] es una herramienta multilingüe[11] de análisis de texto que permite, por un lado, consultar y explotar corpus ya disponibles (con un total de 517 corpus precargados)[12] y, por otro, compilar y explotar corpus propios (Toledo Báez 2020: 142-154). Se ofrecen dos posibilidades a la hora de compilar corpus:

1. Encontrar textos en la web, paso que podría formar parte tanto de la compilación automática como semiautomática de corpus.
2. Subir a Sketch Engine los textos de los que disponga ya el usuario, paso que formaría parte del proceso de compilación semiautomática de corpus.

A la hora de encontrar textos en la web, el usuario tiene la posibilidad de acotar el tema del corpus mediante tres opciones que pueden comprobarse en la Figura 1: 1) facilitando términos o lemas que definan el tema y que se denominan *seed words*; 2) proporcionando una lista de direcciones URL desde las cuales el programa se descarga de forma automática el corpus; y 3) mediante la descarga del contenido completo de una página web.

---

sponible en <http://bootcat.dipintra.it>/.

8. Consúltese la URL <https://www.sketchengine.eu/my_keywords/corpus-architect/> para conocer más detalles acerca de Corpus Architect. Esta herramienta permite la creación de corpus a través de documentos o de páginas *webs* sin la necesidad de que el usuario sea experto en compilación de corpus.

9. <https://www.sketchengine.eu/guide/>.

10. Disponible y accesible desde la URL <https://www.sketchengine.eu/>. Tal y como se detalla en su web, el proyecto ELEXIS financia el acceso a Sketch Engine de instituciones académicas desde 2018 a 2022 (<https://www.sketchengine.eu/elexis/>).

11. Sketch Engine ofrece la posibilidad de trabajar con hasta 143 lenguas, aunque no en todas ellas se dispone de las mismas funcionalidades (<https://www.sketchengine.eu/corpora-and-languages/>).

12. <https://www.sketchengine.eu/corpora-and-languages/corpus-list/>.

Cabe apuntar que resulta imprescindible especificar la lengua del corpus, ya que Sketch Engine únicamente funciona con lenguas que se puedan segmentar de forma automática. Sin embargo, esto no supone un problema, dado que, como se ha especificado anteriormente, Sketch Engine puede trabajar con hasta 143 lenguas, entre las cuales, por supuesto, se encuentran el inglés y el español, nuestras lenguas de trabajo.

Sketch Engine ofrece, asimismo, opciones avanzadas de compilación del corpus que se refieren a: 1) realizar la búsqueda a través del buscador Bing; 2) restringir el tamaño del corpus al especificar el número mínimo y máximo de archivos; 3) añadir dos listados: uno de los términos que han de aparecer en el corpus (*allowlist settings*) y otro de los términos que no deben aparecer en el corpus (*denylist setting*), que suele corresponderse con las palabras contenidas en una *stop list* o lista de palabras vacías (artículos, adverbios, preposiciones, etc.) Una vez compilado el corpus, el usuario puede añadir archivos que se haya descargado previamente mediante la compilación tradicional, lo cual le permite ampliar el corpus y agregar documentos que sean de relevancia.

En el apartado siguiente pasaremos a detallar cómo se ha empleado Sketch Engine de forma semiautomática para la compilación del corpus DISTRIBUCOR.

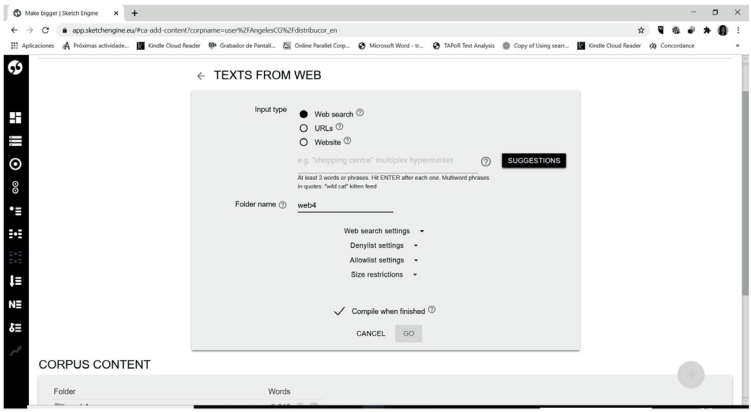


Figura 1. Opciones de búsqueda para compilar corpus con Sketch Engine.

4. El corpus DISTRIBUCOR: protocolo de compilación, características y representatividad

Como se ha especificado en la introducción, el objetivo del presente artículo es compilar un corpus *ad hoc* comparable bilingüe (inglés de Estados Unidos y español de España) de contratos de distribución exclusiva. Para ello, empleamos una metodología semiautomática de compilación basada en parte en la metodología manual de Seghiri (2017), ya que añadimos una fase más a las tres fases ya establecidas por esta autora. Concretamos a continuación las cuatro fases seguidas en la compilación del corpus DISTRIBUCOR.

4.1. Fase 1: parámetros de diseño

En la primera fase diseñaremos el corpus DISTRIBUCOR. Será un corpus *virtual*, pues lo conformarán textos descargados de Internet; *comparable*, ya que lo compondrán contratos de distribución exclusiva originales; *bilingüe*, en concreto de las variedades diatópicas de inglés de Estados Unidos y de español de España.

4.2. Fase 2: compilación manual

La segunda fase de compilación de nuestro corpus será manual y seguirá las cuatro fases propuestas por Seghiri (2017): búsqueda, descarga, formato y almacenamiento. Cada una de estas fases se explicará brevemente a continuación.

1. Búsqueda. La primera fase consiste en la localización de textos para conformar el corpus. Algunos de los textos se han obtenido a través de las páginas webs de algunas empresas, como ISDIN o H&C Animal Health, aunque la mayoría son modelos de contratos puestos a disposición del público en páginas webs como la de Cámara de Comercio o el Instituto de Comercio Exterior (ICEX). En este proceso de búsqueda se ha hecho uso de ecuaciones de búsqueda mediante los operadores booleanos *and*, *or* o *not* como, por ejemplo, exclusive distribution AND agreement o «distribución exclusiva» AND «contrato». Además, en las opciones avanzadas

del buscador se han delimitado los resultados por idiomas, en este caso inglés y español, así como también se han acotado las regiones a Estados Unidos y a España gracias a la especificación del dominio mediante *site*. En cuanto a la búsqueda de formatos, se ha preferido optar por archivos en formato .pdf, si bien se han descargado archivos en otros formatos como .doc, entre otros.

2. Descarga. La descarga de todos los contratos se ha realizado de manera manual.
3. Formato. Una vez que se han descargado y guardado todos los textos, se han convertido al formato .txt, texto plano, y se han guardado con codificación UTF-8.
4. Almacenamiento. Para almacenar los documentos se ha creado una carpeta denominada DISTRIBUCOR; esta carpeta se subdivide a su vez en otras dos, denominadas EN y ES. La subcarpeta EN contiene todos los textos en inglés y la subcarpeta ES, todos los textos en español. Cada una de estas carpetas se subdivide en otras dos, denominadas PDF y TXT. En la primera aparecen los documentos en formato .pdf y en la segunda, los documentos en texto plano .txt. Todos los contratos se han almacenado con un código específico: los textos en inglés se han guardado como EN01, EN02 y así sucesivamente hasta EN14; por su parte, los textos en español se han almacenado como ES01, ES02 hasta llegar a ES19 (véase Anexo I, cfr. Tabla 1 y Tabla 2).

El resultado de nuestra compilación manual es un corpus virtual, comparable y bilingüe (inglés-español) de contratos de distribución exclusiva integrado por 14 textos en inglés (36.655 *tokens*[13] y 2849 *types*[14]) y por 19 textos en español (43.275 *tokens* y 4366 *types*). Somos conscientes de que nuestro corpus DISTRIBUCOR es representativo a nivel cualitativo gracias a los parámetros de diseño y protocolo de compilación seguidos, pero consideramos que su ta-

13. El término *token* hace referencia al número total de palabras que hay en un texto o corpus, independientemente del número de veces que se repite.
14. El término *type* se refiere al número de distintas palabras que aparecen en el texto o corpus.

maño es demasiado reducido, motivo por el cual optamos por ampliarlo mediante la compilación semiautomática ayudándonos de WebBootCat, la herramienta de compilación de corpus integrada en Sketch Engine.

4.3. Fase 3: compilación semiautomática con Sketch Engine

El proceso de compilación semiautomática con Sketch Engine se ha llevado a cabo a través de dos de los métodos que proporciona la herramienta: realizando una búsqueda web y proporcionado listas de URL. En el primer caso, para realizar la búsqueda en la web se han empleado lemas o *seed words* con el objetivo de acotar la búsqueda. Las *seed words* empleadas en español han sido «contrato distribución», «contrato concesión», «distribución exclusiva» y «España». Por su parte, para la búsqueda de los textos en inglés se ha empleado «exclusive distribution», «distribution agreement», «US», «USA» y «United States». Para el método de proporcionar listas de URL, se han proporcionado, entre otras, las siguientes listas: <https://www.sec.gov/Archives/edgar/data/1456857/000121390018011948/f10q0318ex10-5_mjholdingsinc.htm>, <https://www.sec.gov/Archives/edgar/data/1169394/000108503705001442/laseragt.htm>, <https://www.sec.gov/Archives/edgar/data/926617/000119312511215441/dex101.htm>.

Por otro lado, dentro de las opciones de búsqueda se aumentó el número máximo de URL por búsqueda a cuarenta, así como también se amplió el número de términos de búsqueda (*seed words in search*) a cuatro. En los parámetros de búsqueda, se acotó la búsqueda introduciendo los siguientes *black words list* o *denying lists*: «non exclusive» y «no exclusivo». Una vez configurado el corpus se llevó a cabo la extracción de los textos. En este proceso se fueron revisando los textos de manera manual para comprobar que se adaptaban a los parámetros de diseño establecidos en la fase 1. Todo este mismo proceso se repitió varias veces, combinando las diferentes *seed words*, con el objetivo de obtener distintos resultados.

Una vez compilados los textos en inglés y en español de manera semiautomática, se añadieron al corpus compilado de forma manual. En consecuencia, a los catorce textos compilados en inglés manualmente habría que sumarle los veintitrés compilados semiautomáticamente. Del mismo modo, a los diecinueve textos compilados en español de manera manual, habría que sumarle los veinticinco compilados con Sketch Engine (véase Anexo II, cfr. Tabla 3 y Tabla 4).

El resultado de la combinación de la compilación manual descrita en la fase 1 junto con la compilación semiautomática con Sketch Engine descrita en la fase 3 es un corpus virtual, comparable y bilingüe (inglés-español) de contratos de distribución exclusiva integrado por 81 contratos, 44 en la variedad diatópica de español de España (113.562 *tokens* y 7879 *types*) y 37 en la variedad diatópica de inglés de Estados Unidos (122.230 *tokens* y 5107 *types*). Como hemos indicado en la fase 1, nuestro corpus DISTRIBUCOR es representativo a nivel cualitativo gracias a los parámetros de diseño y protocolo de compilación seguidos y consideramos, además, que su tamaño es adecuado para un corpus de estas características. Sin embargo, hemos de comprobar su representatividad con criterios objetivos y, para ello, nos valdremos del programa ReCor, tal y como explicaremos en el apartado siguiente.

4.4. Fase 4: análisis de la representatividad cuantitativa con ReCor

El análisis de la representatividad cuantitativa del corpus, es decir, la comprobación de que se ha cubierto la terminología básica empleada en este género (contratos de distribución exclusiva) y temática (distribución exclusiva), se ha llevado a cabo a través del programa ReCor. Este *software* permite calcular la representatividad de un corpus gracias al algoritmo N-Cor (Seghiri 2006; Corpas Pastor y Seghiri 2007). Como explica Seghiri (2011), ReCor parte de la idea de que el cociente entre las palabras reales de un texto y las totales (*types/tokens*), que da cuenta de la densidad o riqueza léxico de un texto, no aumenta proporcionalmente ni a partir de un número de textos determinado ni a partir de la densidad léxica calculada en base a secuencias de palabras o

n-gramas. De esta forma, las gráficas que arroja ReCor ilustran el punto a partir del cual el corpus en cuestión comienza a ser representativo en términos cuantitativos. En concreto, las representaciones gráficas, a partir de dos líneas que reflejan, de un lado, los documentos incluidos alfabéticamente y, de otro, aleatoriamente, que se unen y se estabilizan a medida que se aproximan al valor cero, muestran el tamaño mínimo de la colección para ser considerada representativa.

Mostramos a continuación en dos figuras (cfr. Figura 2 y Figura 3) la gráfica de representatividad del subcorpus en inglés (Figura 2) y la gráfica de representatividad del subcorpus en español (Figura 3). La gráfica de la Figura 2 muestra que el subcorpus DISTRIBUCOR inglés es representativo a partir de los 34 textos y 50.000 palabras o *tokens*. Por su parte, la gráfica de la Figura 3 indica que el subcorpus español DISTRIBUCOR comienza a ser representativo desde el punto de vista cualitativo a partir de los 43 textos y 60.000 palabras o *tokens* (Figura 3).

Así, tras asegurar la representatividad cualitativa —a través de los parámetros de diseño y el protocolo de compilación— y cuantitativa —mediante el programa ReCor— del corpus DISTRIBUCOR, podemos afirmar que nos encontramos ante un corpus representativo, fiable y de calidad, listo para ser utilizado para cualquier tipo de análisis.

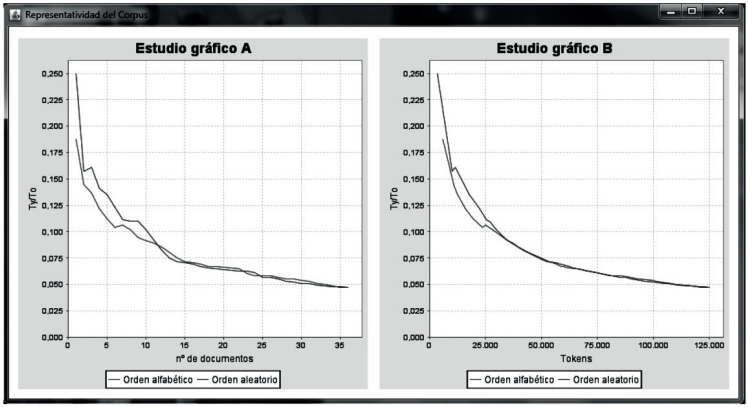


Figura 2. Representatividad del subcorpus en inglés de DISTRIBUCOR.

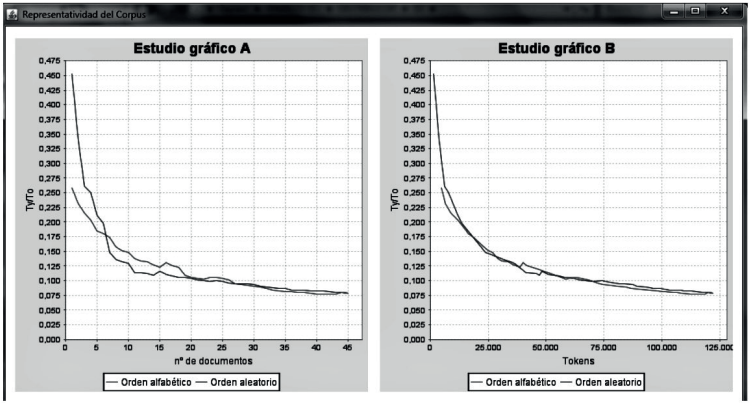


Figura 3. Representatividad del subcorpus español DISTRIBUCOR.

5. Conclusiones

Tal y como se concretó en la introducción, el objetivo principal del presente capítulo era detallar la compilación semiautomática con Sketch Engine del corpus DISTRIBUCOR, un corpus de contratos de distribución exclusiva *ad hoc* comparable bilingüe con las variedades diatópicas de inglés de Estados Unidos y de español de España. Dicho objetivo se ha cumplido a lo largo de nuestro trabajo dando como resultado un corpus virtual, comparable y bilingüe (inglés-español) de contratos de distribución exclusiva integrado por 81 contratos, 44 en la variedad diatópica de español de España (113.562 *tokens* y 7879 *types*) y 37 en la variedad diatópica de inglés de Estados Unidos (122.230 *tokens* y 5107 *types*). Además, hemos asegurado la doble representatividad del corpus DISTRIBUCOR: por un lado, la representatividad cualitativa a través de los parámetros de diseño y el protocolo de compilación y, por otro, la representatividad cuantitativa gracias al programa ReCor. En consecuencia, DISTRIBUCOR es un corpus representativo, fiable y de calidad, listo para ser utilizado para cualquier tipo de análisis del contrato de distribución exclusiva en las variedades diatópicas de inglés de Estados Unidos y de español de España.

Aparte de este objetivo principal, no podemos pasar por alto los cuatro objetivos secundarios planteados al inicio de nuestro trabajo y que también se han cumplido. El primer objetivo secundario se ha alcanzado tras definir el contrato de distribución exclusiva siguiendo a Suirantea Pérez (2006) y tras determinar la especificación de sus características y de su legislación de la mano de Vintimilla Saldaña (2016), Ortega Giménez (2013) y Cruz Rivero (2015) para los contratos españoles y de Gajin (2011) y Hanni (2019) para los contratos estadounidenses.

El segundo objetivo secundario también se ha conseguido tras detallar las herramientas utilizadas a la hora de compilar corpus tanto de forma automática como semiautomática. En lo que atañe a compilación automática de corpus paralelos destacan herramientas como STRAND, WeBiText (Désilets *et al.* 2008), Bitextor (Esplà-Gomis y Forcada 2010) e ILSP-FC (Papavassiliou *et al.* 2013). Respecto a los corpus comparables, donde es más frecuente la compilación semiautomática, sobresalen Terminus, Corpógrafo, Babouk (De Groc 2011) e iCorpora (Costa *et al.* 2014), si bien las más conocidas y empleadas son BootCat (Baroni y Bernardini 2004) y WebBootCat (Baroni *et al.* 2006). Precisamente Sketch Engine, la herramienta de análisis de texto utilizada en nuestro trabajo, emplea WebBootCat, herramienta que hemos descrito como parte del tercer objetivo secundario.

El cuarto objetivo también se ha cumplido tras haber añadido una fase más al protocolo de compilación de corpus propuesto por Seghiri (2011 y 2017): a las tres fases propuestas por esta autora, a saber, 1) parámetros de diseño, 2) compilación manual y 3) análisis de la representatividad cuantitativa, hemos añadido una fase, propuesta entre la compilación manual y el análisis de la representatividad, que consiste en la ampliación de la compilación manual gracias a la compilación semiautomática con Sketch Engine. Consideramos que añadir esta fase de forma adecuada asegura la representatividad cualitativa del corpus, si bien somos conscientes de que es muy necesario cribar y seleccionar con sumo cuidado los resultados que ofrece Sketch Engine. Estima-

mos que la Lingüística de corpus avanza y bebe de nuevas fuentes y nuevos planteamientos, de ahí que propongamos que se beneficie de nuevas herramientas como Sketch Engine, que aportan dos ventajas principales: 1) suponen un ahorro de tiempo y esfuerzo; 2) agilizan en gran medida la compilación de corpus manteniendo un alto nivel de calidad.

Los resultados que presentamos en este trabajo son solo la antesala de las múltiples aplicaciones y los múltiples usos que ofrece un corpus de las características de DISTRIBUCOR. Además de estudios contrastivos terminológicos y fraseológicos, disponer de corpus de contratos de distribución exclusiva permite llevar a cabo trabajos de investigación de corte traductológico, temático o incluso de Derecho comparado que pueden tener tanto aplicaciones didácticas (no solo en el aula de Traducción, sino también en el aula de Interpretación) como profesionales dentro del ámbito de la Traducción, la Interpretación, la Lingüística, el Derecho e incluso el comercio exterior.

Referencias bibliográficas

Baroni, Marco; Bernardini, Silvia (2004): «BootCat: Bootstrapping Corpora and Terms from the Web», en Maria Teresa Lino, Maria Francisca Xavier, Fátima Ferreira, Rute Costa, Raquel Silva (eds.), *Proceedings of LREC*, Lisboa, ELRA, pp. 1313-1316.

Baroni, Marco *et al.* (2006): «WebBootCat: Instant Domain-Specific Corpora to Support Human Translators», en André Lynum, Lena Korsnes (eds.), *Proceedings of EAMT 2006. 11th Annual Conference of the European Association for Machine Translation*, Oslo, EAMT, pp. 247-252.

Bernardini, Silvia; Baroni, Marco; Evert, Stefan (2006): «A Wacky Introduction», en Marco Baroni, Silvia Bernardini (eds.), *WaCky! Working Papers on the Web as Corpus*, Bologna, GEDIT, pp. 1-32.

Bowker, Lynne (1998): «Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study», *Meta*, vol. 43, n. 4, pp. 631-651.

Camero García, Ángeles (2019): *La traducción comentada de un contrato de distribución exclusiva y la elaboración de un corpus* ad hoc *comparable bilingüe y bidireccional (inglés-español/español-inglés)*, Universidad de Córdoba, Trabajo Fin de Máster.

Corpas Pastor, Gloria; Seghiri, Miriam (2007): «Determinación del umbral de representatividad de un corpus mediante el algoritmo N-Cor», *SEPLN: Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*, n. 39, pp. 165-172.

Corpas Pastor, Gloria (2001): «Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada», *TRANS. Revista de Traductología*, vol. 5, n. 1, pp. 155-184.

Corpas Pastor, Gloria (2008): *Investigar con corpus en traducción: los retos de un nuevo paradigma*, Frankfurt am Main, Peter Lang.

Costa, Hernani; Corpas Pastor, Gloria; Seghiri, Miriam (2014): «iCompileCorpora: A Web-based Application to Semi-automatically Compile Multilingual Comparable Corpora», en João Esteves-Ferreira, Juliet Macan, Ruslan Mitkov, Olaf-Michael Stefanov (eds.), *Translating and the Computer 36*, Genève, Éditions Tradulex, pp. 51-55.

Costa, Hernani *et al.* (2015): «iCorpora: Compiling, Managing and Exploring Multilingual Data», en Gloria Corpas Pastor, Miriam Seghiri, Rut Gutiérrez Florido, Miriam Urbano Mendaña (eds.), *Nuevos horizontes en los Estudios de Traducción e Interpretación / New Horizons in Translation and Interpreting Studies*, Genève, Tradulex, pp. 74-76.

Costa, Hernani *et al.* (2016): «Compilaçao de Corpos Comparáveis Especializados: Devemos sempre confiar nas Ferramentas de Compilaçao Semi-automáticas?», *Linguamática*, vol. 8, n. 1, pp. 3-19.

Cruz Rivero, Diego (2015): «La regulación proyectada (y fracasada) de los contratos de distribución. El proyecto de Ley de contratos de distribución y la propuesta de Código Mercantil», en María José Morillas Jarillo, María del Pilar Perales Viscasillas, Leopoldo José Porfirio Carpio (dirs.), *Estudios sobre el futuro Código Mercantil: libro homenaje al profesor Rafael Illescas Ortiz,* Getafe, Universidad Carlos III de Madrid, pp. 1262-1288. En línea, <http://hdl.handle.net/10016/20977> [Consulta: 18/11/2020].

De Groc, Clément (2011): «Babouk: Focused Web Crawling for Corpus Compilation and Automatic Terminology Extraction», en Jomi F. Hübner, Jean-Marc Petit, Einoshin Suzuki (eds.) *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Lyon, Conference Publishing Services, pp. 497-498.

Désilets, Alain *et al.* (2008): «WeBiText: Building Large Heterogeneous Translation Memories from Parallel Web Content», *Proceedings of the Thirtieth International Conference on Translating and the Computer*, 30, pp. 27-28. En línea, <https://nrc-publications.canada.ca/eng/view/ft/?id=a05f4b93-c0e8-4383-97d2-728e08e458e5> [Consulta 24/07/2021].

Esplà-Gomis, Miquel; Forcada, Mikel (2010): «Combining Content-Based and URL-Based Heuristics to Harvest Aligned Bitexts from Multilingual Sites with Bitextor», *The Prague Bulletin of Mathematical Linguistics*, vol. 93, n. 1, pp. 77-86.

Gajin, Dragan (2011): *Antitrust Aspects of Exclusive Distribution Agreements*, Budapest, Central European University, Tesis doctoral dirigida por Tibor Várady y Jules Stuyck. En línea, <http://www.etd.ceu.edu/2011/gajin_dragan.pdf> [Consulta: 20/11/2020].

Gatto, Maristella (2013): *Web as Corpus: Theory and Practice*, London/New York, Bloomsbury.

Gutiérrez Florido, Rut; Corpas Pastor, Gloria; Seghiri, Miriam (2013): «Using semi-automatic compiled corpora for medical terminology and vocabulary building in the healthcare domain», en Guadalupe Aguado de Cea, Nathalie Aussenac-Gilles, Adeline Nazarenko, Sylvie Szulman (eds.), *10th International Conference on Terminology and Artificial Intelligence (TIA'13)*, Paris, Université Paris 13-Paris Sorbonne Cité, pp. 197-200.

Hanni, Noona (2019): «Exclusive Transnational Distribution Agreements and Non-Compete Clauses in Trade Between European Union and United States», *Udayana Journal of Law and Culture,* vol. 3, n. 2, pp. 141-163. En línea, <https://ojs.unud.ac.id/index.php/UJLC/article/view/47356/30518> [Consulta: 20/11/2020].

Kilgarriff, Adam *et al*. (2014): «The Sketch Engine: ten years on», *Lexicography*, vol. 1, n. 1, pp. 7-36.

Ortega Giménez, Alfonso (2013): *Contratación internacional práctica. Cómo evitar los "riesgos contractuales" en el comercio internacional*, Madrid, Publicaciones ICEX.

Papavassiliou, Vassilis; Prokopidis, Prokopis; Thurmair, Gregor (2013): «A modular open-source focused crawler for mining monolingual and bilingual corpora from the web», en Serge Sharoff, Pierre Zweigenbaum, Reinhard Rapp (eds.), *6th Workshop on Building and Using Comparable Corpora*, Sofia, Association for Computational Linguistics, pp. 43-51.

Ramos Ibos, Teresa María (2006): «El marco legislativo comunitario de los contratos de colaboración comercial», *Cuadernos de derecho judicial*, n. 21, pp. 15-52.

Sánchez Gijón, Pilar (2009): «Developing documentation skills to build do-it-yourself corpora in the specialized translation course», en Allison Beeby, Patricia Rodríguez Inés, Pilar Sánchez Gijón (eds.), *Corpus use and translating. Corpus use for learning to translate and learning corpus use to translate,* Amsterdam, John Benjamins Publishing, pp. 109-127.

Seghiri, Miriam (2006): *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*, Málaga, Universidad de Málaga.

Seghiri, Miriam (2011): «Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad», *Revista de Lingüística Teórica y Aplicada*, vol. 49, n. 2, pp. 13-30.

Seghiri, Miriam (2017): «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores», *Babel*, vol. 63, n. 1, pp. 43-64.

Suirantea Pérez, David (2006): *Contratos mercantiles más frecuentes en la empresa*, Madrid, Grupo Difusión.

Toledo Báez, María Cristina (2020): «Sketch Engine en traducción científico-técnica (francés-español): creación y explotación del corpus *ad hoc* comparable *GeneCorp*», en Miriam Seghiri (ed.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*, Frankfurt am Main, Peter Lang, pp. 133-161.

Vintimilla Saldaña, Jaime (2016): «El contrato de distribución: de la atipicidad a la formalización», *Iuris Dictio,* vol. 15, n. 17, pp 207-230. En línea, <https://revistas.usfq.edu.ec/index.php/iurisdictio/article/view/743> [Consulta: 20/11/2020].

Zanettin, Federico (1998): «Bilingual Corpora and the Training of Translators», *Meta*, vol. 43, n. 4, pp. 616-630.

Recursos en línea

BootCat. En línea, <https://bootcat.dipintra.it> [Consulta: 05/11/2020].

Expansión/Datosmacro.com, «Exportaciones de mercancía». En línea, <https://datosmacro.expansion.com/comercio/exportaciones> [Consulta: 10/11/2020].

ICEX, «Red de oficinas económicas y comerciales de España en el exterior». En línea, <https://www.icex.es/icex/es/navegacion-principal/todos-nuestros-servicios/informacion-de-mercados/paises/navegacion-principal/el-pais/relaciones-bilaterales/index.html?idPais=US#:~:text=Las%20exportaciones%20espa%C3%B1olas%20de%20servicios,Reino%20Unido%2> [Consulta:10/11/2020].

Linguateca, «Corpógrafo». En línea, <https://www.linguateca.pt/corpografo/> [Consulta: 28/10/2020].

Sketch Engine, «Corpora by Language». En línea, <https://www.sketchengine.eu/corpora-and-languages/> [Consulta: 30/11/2020].

Sketch Engine, «Corpus Architect». En línea, <https://www.sketchengine.eu/my_keywords/corpus-architect/> [Consulta: 29/10/2020].

Sketch Engine, «List of Corpora». En línea, <https://www.sketchengine.eu/corpora-and-languages/corpus-list/> [Consulta: 30 /11/2020].

Sketch Engine, «Sketch Engine funded by ELEXIS». En línea, <https://www.sketchengine.eu/elexis/> [Consulta: 29/10/2020].

Sketch Engine, «Sketch Engine User Guide». En línea, <https://www.sketchengine.eu/guide/> [Consulta: 29/10/2020].

Sketch Engine, «Sketch Engine». En línea, <https://www.sketchengine.eu/> [Consulta: 29/10/2020].

SourceForge, «SourceForge». En línea, <https://sourceforge.net/p/bitextor/wiki/Home/> [Consulta: 28/10/2020].

STRAND, «The STRAND Bilingual Databases». En línea, <http://www.umiacs.umd.edu/~resnik/strand/> [Consulta: 1/11/2020].

Terminus 2.0. En línea, <http://terminus.iula.upf.edu//cgi-bin/terminus2.0/terminus.pl> [Consulta: 28/11/2020].
Unión Europea, «La economía». En línea, <https://europa.eu/european-union/about-eu/figures/economy_es> [Consulta: 8/11/2020].

Anexo I

Se muestran en las siguientes tablas (cfr. Tabla 1 y Tabla 2) los nombres de las páginas webs de donde se han extraído los textos de manera manual, sus correspondientes URL, la procedencia del documento y el código de almacenamiento.

Nombre de las páginas web	URL	Procedencia del texto	Código
Centric Brands Inc.	https://www.centricbrands.com/investors/secfilings/annual-reports/content/0000844143-03-000018/innoex10_94.htm?TB_iframe=true&height=auto&width=auto&preload=false	California, EE. UU.	EN01
Cohen International	http://www.rogercohen.com/distributor_agreement.shtml	Nueva York, EE. UU.	EN02
H&C Animal Health	http://bit.ly/3dFjRScHCAnimals_Distribution_Agreement	Colorado, EE. UU.	EN03
LaCount Law	http://www.lacountlaw.com/articles/international2.pdf	California, EE. UU.	EN04
LawsDocBox	https://lawsdocbox.com/Legal_Issues/66835768-Distribution-agreement-between-danik-hook-llc-and.html	California, EE. UU.	EN05
Scribd	https://es.scribd.com/document/386202117/Distribution-Agreement	California, EE.UU.	EN06
	https://es.scribd.com/document/39180367/Exclusive-Distribution-Contract-sample	(Sin especificar) EE. UU.	EN07
	https://es.scribd.com/document/121496984/International-Distribution-Agreement	Florida, EE.UU.	EN08
	https://es.scribd.com/document/1041911/Distribution-Agreement-Short-Form	(Sin especificar) EE. UU.	EN09

Shefan Solutions, LLC	http://bit.ly/3ar8mvJShefanSolutionExclusiveDistrAgreement	Pensilvania, EE. UU.	EN10
SAP&DC	http://www.sapdc.org/documents/Distribution_Agreement.pdf	Pensilvania, EE. UU.	EN11
SparkFun	https://sparkfun.com/download/PDFs/SFE_Distributor_Agreement.pdf	Colorado, EE. UU.	EN12
Tavern Craft Llc.	https://taverncraft.com/wp-content/uploads/2018/12/TC-DA-112718.6.pdf	California, EE. UU.	EN13
Vecenie's Distributing Co.	http://bit.ly/3aNPCXB_Brewery_DistrAgreement	Pensilvania, EE. UU.	EN14

Tabla 1. Nombre de las páginas web, URL, procedencia y código del subcorpus en inglés compilado de forma manual.

Nombre de las páginas web	**URL**	**Procedencia del texto**	**Código**
Boyacá	https://www.boyaca.es/assets/ccomerciales/dima/ctto_distribucion.pdf	Madrid, España	ES01
Cámara de Comercio de Sevilla	https://camaradesevilla.com/wp-content/uploads/2016/07/Distribucion_exclusiva.docx	Sevilla, España	ES02
Contacto negocio	http://www.contactonegocio.com	Oviedo, España	ES03
D & V Asesores	http://www.fiscalasesor.com/archivos/contrato-de-distribucion.html	Madrid, España	ES04
Espacio Pymes	https://espaciopymes.com/wp-content/uploads/2018/11/Contrato-de-distribucion.pdf	Madrid, España	ES05
Global Economist & Jurist	http://bit.ly/2QlECtaContratosMercantilesfrecuentesEmpresa	Madrid, España	ES06 ES07
Icex	http://bit.ly/3va99JdContratación_Internacional	Madrid, España	ES08
Inteco	http://www.inteco.es/Seguridad/Observatorio/area_juridica_seguridad/Modelos_de_Contratos_Tecnologicos/	León, España	ES09
Lefebvre	http://bit.ly/3eCDjhVContratosuministro	Madrid, España	ES10

Modelos de contratos	https://www.modelocontrato.net/wp-content/uploads/2015/04/Modelo-contrato-de-distribución-en-exclusiva.pdf	Valencia, España	ES11
Prodetur	https://www.prodetur.es/prodetur/AlfrescoFileTransferServlet?action=download&ref=164ec4bf-10bb-4f3f-a4ff-f09ae707c7b6	Sevilla, España	ES12
Reexporta	https://www.clonica.net/usuario/img_usuario/comercio-exterior.es/Herramientas_Doc/145.pdf	Barcelona, España	ES13
Revistas Wait & See	https://waitandsee.net/wp-keledra/uploads/2016/05/contrato.pdf	Madrid, España	ES14
Scribd	https://es.scribd.com/document/296292735/Modelo-Contrato-Distribucion	(Sin especificar) España	ES15
	https://es.scribd.com/document/139704425/CONTRATO-DE-DISTRIBUCION-NACIONAL	Madrid, España	ES16
	https://es.scribd.com/document/187281138/CONTRATO-DE-DISTRIBUCION-EXCLUSIVA	Navarra, España	ES17
TPV inforpyme	www.tpvinforpyme.com/DOCS/contrato_distribuidor.doc	Valencia, España	ES18
Vozelia	http://wiki.vozelia.com/lib/exe/fetch.php?media=wiki:contratodistribucion.pdf	Madrid, España	ES19

Tabla 2. Nombre de las páginas web, URL, procedencia y código del subcorpus en español compilado de forma manual.

Anexo 2

Se presentan dos tablas (cfr. Tabla 3 y Tabla 4) con los textos compilados semiautomáticamente con Sketch Engine. La primera tabla (Tabla 3) se corresponde con los textos compilados en inglés y la segunda (Tabla 4) con los textos en español.

Nombre de las páginas web	URL	Procedencia del texto	Código
Absoft	https://www.absoft.com/wpcontent/uploads/2015/09/DistributorAgreement.pdf	Michigan, EE. UU.	EN15
Biobased	http://www.biobased.us/pdf/Agriculture%20Biobased%20USA%20Agreement.pdf	Florida, EE. UU.	EN16
General Counsel, P.C.	http://www.generalcounselpc.com/docs/Software_Distribution_Agreement_(International).pdf	Virginia, EE. UU.	EN17
LB Design LLC	https://lbolineperformance.com/wpcontent/uploads/2016/05/Untitled-2.pdf	Arizona, EE. UU.	EN18
Liquor Idaho	https://liquor.idaho.gov/pdf/2020_Sample_Contract.pdf	Idaho, EE. UU.	EN19
Patent Laboratories Inc.	http://bit.ly/3dHHtpgpharmaceutical_distributionagreement	Nueva Jersey, EE. UU.	EN20
Scribd	https://es.scribd.com/document/86690560/Distribution-Agreement-Short-Form	Arizona, EE. UU.	EN21
	https://es.scribd.com/document/409265918/Distribution-Agreement	(Sin especificar) EE. UU.	EN22

U.S. Securities and Exchange Commission	https://www.sec.gov/Archives/edgar/data/1761055/000119312519062427/d717594dex995a.htm	Delaware, EE. UU.	EN23
	https://www.sec.gov/Archives/edgar/data/815094/000119312507024227/dex1024.htm	Massachussetts, EE. UU.	EN24
	https://www.sec.gov/Archives/edgar/data/1386765/000101968707001441/donmarcos_sb2-ex1003.htm	Florida, EE. UU.	EN25
	https://www.sec.gov/Archives/edgar/data/1286733/000119312504075854/dex102.htm	California	EN26
	https://www.sec.gov/Archives/edgar/data/895419/000119312503054147/dex1019.htm	Carolina del Norte, EE. UU.	EN27
	https://www.sec.gov/Archives/edgar/data/1463959/000094344012001087/dskx_ex10z15.htm	Miami, EE. UU.	EN28
	https://www.sec.gov/Archives/edgar/data/1527702/000107878212000972/s1a4_ex10z4.htm	California, EE. UU.	EN29
	https://www.sec.gov/Archives/edgar/data/1408299/000146970910000301/exhibit101distributionagreem.htm	Carolina del Sur, EE. UU.	EN30
	https://www.sec.gov/Archives/edgar/data/1330340/000119312508071128/dex101.htm	Connecticut, EE. UU.	EN31
	https://www.sec.gov/Archives/edgar/data/1456857/000121390018011948/f10q0318ex105_mjholdingsinc.htm	Delaware, EE. UU.	EN32
	https://www.sec.gov/Archives/edgar/data/1169394/000108503705001442/laseragt.htm	Texas, EE. UU.	EN33
	https://www.sec.gov/Archives/edgar/data/926617/000119312511215441/dex101.htm	Delaware, EE. UU.	EN34
	https://www.sec.gov/Archives/edgar/data/1580883/000107878217000196/f10q123116_ex10z1.htm	Nevada, EE. UU.	EN35
	https://www.sec.gov/Archives/edgar/data/1139614/000110313204000026/wlqx101.pdf	Texas, EE. UU.	EN36

Upcounsel	https://www.upcounsel.com/distribution-agreement	California, EE. UU.	EN37

Tabla 3. Nombre de las páginas web, URL, procedencia y código del subcorpus en inglés compilado de forma semiautomática con Sketch Engine.

Nombre de las páginas web	**URL**	**Procedencia del texto**	**Código**
Amazon	http://bit.ly/2QKO4WX_Amazon_Distr_Agreement	(Sin especificar) España	ES20
Aosom S.L.	https://m2.aosom.es/media/wysiwyg/2020/Contrato-Distribuidores-con-Dropshipping-Pago-Anticipado-AOSOM-ES.pdf	Barcelona	ES21
Asociación española de gerentes del golf	http://www.aeggolf.org/enlaces/servicios/informes_golf/contratos_alquiler_restaurante/Alquiler_restaurante_campo_golf_6.pdf	Tarrasa, España	ES22
Cámara de Comercio de Oviedo	https://www.camara-ovi.es/documentos/aempresarial/Modelo%20contrato%20Anexo%20IV%202012.pdf	Oviedo, España	ES23
Cámara de Comercio de Sevilla	https://camaradesevilla.com/servicios-online/modelos-tutores-documentales/modelos-contratos-internacionales/	Sevilla, España	ES24
Cámara de Sevilla	https://camaradesevilla.com/servicios-online/modelos-tutores-documentales/modelos-contratos-internacionales/compraventa-internacional-mercancias/	Sevilla, España	ES25
Comunidad de Madrid	http://www.madrid.org/contratos-publicos/1354780717734/1109266180653/1354791439270.pdf	Madrid, España	ES26
Enagás S.A.	http://bit.ly/3atdiQQ_enagas_contrato_de_distribucion	Madrid, España	ES27
Global Economist & Jurist	https://global.economistjurist.es/BDI/KS/cliente/descargar_documento.php?id=13115	Madrid, España	ES28
IMTLazarus	http://www.imtlazarus.com/wp-content/uploads/2019/01/Acuerdo-comercial-distribuidor-.pdf	Baracaldo, España	ES29

Indra Company	https://www.indracompany.com/sites/default/files/modelo_acuerdo_de_distribucion_isocloud.pdf	Madrid, España	ES30
ISDIN	https://www.isdin.com/pdf/Contrato_Distribuidor_Autorizado.pdf	(Sin especificar) España	ES31
Luque y Martín	https://www.pausaseleccion.com/Mambo/DOCUMENTOS/Contrato-Distribucion-Distribuidor-07.pdf	Madrid, España	ES32
Scribd	https://es.scribd.com/document/33054217/Contrato-de-distribucion-en-exclusiva-15-01-09	Madrid, España	ES33
	https://es.scribd.com/document/435888431/Modelo-de-Contrato--de-Distribucion-de-Vinos	(Sin especificar) España	ES34
	https://es.scribd.com/document/399048069/Acuerdo-de-Distribucion-tiposs	Sevilla, España	ES35
	https://es.scribd.com/document/276540862/DISTRIBUCION	(Sin especificar) España	ES36
	https://es.scribd.com/document/373083274/Contrato-de-Distribucion-Digital-de-Contenidos	(Sin especificar) España	ES37
	https://es.scribd.com/document/339493155/GALACTIC-RECORDS-Contrato-de-Distribucion	Valencia, España	ES38
	https://es.scribd.com/document/484283339/Distribucion-exclusiva-esp	(Sin especificar) España	ES39
	https://es.scribd.com/document/139596209/Concesion-de-distribucion-exclusiva-de-marcas-cosmeticos	(Sin especificar) España	ES40
	https://es.scribd.com/document/47112934/contrato-distribuidor-1	(Sin especificar) España	ES41
	https://es.scribd.com/document/432881754/Modelo-Contrato-de--Distribucion-en-Exclusiva	(Sin especificar) España	ES42
	https://es.scribd.com/document/262331651/Contrato-de-Distribucion	(Sin especificar) España	ES43

Supercontable	https://www.supercontable.com/informacion/ley_gestion/Contrato_de_distribucion_en_exclusiva.html	(Sin especificar) España	ES44

Tabla 4. Nombre de las páginas web, URL, procedencia y código del subcorpus en español compilado de forma semiautomática con Sketch Engine.

Extracción fraseológica basada en un corpus de fichas técnicas de impresoras 3D en español: colocaciones y locuciones*

Ángela Luque-Giráldez
Universidad de Málaga

Miriam Seghiri
Universidad de Málaga

1. Introducción

En mercados como el de la traducción profesional de España, destaca especialmente la traducción técnica, considerada como una de las especialidades con mayor demanda (Bruno *et al.* 2016) y, dentro de ella, sobresale, además, el área de la tecnología frente a otras (Rico Pérez y García Aragón 2016). A ello cabe sumar que las lenguas desde las que más se traduce en España son, por este orden, el inglés, el alemán y el francés debido a la relevancia de sus potencias industriales en España (Rico Pérez y García Aragón 2016). De hecho, en las últimas décadas, y precisamente en el ámbito de la tecnología, se ha desarrollado un nuevo dispositivo que hoy conocemos como impresora 3D. Esta impresora ha supuesto un gran avance para la ciencia y la tecnología, y se ha incorporado con éxito a distintos ámbitos como el de la automoción, la medicina, la ingeniería o el sector aeroespacial, pues sus aplicaciones son infinitas. La impresión 3D se está diseñando de ma-

* El presente trabajo se enmarca en el seno de las redes «T3: Terminología, Técnica y Traducción», «INTEC» y «TACTRAD» (Ref. 719/2018), coordinadas por la Universidad de Málaga (España), así como en el marco de los proyectos «INMOCOR» (Ref. P20-00109), «VIP II» (Ref. PID2020-112818GB-I00), «Training network on language technologies for interpreters» (Ref. EUIN2017-87746), «Aplicación de técnicas avanzadas de PLN al ámbito de las Tecnologías de la Traducción y la Interpretación» (Ref. EQC2018-004572-P), «MI4ALL» (AndalucíaTech), «Introducción al mercado profesional de la traducción especializada» (Acción 731, UMA 2020), «POSTrad IV» (UVa), «TRIAJE» (Ref. UMA18-FEDERJA-067) y «PROFETA» (Ref. PIE19-033, UMA).

nera que cada vez sea más accesible económicamente para cualquier tipo de usuario. De hecho, según un estudio llevado a cabo por TMT Deloitte (2019), la industria de la impresión 3D está creciendo a un ritmo aproximado del 12,5 % cada año y, además, pudo estimar que en 2019 las ventas superarían los 2700 millones de dólares, como así ha sido, y prevé que en 2020 superarán los 3000 millones.

En este contexto de alta demanda es donde nace la necesidad de estudiar documentos relacionados con la venta de impresoras 3D —como pueden ser manuales de instrucciones o fichas técnicas, por ejemplo—, pues cada vez son más las empresas tecnológicas que desean comercializarlas por el mundo. Sin ir más lejos, la European Union of Associations of Translation Companies (EUATC 2018) señala que, en Europa, la manufacturación es una de las áreas que más traducción genera, siendo el segundo campo más traducido por empresas de traducción y el tercero por traductores autónomos. Además, cabe destacar que la traducción, y por ende redacción en español de este género —en el caso del ámbito técnico, no se puede olvidar que la mayor parte de los documentos para su comercialización, como es el caso de las fichas técnicas, provienen de traducciones efectuadas desde la lengua inglesa—, viene dada por imperativo legal con el fin de preservar los derechos de los consumidores, a través de la implantación de la normativa *Decisión N.º 768/2008/CE del Parlamento Europeo y del Consejo de 9 de julio de 2008 sobre un marco común para la comercialización de los productos y por la que se deroga la Decisión 96/465/CEE del Consejo*[1] que regula la comercialización de productos en la Unión Europea. En el artículo R19 del capítulo 3, se indica que toda descripción técnica deberá traducirse a la lengua o lenguas requeridas por el Estado miembro en el que se comercialice el producto.

Sin embargo, a pesar del éxito de ventas de este producto, su reciente irrupción en el mercado hace que estemos ante un terreno

1. La normativa *Decisión Nº 768/2008/CE del Parlamento Europeo y del Consejo de 9 de julio de 2008 sobre un marco común para la comercialización de los productos y por la que se deroga la Decisión 96/465/CEE del Consejo* puede consultarse en: https://eur-lex.europa.eu/eli/dec/2008/768(1)/oj [Consulta 24/11/20].

yermo en investigación terminológica y fraseológica. Llegados a este punto, no cabe duda de que el uso de corpus supone un gran aliado para la extracción fraseológica (Seghiri 2016 y 2017a), dando lugar a los estudios de fraseología basados en corpus. Dentro de los distintos tipos de corpus existentes, destacan aquellos llamados virtuales (Seghiri 2006), es decir, corpus integrados por textos descargados únicamente de Internet, lo que los convierte en una herramienta rápida y eficaz que permite estudiar un género y temática específicos y, entre otros aspectos, la terminología y fraseología utilizadas, el nivel de uso, la frecuencia o la variación de un lema, todo ello en contextos originales (Pitkowski y Vásquez Gamarra 2009).

Así, el objetivo principal de este trabajo es el de llevar a cabo una extracción fraseológica —en particular, de colocaciones y locuciones— basada en corpus de fichas técnicas de impresoras 3D en español.

2. Aspectos fraseológicos: colocaciones y locuciones

La fraseología se define en el *DLE* como la «parte de la lingüística que estudia las frases, los refranes, los modismos, los proverbios y otras unidades de sintaxis total o parcialmente fija» y, además, añade que puede definirse como un «conjunto de frases hechas, locuciones figuradas, metáforas y comparaciones fijadas, modismos y refranes, existentes en una lengua, en el uso individual o en el de algún grupo». Así, este conjunto de combinaciones se conoce como *unidades fraseológicas* o *fraseologismos*. En esta misma línea, Corpas Pastor describe las *unidades fraseológicas* (UFS) como:

> Combinaciones, cuyos rasgos distintos son: (1) su polilexicalidad; (2) su alta frecuencia de aparición como unidades habituales de la lengua y de coaparición por parte de sus elementos integrantes; (3) su institucionalización o convencionalización derivada de su reproducción reiterada; (4) su estabilidad entendida en términos de fijación (interna o externa con diversos subgrupos) y de especialización semántica; (5) su idiomaticidad y variación potenciales. Los últimos cuatro rasgos se dan en diverso grado (gradación) en los distintos tipos de unidades. (Corpas Pastor 1996: 169)

Siguiendo los pasos de Corpas Pastor (1996), Montoro del Arco define también las unidades fraseológicas como:

> [...] segmentos lingüísticos pluriverbales (formados por dos o más palabras gráficas) cuyos elementos componentes han perdido parte de sus propiedades paradigmáticas y sintagmáticas propias (que pueden seguir manteniendo en otros contextos) y que se comportan como una unidad institucionalizada: en el plano morfosintáctico, manifiestan un determinado grado de fijación formal, y un determinado grado de variación potencial fija; desde el punto de vista semántico, presentan un determinado grado de idiomaticidad o especialización semántica; desde el punto de vista pragmático, se caracterizan por haber adquirido valores que afectan especialmente a alguna de las instancias de la comunicación: relación interactiva (cortesía, etc.), relación entre las partes del discurso (conexión) o actitud del hablante ante el discurso (modalidad). Cada una de ellas se define además, por una función que puede hallarse dentro del marco estructural o bien en un marco discursivo. (Montoro del Arco 2006: 110-111)

En el *Manual de Fraseología Española* de Corpas Pastor (1996), se encuentra, además, una clasificación de las unidades fraseológicas —siendo una de las más aceptadas hasta la fecha— que la autora divide en tres esferas, a saber, colocaciones, locuciones y enunciados fraseológicos. En la presente investigación, nos centraremos en estas dos primeras, de carácter nominal, debido al hecho de que en los textos técnicos suelen predominar los tecnicismos en forma de sustantivo —según comprobaron Pérez Carrasco y Seghiri (2020a en prensa) en un estudio aplicado a textos de automoción— y predecimos que es lo que posiblemente pueda ocurrir en este trabajo.

Así, por un lado, las colocaciones son unidades fraseológicas que, desde el punto de vista del sistema de la lengua, actúan como sintagmas completamente libres, pero a la vez siguen unas reglas que restringen su combinación (Corpas Pastor 1996: 53). Estas construcciones semifraseológicas están formadas por dos o más unidades léxicas, entre las que se distinguen la base y el colocativo. El colocativo se escoge de forma (parcialmente) arbitraria para expresar un sentido dado o un papel sintáctico respecto a la unidad principal, conocida como la base (Alonso Ramos 2002). Otros autores, como

Osorio Olave y Serra Sepúlveda (2012: 105) afirman que son «combinaciones frecuentes y presentes de dos o más palabras, cuya particularidad es construir una categoría intermedia entre las combinaciones llamadas "libres" y las combinaciones "fijas" o fraseológicas de la lengua». En nuestro caso, entendemos las colocaciones en la misma línea que lo hace García Rodríguez (2019: 76), al afirmar que son «combinaciones frecuentes formadas por una base, es decir, una palabra semánticamente autónoma, y un colocativo, una palabra seleccionada de la propia base».

De este modo, autores como Benson *et al.* (1986), Corpas Pastor (1996: 270) o Koike (2001) distinguen los siguientes seis tipos de colocaciones:

(1) Verbo + sustantivo (sujeto)
Asestar un golpe, estallar una guerra
(2) Verbo + (preposición +) sustantivo (objeto)
Poner a prueba, entablar amistad
(3) Adjetivo + sustantivo/Sustantivo + sustantivo
Éxito fulgurante, visita relámpago
(4) Sustantivo + preposición + sustantivo
Un banco de peces, un enjambre de abejas
(5) Verbo + Adverbio
Desear fervientemente, negar rotundamente
(6) Adjetivo + Adverbio
Opuesto diametralmente, estrechamente ligado.

Por otro lado, Penadés Martínez (2012: 23) define la locución como «una combinación fija de palabras que funciona como elemento de la oración y cuyo significado no se corresponde con la suma de los significados de sus componentes». Esta idea coincide con las afirmaciones de Corpas Pastor (1996: 270-271), quien también entiende las locuciones como unidades fraseológicas formadas por combinaciones de palabras cuyo significado no se basa en la suma de los elementos léxicos constituyentes y que tienen carácter idiomático. Así, Corpas Pastor (1996: 270) establece los siguientes tipos:

(1) Locuciones nominales: mosquita muerta, paño de lágrimas, el qué dirán
(2) Locuciones adjetivas: corriente y moliente, más papista que el Papa
(3) Locuciones adverbiales: gota a gota, de tapadillo, a raudales
(4) Locuciones verbales: revolver cielo y tierra, nadar y guardar la ropa
(5) Locuciones prepositivas: a causa de, a pesar de
(6) Locuciones conjuntivas: antes bien, a fin de que
(7) Locuciones clausales: salirle a alguien el tiro por la culata, como quien oye llover.

Una vez abordados sucintamente los distintos tipos de colocaciones y locuciones, procederemos a estudiar qué combinaciones son las más frecuentes en las fichas técnicas de impresoras 3D a través de un corpus que se compilará a tal efecto. Esta búsqueda será más sencilla gracias a la anotación del corpus, que permite la búsqueda de combinaciones de etiquetas que designan distintas categorías gramaticales. En este sentido, la ficha técnica, como género textual, se caracteriza por ser de tipo expositivo, en tanto que pretende presentar una información clara y objetiva de forma neutral; además, esta información suele presentarse de forma esquematizada y estructurada, por lo que se prevé que exista una alta tendencia a la sustantivación.[2]

A continuación, procederemos a exponer los pasos para la compilación de un corpus de fichas técnicas 3D para poder estudiar, seguidamente, sus colocaciones y locuciones más frecuentes en la lengua española.

3. Protocolo de compilación de un corpus de fichas técnicas de impresoras 3D

Para la compilación de un corpus virtual de fichas técnicas de impresoras 3D, seguiremos la metodología propuesta por Seghiri (2006, 2015 y 2017b), que se divide en dos fases bien diferenciadas: la prime-

2. De igual forma, puede observarse elipsis de artículos o preposiciones y ausencia de conectores y, debido a la temática elegida de las impresoras 3D, destacará el empleo de una terminología especializada y, teniendo en cuenta la novedad de la temática, es probable que se dé una alta presencia de préstamos provenientes de la lengua inglesa.

ra fase está destinada a asegurar la representatividad cualitativa de la muestra compilada y, para ello, estableceremos, por un lado, el diseño del corpus y, por otro, un protocolo de compilación; la segunda fase está destinada a asegurar la representatividad cuantitativa, gracias al uso del programa informático ReCor.

Durante la primera fase de compilación, y con el fin de asegurar la representatividad cualitativa, procedemos a diseñar el corpus, que se caracteriza por ser: (1) *especializado* en el ámbito de la *técnica*, ya que los textos que se incluirán en él serán fichas técnicas de impresoras 3D redactadas en español; (2) *monolingüe*; (3) *virtual*, pues contiene exclusivamente textos descargados de Internet; a la par que (4) *textual*, ya que se trata de fichas técnicas al completo; y, por último, también será un (5) corpus *etiquetado*, aspecto que nos permitirá realizar búsquedas más rápido y estudiar con mayor precisión las colocaciones y locuciones.

Una vez diseñado el corpus, seguimos cuatro pasos para la compilación de este, a saber, búsqueda, descarga, formato y almacenamiento.

Así, para la búsqueda de textos[3] (Paso 1), hemos acudido directamente a las páginas web de las empresas que comercializan impresoras 3D —como HP, BQ o BCN3D, entre otras— y se han seleccionado las fichas técnicas disponibles en lengua española. La descarga de estos documentos (Paso 2) se puede realizar tanto de forma manual, como de manera automatizada a través de programas como BootCat, que descarga textos en lotes desde una página web concreta. Seguidamente, debe tenerse en cuenta que los programas de gestión de corpus pueden interrogar documentos en formato de texto plano o TXT (Paso 3), por lo que necesitaremos modificar el formato de las fichas técnicas (que suelen estar en PDF o HTML) y para ello nos serviremos de Abbyy FineReader, un programa de reconocimiento OCR que puede transformar prácticamente cualquier formato de texto. Por último, para almacenar todos los documentos (Paso 4), se ha creado una carpeta, internamente organizada en función del formato de la ficha técnica,

3. Para más información en torno a cómo garantizar la calidad en la búsqueda de documentos a través de la red Internet, véase Seghiri (2016).

para poder archivar cada documento. Así, en primer lugar, se ha creado una carpeta llamada «Fichas técnicas» que se divide en dos subcarpetas: una bajo el nombre de «PDF», en la que se incluirán las fichas en su formato original, y otra denominada «TXT», que albergará los textos ya convertidos a texto plano. Adicionalmente, se ha ideado una codificación que indicará la lengua del texto (ES), si se trata de un texto original (TO) o de un texto meta (TM) y, además, se contempla el género textual (FT) y un número por cada modelo de impresora a la que se le ha asignado una numeración (1001/1100, etc.). De este modo, la codificación se iniciará tal y como aparece en el siguiente ejemplo: «1101FTTOES». Esta estructura y codificación facilitará el almacenaje de los documentos que integran el corpus, así como su identificación y posibles futuras ampliaciones de la muestra compilada.

Tras la aplicación de los criterios de diseño y del protocolo de compilación, el resultado ha sido un corpus formado por 72 fichas técnicas en español (y 18.119 *tokens* o palabras), que puede considerarse representativo desde el punto de vista cualitativo.

A continuación, abordamos la segunda fase con el fin de comprobar la representatividad cuantitativa de la muestra compilada. Para ello, nos hemos valido de la herramienta ReCor,[4] diseñada para tal fin (cfr. Seghiri 2006; Corpas y Seghiri 2007), que sirve para determinar el tamaño mínimo de un corpus dado. Cuando el programa ha analizado el corpus, devuelve dos gráficas que indican si se ha cubierto la terminología básica de ese campo de especialidad, es decir, si hay un momento en el que no entran palabras nuevas (*types*) al corpus. ReCor, además, permite analizar la representación tanto a nivel de palabra (1 grama), como a nivel fraseológico (de 2 hasta 10 gramas). En nuestro caso, como realizaremos un estudio colocacional y locucional, se ha verificado la representatividad del corpus con 1[5] grama (cfr. Figura 1), así como con 2 y 3 gramas.

4. La licencia de uso del programa ReCor puede solicitarse de manera gratuita en la siguiente dirección URL: https://umapatent.uma.es/es/patent/metodo-para-la-determinacion-de-la-representae40/.

5. Para más información en torno a la explotación terminológica (de 1 grama) del corpus 3DCOR, véase Luque Giráldez y Seghiri (2019, 2020a y 2020b en prensa).

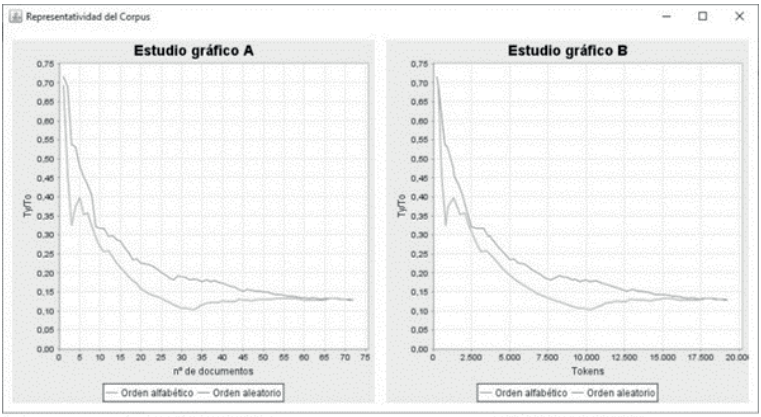


Figura 1. Gráficas del corpus en español ES3DCOR (1 grama) generadas por ReCor.

Al interpretar los datos que devuelve ReCor, hemos podido comprobar la representatividad en términos de cantidad para 1 grama, ya que las líneas de ambas gráficas se unen (cfr. Figura 1) y se aproximan a cero. Lo mismo ocurre con el análisis de 2 y 3 gramas. En consecuencia, podemos afirmar que el resultado final es un corpus virtual, monolingüe, textual y completo de fichas técnicas de impresoras al que denominaremos ES3DCOR, integrado por 72 fichas técnicas, y que es representativo tanto a nivel cualitativo, como cuantitativo. Llegados a este punto, se procederá al etiquetado de ES3DCOR con TagAnt para, después, continuar con su explotación.

En esta misma línea, véanse los trabajos de Arce Romeral y Seghiri (2018a y 2018b, 2019, 2020a, 2020b, 2021a en prensa, 2021b y 2021c en prensa), Pérez-Carrasco y Seghiri (2020a en prensa, 2020b y 2020c), Seghiri (2019) o Seghiri y Arce (2020).

4. Estudio fraseológico basado en el corpus ES3DCOR

En esta sección, nos ocuparemos de la explotación del corpus ES3DCOR para estudiar el tipo de colocaciones y locuciones empleadas en las fichas técnicas de impresoras 3D en español. A tal efecto será necesario anotar —o etiquetar— previamente el corpus mediante el uso del programa TagAnt.

4.1. Etiquetado de ES3DCOR con TagAnt

TagAnt[6] es un programa gratuito que se utiliza para el etiquetado automático de un corpus. Permite anotar distintas lenguas, entre ellas el español, y, para ello, utiliza etiquetas diferentes según la lengua con la que se trabaje.[7]

Para empezar el proceso de anotación, abriremos el programa y elegiremos en la barra inferior la lengua con la que queremos trabajar (1). TagAnt ofrece dos opciones para subir documentos: por un lado, se puede escribir o copiar un texto (Input Text) o, por otro lado, es posible subir un conjunto de textos (Input Files). En nuestro caso, seleccionaremos Input Files (2) para subir al programa todas las fichas técnicas del corpus en español en texto plano haciendo clic en Load (3). Una vez subidas todas fichas, haremos clic en el botón Start (4) (cfr. Figura 2).

6. El programa TagAnt puede descargarse de forma gratuita en https://www.laurenceanthony.net/software/tagant.

7. Es conveniente consultar las etiquetas utilizadas para cada lengua de trabajo, pues son diferentes, por ejemplo, la etiqueta en español para adjetivo es «ADJ», mientras que en inglés es «JJ». De este modo, las etiquetas generadas por el programa AntConc para este corpus en español pueden consultarse en: https://www.sketchengine.eu/spanish-treetagger-part-of-speech-tagset/.

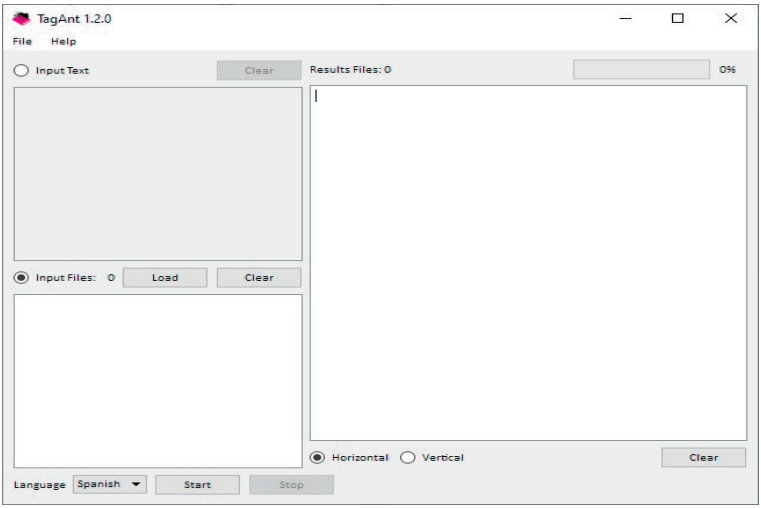


Figura 2. Subida de corpus en TagAnt.

El programa llevará a cabo automáticamente y en cuestión de segundos el etiquetado de los documentos. Además, cuando haya finalizado, almacenará todos los textos en la misma carpeta (manteniendo la codificación del texto original, a la que le añade la denominación «_tagged» al final)(cfr. Figura 3).

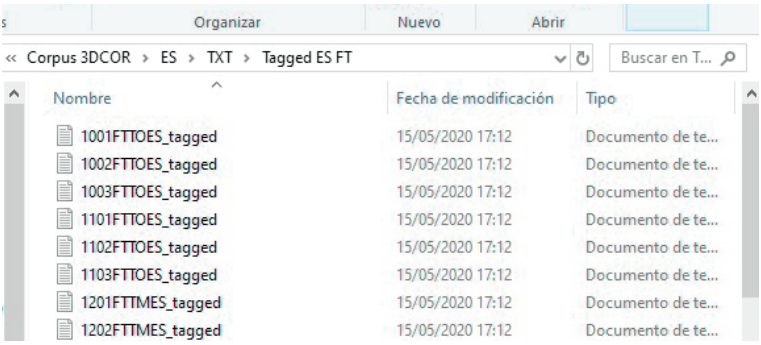


Figura 3. Producto final tras el etiquetado con TagAnt.

Una vez seguidos estos pasos, habremos obtenido el corpus ES3D-COR etiquetado al completo.

4.2. Extracción colocacional y locucional de ES3DCOR con AntConc

Ya etiquetado el corpus ES3DCOR, abordaremos la extracción colocacional y locucional de la muestra compilada con ayuda del programa de concordancias AntConc. Esta herramienta es de uso gratuito y permite observar palabras o grupos de palabras en un contexto determinado a través de las diferentes funciones que ofrece. Una de las opciones más interesantes que brinda el citado programa para el fin que nos proponemos es la de Clusters (cfr. Figura 4), que permite buscar una palabra o patrón y agrupar los resultados junto con las palabras que hay a la izquierda o a la derecha de una palabra dada. Este grupo de palabras se puede ordenar por frecuencia, por el inicio de la palabra, por el rango del grupo de palabras o por la probabilidad de que la primera palabra del grupo preceda a las palabras restantes.

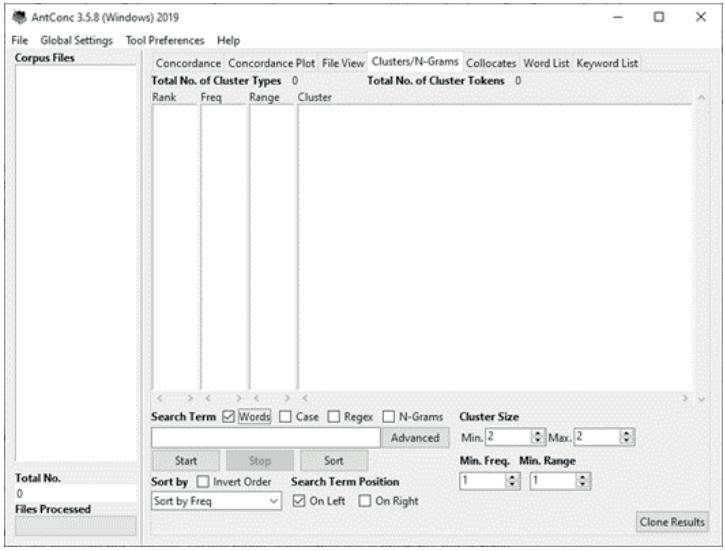


Figura 4. Función Clusters en la interfaz de AntConc.

Así, cuando subimos el corpus ES3DCOR etiquetado, y gracias a la función de Clusters, procederemos a extraer las colocaciones y locuciones más frecuentes. Para ello, llevaremos a cabo una búsqueda por etiquetas. De este modo, si deseamos que el programa, por ejemplo, extraiga la combinación sust. + verb., es necesario que la ecuación de búsqueda permita la localización de sustantivos comunes, tanto en singular, como en plural y verbos de cualquier tiempo y modo; para ello, escribiremos en la barra de búsqueda un asterisco (para recoger cualquier palabra), un guion bajo (que precede a cada etiqueta), la etiqueta deseada (la de sustantivo, en este caso, «nc», que corresponde a la abreviatura de nombre común y la de verbo, «v», que corresponde a cualquier modo y forma verbal) y finalmente un asterisco (para que recoja todas las variantes), como sigue «*_nc* *_v*» (incluyendo así las etiquetas: VEadj, VEger, VEinf, VLger, etc.).

A continuación, se ilustrará la búsqueda del ejemplo precedente (cfr. Figura 5).

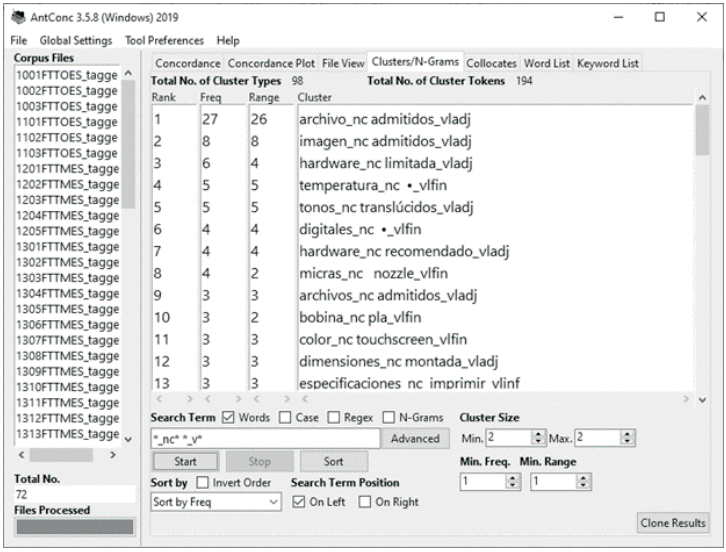


Figura 5. Interrogación del corpus etiquetado con AntConc.

Siguiendo este método, repetiremos la búsqueda atendiendo a la clasificación propuesta por Benson *et al.* (1986), Corpas Pastor (1996) y Koike (2001) para recoger todas las variantes posibles de colocaciones y locuciones según la siguiente muestra:

(1) Verbo + sustantivo (sujeto) / Sustantivo (sujeto) + verbo (para lo que se empleará la ecuación de búsqueda: *_v* *_nc* / *_nc* *_v*).
(2) Verbo + sustantivo (objeto) / Verbo + preposición + sustantivo (objeto) (para lo que se emplearán las ecuaciones de búsqueda: *_v* *_nc* / *_v* *_prep* *_nc*).
(3) Adjetivo + sustantivo / Sustantivo + adjetivo / Sustantivo + sustantivo (para lo que se emplearán las ecuaciones de búsqueda: *_adj* *_nc* / *_nc* *_adj* / *_nc* *_nc*).
(4) Sustantivo + preposición + sustantivo (para lo que se empleará la ecuación de búsqueda *_nc* *_prep* *_nc*).
(5) Verbo + adverbio / Adverbio + verbo (para lo que se emplearán las ecuaciones de búsqueda: *_v* *_adv* / *_adv* *_v*).
(6) Adjetivo + adverbio / Adverbio + adjetivo (para lo que se emplearán las ecuaciones de búsqueda: *_adj* *_adv* / *_adv* *_adj*).

Una vez establecidas las diferentes variantes existentes, llevaremos a cabo las distintas búsquedas a través de las ecuaciones y etiquetas supraindicadas para localizar las colocaciones y locuciones en el corpus con objeto de observar la frecuencia de uso de las mismas. El resultado ha sido un total de 1801 posibles colocaciones y locuciones entre las que destaca, en un principio, la construcción adj. + sust. / sust. + sust. con un 44,41 % de apariciones, de las cuales un 21,04 % son locuciones nominales o colocaciones sust. + sust., mientras que un 23,37 % son combinaciones sust. + adj. A continuación, se refleja en una gráfica la frecuencia de uso de cada combinación (cfr. Figura 6):

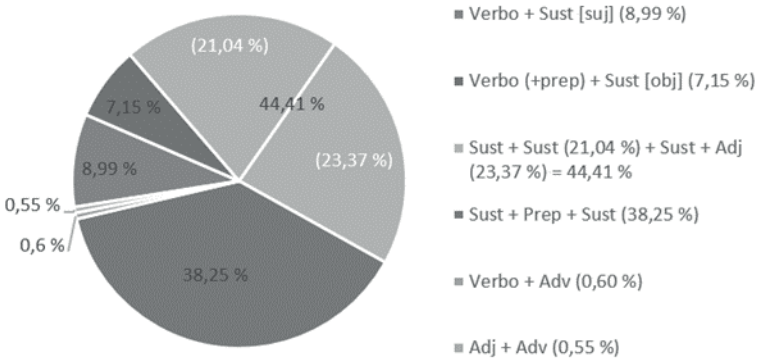


Figura 6. Frecuencia de uso de colocaciones y locuciones en el corpus ES3DCOR (I).

No obstante, a pesar de que el programa AntConc recoge un gran número de combinaciones sust. + sust. / sust. + adj., se han examinado estas combinaciones y se ha apreciado que muchos de los elementos de estas combinaciones forman parte, en realidad, de otras secuencias (como sucede con «operativos requisitos» o «material compatibilidad», ambos con una frecuencia de uso de 22 veces). Esto ha ocurrido porque AntConc no diferencia el final de una oración o enunciado y el principio de otra si no detecta un punto —a este respecto conviene subrayar el hecho de que en las fichas técnicas abundan listados de elementos, por ejemplo, que no se finalizan con punto—, por lo que ha identificado como una colocación o locución elementos que se encuentran en diferentes líneas. Esto ha ocurrido en 510 casos para las combinaciones del tipo sust. + sust. / sust. + adj.; sin embargo, esta incidencia no se da con las combinaciones del tipo sust. + prep. + sust. A vistas de lo ocurrido, hemos podido comprobar que la combinación sust. + prep. + sust. es la de mayor frecuencia en el corpus.

En consecuencia, el corpus recoge las siguientes cifras reales: un total de 1291 posibles colocaciones o locuciones entre las que destaca la combinación sust. + prep. + sust. con un 53,36 % de apariciones,

seguida de la combinación sust. + adj. con un 22,46 % de apariciones, mientras que la combinación sust. + sust. no tiene ninguna aparición.

Ilustraremos estas cifras reales a través de la siguiente gráfica (cfr. Figura 7) con la frecuencia de uso de cada combinación en el corpus ES3DCOR, una vez eliminados los casos no pertinentes (cfr. Figura 6).

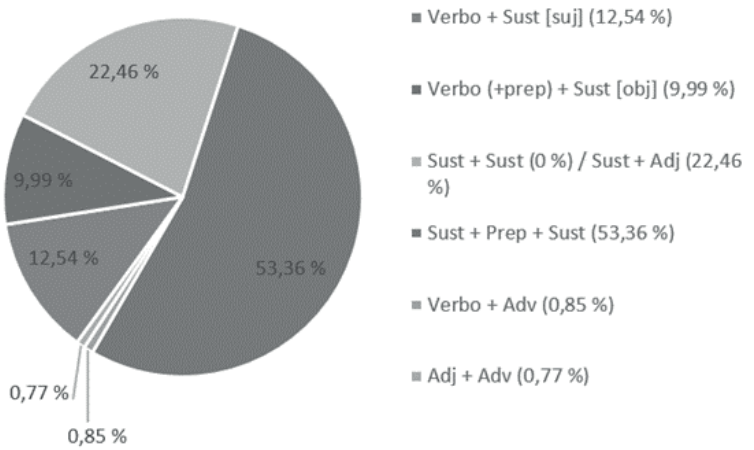


Figura 7. Frecuencia de uso de colocaciones y locuciones en el corpus ES3DCOR (II).

5. Discusión

Una vez estudiadas las colocaciones y locuciones del corpus ES3DCOR, procederemos a la elaboración de un listado que contendrá el tipo de colocación o locución más frecuente, a saber, sust. + prep. + sust., seguido de sust. + adj. Además, añadiremos la lista de frecuencia de estas combinaciones generada por la función de Clusters de AntConc que servirá de criterio a partir del cual ordenaremos la citada lista. Asimismo, hemos establecido un mínimo de frecuencia de aparición de 5 veces para incluir la locución o colocación en el listado. A continuación, se recoge una muestra de las combinaciones más frecuentes contenidas en la tabla final (cfr. Tabla 1).

Colocaciones y locuciones (sust. + prep. +sust.)	Frecuencia
Plataforma de impresión	47
Paso a paso	45
Espacio en disco	41
Tecnología de impresión	41
Temperatura de funcionamiento	36
Diámetro de filamento	32
Requisitos de alimentación	30
Temperatura de almacenamiento	29
Formatos de archivo	28
Resolución de capa	27
Compatibilidad de material	26
Dimensiones del producto	25
Interfaz de usuario	25
Propiedades de impresión	25
Dimensiones con embalaje	23
Humedad de funcionamiento	23
Peso con embalaje	23
Área de construcción	23
Requisitos del sistema	22
Volumen de impresión	21
Fabricación de filamento	20
Número de boquillas	20
Velocidad de movimiento	20

Tabla 1. Muestra del listado compuesto por locuciones nominales y colocaciones sust. + prep. + sust. en el corpus ES3DCOR.

La tabla completa con todas las colocaciones y locuciones extraídas se recoge al final del trabajo, en el Anexo 1, donde se incluyen las cifras reales una vez eliminados los falsos candidatos a colocación y locución (cfr. Figura 7). De este modo, la citada tabla contiene 71 colocaciones y locuciones, de las cuales 52 son del tipo sust. +

prep. + sust. y 19 son del tipo sust. + adj., pues ambas tienen una alta representación en el corpus ES3DCOR, 53,36 % y 22,46 % respectivamente; desafortunadamente, no se ha podido incluir la combinación sust. + verbo (12,54 %), ya que todas ellas tienen una frecuencia menor de 5. También, se ha podido observar que las unidades fraseológicas incluidas designan principalmente piezas o funciones del dispositivo (número de boquillas, resolución de capa, plataforma de impresión, etc.), compatibilidad del dispositivo (requisitos del sistema, interfaz de usuario, sistema operativo, etc.) y las dimensiones físicas de la impresora (dimensiones con embalaje, tamaño máximo, temperatura máxima, etc.).

6. Conclusiones

En el presente trabajo nos propusimos llevar a cabo una extracción fraseológica —en particular, de colocaciones y locuciones— basada en corpus de fichas técnicas de impresoras 3D en español. El resultado final ha sido un listado que contiene colocaciones y locuciones, en concreto, 71 entradas. Tras el análisis llevado a cabo, hemos podido comprobar que las combinaciones del tipo sust. + prep. + sust. en el ES3DCOR son las más utilizadas, seguidas de sust. + adj., verbo + sust. y por último verbo + adv. y adj. + adv. Estos resultados vendrían a confirmar la hipótesis inicial de que el género de ficha técnica cuenta efectivamente con una alta tendencia a la sustantivación (cfr. 2. Aspectos fraseológicos: colocaciones y locuciones).

Por último, cabe subrayar que la metodología aquí presentada puede ser aplicada para la creación, etiquetado y estudio colocacional y locucional de cualquier otro corpus o subcorpus, ya sea monolingüe, bilingüe o multilingüe, así como para su aplicación para el análisis monolingüe y monocultural de fichas técnicas de impresoras 3D, otros géneros análogos (como, por ejemplo, los manuales de instrucciones), temáticas (como aquellas de lápices o escáneres 3D) e incluso otras lenguas, además de estudios traductológicos y de corte contrastivo.

Referencias bibliográficas

Alonso Ramos, Margarita (2002): «Colocaciones y contorno de la definición lexicográfica», *Lingüística Española Actual*, vol. xxiv, n. 1, pp. 63-96.

Anthony, Laurence (2015): *TagAnt (Version 1.1.0)*, Tokyo, Waseda University [programa informático]. En línea, <https://www.laurenceanthony.net/software> [Consulta: 29/11/2020].

Anthony, Laurence (2019): *AntConc (Version 3.5.8)*, Tokyo, Waseda University [programa informático]. En línea, <https://www.laurenceanthony.net/software> [Consulta: 29/11/2020].

Arce Romeral, Lorena; Seghiri, Miriam (2018a): «Booth-friendly term extraction methodology based on parallel corpora for training medical interpreters», *Current Trends in Translation Teaching and Learning E*, vol. 5, pp. 1-46.

Arce Romeral, Lorena; Seghiri, Miriam (2018b): «Determinación de la representatividad cualitativa y cuantitativa de un corpus virtual de contratos de compraventa de viviendas (inglés-español)», *Traducción literaria y discursos traductológicos especializados*, Amsterdam, Peter Lang, pp. 309-330.

Arce Romeral, Lorena; Seghiri, Miriam (2019): «Diseño de plantillas de redacción y traducción al inglés (variedades británica y estadounidense) de contratos de compraventa de viviendas basadas en corpus», en Miriam Seghiri (eds.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*, Amsterdam, Peter Lang, pp. 69-106.

Arce Romeral, Lorena; Seghiri, Miriam (2020a): «LEXCOR: un corpus legislativo de compraventa inmobiliaria como herramienta de traducción», en Susana Álvarez Álvarez, María Teresa Ortego Antón (eds.), *Perfiles estratégicos de traductores e intérpretes en la transmisión de la información experta multilingüe en la sociedad del conocimiento del siglo XXI*, Granada, Comares, pp. 125-141.

Arce Romeral, Lorena; Seghiri, Miriam (2020b): «Creación de una plantilla de traducción de contratos de compraventa de viviendas. Un estudio aplicado a Irlanda», *Transatlantic Studies Network* (TNS), n. 9, pp. 65-76.

Arce Romeral, Lorena; Seghiri, Miriam (2021a, en prensa): «Implementación de plantillas de redacción de contratos de compraventa de viviendas en español: variedades peninsular y argentina», en Miguel Ángel García Peinado, Juan de Dios Torralbo (eds.), *La práctica de la traducción especializada jurídica*, Amsterdam, Peter Lang.

Arce Romeral, Lorena; Seghiri, Miriam (2021b): «COMPRAVENTACOR: un sitio *web* para la traducción de contratos de compraventa de viviendas en España, Argentina, Reino Unido y Estados Unidos», *Hikma: Revista de Traducción*, vol. 20, n. 1, pp. 221-254.

Arce Romeral, Lorena; Seghiri, Miriam (2021c, en prensa): «Implementation of a Glossary for the Translation of Housing Purchase and Sale Agreements in

Spain, Argentina, the United Kingdom and the United States: a Methodology Based on Corpus», *Hermēneus. Revista de Traducción e Interpretación*, vol. 24.

Benson, Morton; Benson, Evelyn; Ilson, Robert (1986): *The BBI Combinatory Dictionary of English. A guide to Word Combinations*, Amsterdam/Philadelphia, John Benjamins.

Bruno, Laura; Luque, Ileana; Ferreyra, Laura E. (2016): *La traducción de textos técnicos*, Córdoba, Universidad Nacional de Córdoba.

Corpas Pastor, Gloria (1996): *Manual de fraseología española*, Madrid, Gredos.

Corpas Pastor, Gloria; Seghiri, Miriam (2007): «Determinación del umbral de representatividad de un corpus mediante el algoritmo N- Cor», *SEPLN: Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural,* vol. 39, pp. 165-172.

DLE = Real Academia Española; Asociación de Academias de la Lengua Española (2019): *Diccionario de la Lengua Española*, versión electrónica 23.3. En línea, http://www.rae.es [Consulta: 29/11/2020].

EUATC 2018. *Expectations and concerns of the European language industry*. En línea, https://euatc.org/wp-content/uploads/2019/10/2018_Language_Industry_Survey_Report.pdf [Consulta: 29/11/2020].

García Rodríguez, Joseph (2019): *Las unidades fraseológicas del español y el catalán con elementos de la naturaleza: estudio cognitivo-contrastivo y propuesta de un diccionario electrónico de fraseológico bilingüe*, Barcelona, Universitat Autònoma de Barcelona.

Koike, Kazumi (2001): *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*, Madrid, Universidad de Alcalá.

Luque Giráldez, Ángela; Seghiri, Miriam (2019): «Extracción terminológica basada en corpus para la traducción de fichas técnicas de impresoras 3D», en Gloria Corpas Pastor, Ruslan Mitkov, Maria Kunilovskaya, María Araceli Losey León (eds.), *Computational and Corpus-based Phraseology*, Genève, Tradulex, pp. 99-106.

Luque Giráldez, Ángela; Seghiri, Miriam (2020a): «El corpus 3DCOR como herramienta para traducir fichas técnicas de impresoras 3D», en Pedro Luis Luchini, Ubirata Kickhöfel Alves (eds.), *Cuestiones del lenguaje: Desarrollo de lenguas extranjeras, enseñanza y traducción*, Mar del Plata, Universidad Nacional de Mar del Plata, pp. 125-140.

Luque Giráldez, Ángela; Seghiri, Miriam (2020b, en prensa): «Colocaciones y locuciones en las fichas técnicas de impresoras 3D (inglés-español): un estudio basado en el corpus 3DCOR», en Gloria Corpas Pastor *et al.*, *Fraseología basada en corpus*, Granada, Comares.

Montoro del Arco, Esteban Tomás (2006): *Teoría fraseológica de las locuciones particulares. Las locuciones prepositivas, conjuntivas y marcadoras en español*, Frankfurt am Main, Peter Lang.

Osorio Olave, Gabriela; Serra Sepúlveda, Susana (2012): «Colocaciones, compuestos sintagmáticos y locuciones nominales: hacia un intento de delimitación conceptual», *Lenguas modernas*, vol. 39, pp. 103-116.

Penadés Martínez, Inmaculada (2012): «La fraseología y su objeto de estudio», *Linred*, n. 10, pp. 1-17

Pérez Carrasco, Miriam; Seghiri, Miriam (2020a, en prensa): «MOTOCOR Y MOTOLEX: Estudio comparativo de la terminología técnica empleada en un subcorpus de manuales de usuario y un subcorpus legislativo sobre automoción», en Mar Ogeda, Francisco Rodríguez (eds.), *TRAK*, Sevilla, Universidad de Sevilla.

Pérez Carrasco, Miriam; Seghiri, Miriam (2020b): «Estudio diacrónico de terminología técnica de manuales de usuario de automoción a través de recursos de la Real Academia Española: *CORDE*, *CREA*, *CORPES XXI*, *NTTLE* y *MDA*», en Matteo De Beni, Dunia Hourani-Martín (eds.), *Corpus y estudio diacrónico del discurso especializado en español*, Frankurt am Main etc., Peter Lang, pp. 263-292.

Pérez Carrasco, Miriam; Seghiri, Miriam (2020c): «MOTOCOR: Compilación de un corpus técnico paralelo bilingüe (inglés-español) y su aplicación en la traducción directa de un texto sobre automoción», *Anales de Lingüística (Segunda época)*, n. 5, pp. 123-148.

Pitkowski, Elena Fabiana; Vásquez Gamarra, Javier (2009): «El uso de los corpus lingüísticos como herramienta pedagógica para la enseñanza y aprendizaje de ELE», *Instrumentos de trabajo sobre el español*, Montréal, Université de Montréal, pp. 31-51.

Rico Pérez, Celia; García Aragón, Álvaro (2016): *Análisis del sector de la traducción en España (2014-2015)*, Universidad Europea de Madrid, Campus de Villaviciosa de Odón.

Seghiri, Miriam (2006): *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*, Málaga, Universidad de Málaga.

Seghiri, Miriam (2015): «Determinación de la representatividad cuantitativa de un corpus *ad hoc* bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos/ Establishing the quantitative representativiness of an E-Reader User's Guide ad hoc corpus (English-Spanish)», en María Teresa Sánchez Nieto (ed.), *Corpus-based Translation and Interpreting Studies: From Description to Application,* Berlin, Frank & Timme, pp. 125-146.

Seghiri, Miriam (2016): «Diseño de una plantilla electrónica de evaluación de sedes web científicas para la creación de recursos de enseñanza-aprendizaje (alemán-inglés-español) / Designing an Evaluation Template of Scientific Web Sites for the Implementation of Teaching and Learning Materials (German-English-Spanish)», *Educatio siglo XXI*, vol. 34, n. 3, pp. 9-26.

Seghiri, Miriam (2017a): «Corpus e interpretación biosanitaria: extracción terminológica basada en bitextos del campo de la Neurología para la fase documental

del intérprete (Corpora and Medical Interpreting: Terminology Extraction Based on Bitexts for the Interpreter's Documentation Process)», *Panacea*, vol. 18, n. 46, pp. 123-132 [Volumen monográfico de interpretación en el ámbito biosanitario].

Seghiri, Miriam (2017b): «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores», *Babel*, vol. 63 n. 1, pp. 43-64.

Seghiri, Miriam (ed.) (2019): *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*, Amsterdam, Peter Lang.

Seghiri, Miriam; Arce-Romeral, Lorena (2020): *La traducción de contratos de compraventa inmobiliaria: un estudio basado en corpus aplicado a España e Irlanda*, Amsterdam, Peter Lang.

Technology, Media & Telecommunications Deloitte (2019): *Predicciones TMT 2019*. En línea, <https://www2.deloitte.com/content/dam/Deloitte/ec/Documents/technology-media-telecommunications/Deloitte-ES-TMT-Trends-2019-Folleto.pdf> [Consulta: 29/11/2020].

Anexo I

Colocaciones y locuciones	
1. Combinación sust. + prep. + sust.	**Frecuencia**
Plataforma de impresión	47
Paso a paso	45
Espacio en disco	41
Tecnología de impresión	41
Temperatura de funcionamiento	36
Diámetro de filamento	32
Requisitos de alimentación	30
Temperatura de almacenamiento	29
Formatos de archivo	28
Resolución de capa	27
Compatibilidad de material	26
Dimensiones del producto	25
Interfaz de usuario	25
Propiedades de impresión	25
Dimensiones con embalaje	23

Humedad de funcionamiento	23
Peso con embalaje	23
Área de construcción	23
Requisitos del sistema	22
Volumen de impresión	21
Fabricación de filamento	20
Número de boquillas	20
Velocidad de movimiento	20
Sistema de alimentación	17
Armario de material	16
Longitud de onda	13
Base de impresión	13
Placa de impresión	10
Velocidad de impresión	10
Altura de capa	9
Capas de impresión	9
Chorro de agua	9
Espesor de capa	9
Ram para gráficos	9
Tiempo de calentamiento	9
Tamaño de modelo	9
Formatos de imagen	8
Grabado por láser	8
Línea a línea	8
Potencia de salida	8
Área de grabado	8
Parámetros de fabricación	8
Cabezal de impresión	7
Bobinas de filamento	7
Área de funcionamiento	6
Fuente de alimentación	6
Integración de complementos	5

Diámetro de boquilla	5
Tanque de resina	5
Temperatura sin funcionamiento	5
Transferencia de archivos	5
Voltaje de entrada	5
2. Combinación sust. + sust.	**Frecuencia**
Disco duro	50
Sistema operativo	37
Filamento fundido	23
Requisitos operativos	22
Versiones posteriores	21
Requisitos mínimos	19
Especificaciones técnicas	13
Tamaño máximo	11
Temperatura máxima	10
Dispositivos móviles	9
Cama caliente	8
Pantalla táctil	8
Temperatura ambiente	8
Archivos compartibles	6
Panel táctil	6
Compatibilidad electromagnética	5
Emisión sonora	5
Filamento compatible	5
Información medioambiental	5

La terminología comercial a través de la fraseología: estudio de caso a partir de un corpus de especialidad

Dunia Hourani-Martín
Universität Leipzig / Università degli Studi di Verona

1. Introducción

La evolución de los dominios de especialidad conlleva la aparición de términos que den cuenta de las nuevas necesidades denominativas. El ámbito comercial no ha sido ajeno a los cambios acontecidos en la economía mundial en las dos últimas décadas del siglo xx (cfr. Gómez de Enterría y Martí 2009: 15-18). Estos acontecimientos han configurado un nuevo entorno comunicativo en el que se produce una

> interacción constante entre un vocabulario fuertemente especializado como es el de la economía y las finanzas, y el vocabulario de la lengua común o estándar, cuya dependencia mutua se pone de manifiesto con el nacimiento de un léxico renovador que acusa la presencia de numerosos procesos de carácter neológico, condicionados por los usos divulgativos que hoy ponen en práctica los medios de comunicación e Internet. (Gómez de Enterría y Martí 2009: 15-16)

Así pues, el lenguaje del comercio se caracteriza por su naturaleza innovadora y evolutiva, así como por su capacidad de generar nuevos significados (Mateo Martínez 2007: 196). Su dinamismo neológico se debe, como ocurre en otros lenguajes semitécnicos, a su procedencia de un ámbito de aplicación —no de estudios teóricos— y su posterior divulgación en los medios de comunicación (cfr. *ibid*. 2007: 193-195).

Los procedimientos de creación neológica en los ámbitos de especialidad en español no difieren de aquellos de los que dispone el sistema de la lengua general; para el campo de la economía actual se han constatado como especialmente productivos los préstamos —crudos, híbridos y adaptados— y los calcos, principalmente procedentes del

inglés,[1] la derivación (prefijación, sufijación y parasíntesis), la composición, las siglas y los acortamientos, además del trasvase de términos de otras especialidades debido a su carácter interdisciplinar (cfr. Gómez de Enterría 2003; Gómez de Enterría y Martí 2009).

Para estudiar la producción neológica en un lenguaje de especialidad novedoso y dinámico, se puede recurrir, a su vez, a la fraseología, pues su análisis puede ayudar a explicar los procesos de lexicalización y especialización que tienen lugar en el discurso de especialidad (Pavel 1993: 29; Kjær 2007: 508-509). De hecho, la fraseología desempeña un papel esencial en la creación y la evolución de la terminología; a través de ella se representan nuevos conceptos que, con el paso del tiempo, se establecen como unidades fraseológicas especializadas y pueden incluso llegar a convertirse en unidades terminológicas.

Teniendo en cuenta lo anterior, en el presente capítulo nos proponemos mostrar el potencial que tiene el análisis de la fraseología para comprender el desarrollo de la terminología en el ámbito comercial. Lo ilustraremos con un estudio de caso del verbo fraseológico *comercializar*. Siguiendo a Lorente Casafont (2002), entendemos por *verbo fraseológico* aquella unidad verbal que aparece «en los textos de especialidad para expresar acciones, procesos y estados, y cuyo significado no se distingue de su uso en contextos de lengua no especializada». Estos verbos forman unidades fraseológicas especializadas, en concreto, *construcciones verbonominales* (cfr. Tabares Plasencia 2012),[2] al combinarse necesariamente

1. La terminología española referida a la economía actual, en toda su amplitud, es clara deudora del inglés, de ahí que, entre los diversos procedimientos de formación de neologismos, sobresalgan los préstamos y calcos (Gómez de Enterría 2003: 204, 206-208; Gómez de Enterría y Martí 2009: 22-23). Este hecho también ha despertado un gran interés por su estudio desde una perspectiva lingüístico-terminológica (Gómez de Enterría 1992; Vélez Barreiro 2003; Diéguez Morales 2004 y 2005; Vázquez Amador 2017; López Zurita 2018, entre otros).

2. Seguimos aquí la clasificación de unidades fraseológicas especializadas propuesta por Tabares Plasencia (2012) que integra tres categorías: (i) cadenas gramaticales con valor preposicional (*a efectos de*; *de oficio*; *salvo que expresamente se disponga lo contrario*); (ii) construcciones verbonominales (*librar un cheque*; *celebrar matrimonio*; *entrar en vigor*) y (iii) textos formularios (*así, por esta nuestra Sentencia,*

con términos del dominio de especialidad: *comercializar productos*, *comercializar bienes*, etc. Las construcciones verbonominales, por lo tanto, al estar formadas por un núcleo terminológico (sujeto, objeto o complemento directo) y un verbo fraseológico, tienen la peculiaridad de poder actualizar transformaciones en el eje sintagmático y paradigmático, que están relacionadas en este último caso con los procedimientos de formación de palabras, a saber, la nominalización y la adjetivación deverbales. Precisamente, entre los diferentes mecanismos lexicogenéticos reseñados *supra*, la sufijación presenta especial recurrencia en el ámbito económico (Gómez de Enterría y Gallardo 2004: 412) y constituye un recurso muy provechoso para la creación de nuevas unidades por cuanto amplía el léxico con tan solo la incorporación de un afijo. Así pues, partimos del hecho de que las posibles transformaciones en el paradigma de los verbos fraseológicos pueden acabar fosilizándose en unidades terminológicas nuevas.

A continuación, presentaremos el análisis de las posibilidades de transformación del verbo fraseológico *comercializar* constatadas en el discurso real, esto es, en un corpus especializado del ámbito comercial (*vid. infra*). Extraeremos las unidades de la familia léxica de *comercializar* encontradas en nuestro corpus, atenderemos a los sufijos más productivos y a las estructuras más frecuentes en las que los derivados deverbales se insertan; también estudiaremos el valor conceptual de las transformaciones y la posible concurrencia con otras unidades terminológicas dentro del campo de especialidad. Por añadidura, expondremos los sucesivos pasos que hemos seguido en el análisis, de forma que puedan ser reproducidos como itinerario de aprendizaje en un contexto didáctico.[3] Finalizaremos con las conclusiones extraídas del análisis de las transformaciones fraseológicas y sus consecuencias terminológicas.

definitivamente juzgado, lo pronunciamos, mandamos y firmamos).

3. Si bien la terminología de un campo determinado se aprende estudiando conscientemente el propio ámbito de especialidad, los corpus se revelan como un recurso de gran valor y utilidad en el proceso de aprendizaje de terminología (*vid*. Hourani-Martín 2021). No obstante, considerando la amplitud de géneros profesionales en los que se emplea el lenguaje comercial y, con ello, la cantidad de usuarios que hacen uso de él (Mateo Martínez 2007: 194-195), un análisis como

2. Metodología

Para el análisis del verbo fraseológico *comercializar* hemos tenido en cuenta, por un lado, su aparición en el corpus de estudio en su *forma base* (cfr. Hourani Martín 2017; Tabares Plasencia e Ivanova 2020), esto es, aquella en la que las funciones sintácticas de la construcción verbonominal en la que se inserta se mantienen estables y no presentan alteración en el esquema [sujeto + verbo + objeto o complemento preposicional]. Por otro lado, hemos considerado las transformaciones paradigmáticas del elemento verbal, a saber, los cambios de categoría morfológica en forma de nominalización y adjetivación.

El estudio se lleva a cabo en un corpus especializado compuesto por textos académicos del ámbito del comercio procedentes de Argentina, España y México que han sido redactados entre 1990 y 2018. El corpus constituido cuenta con un total de 1.064.140 *tokens* (34.124 *types*) (cfr. Hourani-Martín 2021).[4] Para la gestión y el análisis del corpus se ha utilizado el programa Sketch Engine.[5]

Dado que la contribución va enfocada a la formación terminológica, se ha priorizado la selección de una muestra textual sincrónica que refleje el estado actual del lenguaje comercial en un periodo lo suficientemente amplio para dar cuenta de su dinamicidad. Al mismo tiempo, con ánimo de completar el análisis, así como para entender la productividad actual de la unidad estudiada, determinar su concurrencia con otras formas y ampliar, contrastar o respaldar los datos extraídos de nuestro corpus, hemos utilizado en algunos casos recursos lexicográficos y corpus de referencia del español.

---

el que presentamos podría ser de utilidad en el aula de español de los negocios, de terminología o de traducción, entre otros.

4. Se trata de un subcorpus procedente del corpus DIACOM (cfr. De Beni y Hourani-Martín 2020 para más información sobre DIACOM y el proyecto en el que se enmarca).

5. *Vid.* Arias Rodríguez *et al.* (2019) para más información acerca de la creación de un corpus y su explotación con Sketch Engine.

3. Estudio de caso: *comercializar*

El verbo objeto de estudio se construye uniendo al adjetivo *comercial* el sufijo *-izar*. Este afijo se ha destacado como especialmente productivo en los lenguajes de especialidad en general (Lang 1992: 217; Serrano-Dolader 1999: 4693) y en el lenguaje económico en particular (Gómez de Enterría 2003: 208).[6] Su capacidad de combinación le permite adherirse a bases adjetivales y nominales, incluso, también, a nombres propios (Lang 1992: 218), otorgándoles, por lo general, un valor causativo (Serrano-Dolader 1999: 4694). Es también interesante para el presente estudio por su carácter internacional, ya que se encuentran correspondencias en otras lenguas como el inglés (*-ize*/*-ise*), el francés (*-iser*) o el alemán (*-isieren*) (*ibid*. 1999: 4693).

La extracción de los derivados del verbo *comercializar* en el corpus puede realizarse bien a través de una consulta manual, bien a partir de la función *Lista de palabras* de Sketch Engine. En nuestro caso hemos optado por este último recurso, dado el carácter inductivo del análisis que presentamos.[7] Para averiguar todas las posibles transformaciones mediante la herramienta de análisis del corpus, es necesario reducir la consulta a las formas que comienzan por *comercializ* (Figura 1). Asimismo, la herramienta de análisis del corpus permite agrupar los resultados por palabras o por lemas; en el primer caso, el programa devuelve una lista de todas las formas que contengan el segmento introducido, en el segundo, se agrupan en una única entrada todas las formas del paradigma derivativo. Nosotros hemos seleccionado esta segunda opción; no obstante, es necesaria su comprobación en el contexto, ya que el programa silencia algunos resultados, como veremos más adelante.

6. Este dato se constata también en el corpus de estudio, en el que se pone de relieve la gran frecuencia de verbos denominales o deadjetivales del tipo *optimizar, fidelizar, maximizar, internacionalizar, liberalizar, centralizar, rentabilizar, externalizar, revitalizar, operacionalizar, capitalizar, estandarizar, etc.*

7. Para más información sobre el enfoque inductivo (*corpus-driven*) o deductivo (*corpus-based*) en los estudios de corpus *vid*. Tognini-Bonelli (2001: 65-100).

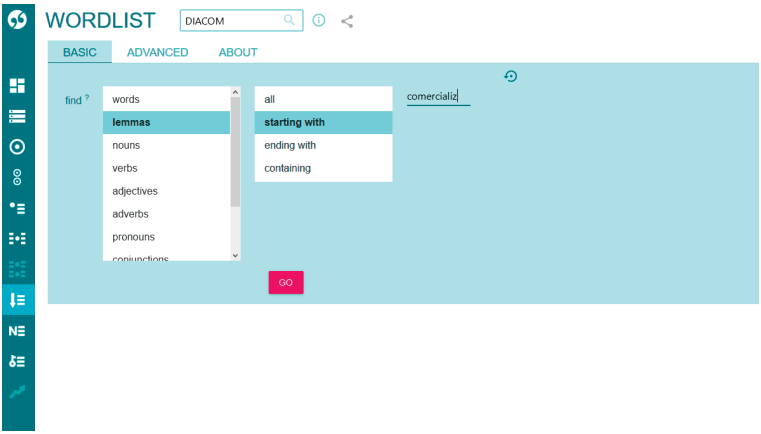


Figura 1. Búsqueda de la familia léxica de *comercializar* mediante *Wordlist*.

La búsqueda da como resultado un total de 4 lemas (Figura 2), entre los que se encuentran representadas las categorías nominal, verbal y adjetival, aunque con frecuencias dispares. A pesar de que, cuantitativamente, los resultados son, en algunos casos, reducidos, como mostraremos en los siguientes apartados, el análisis cualitativo ha puesto de manifiesto la utilidad del corpus a la hora de estudiar los procesos de formación terminológica en el ámbito comercial y mostrar la competencia entre distintas unidades.

Figura 2. Lista de lemas que comienzan por *comercializ-*.

Algunos de los lemas se pueden clasificar en diferentes categorías morfológicas, es el caso de la forma de participio del verbo, *comercializado*, que puede formar parte de un tiempo verbal o manifestarse como adjetivo, y *comercializador*, que puede actuar como sustantivo y adjetivo. Por ello ha sido necesario combinar diferentes funciones del programa de análisis de corpus para determinar la transformación representada. Se han empleado, para ello, las herramientas *Word Sketch*, que analiza los patrones sintácticos y las combinaciones léxicas de la unidad consultada, y *Concordance*, que extrae los contextos en los que esta aparece.

Para la distinción entre el participio verbal y el adjetivo, se han seleccionado a partir de *Word Sketch* los contextos en los que el verbo *comercializar* aparece en forma de participio (Figura 3), que posteriormente se muestran en la herramienta de concordancias (Figura 4). Esto ha restringido mucho más los contextos —23 casos en total—, que ahora pueden ser analizados manualmente de manera más sencilla.

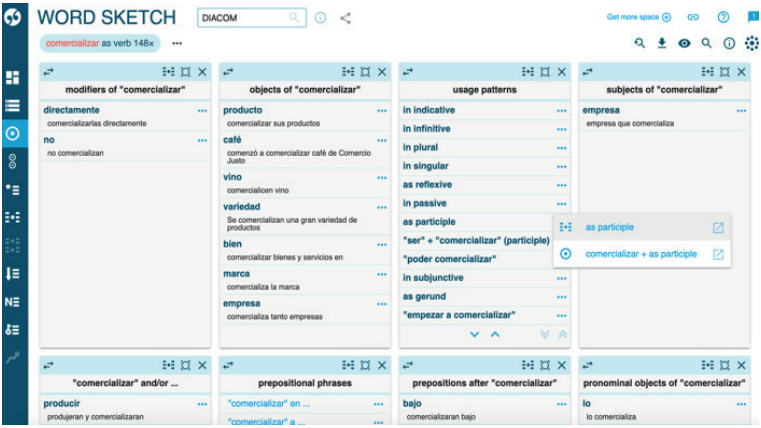


Figura 3. Búsqueda de contextos de *comercializar* como participio.

Figura 4. Concordancias para *comercializar* como participio.

En el caso de *comercializador*, se ha tenido que recurrir a la herramienta *Concordance* para extraer las concordancias, pues la unidad se encuentra lematizada como adjetivo (Figura 5) y la función *Word Sketch* no genera resultados como sustantivo (Figura 6).[8] Esto pone de relieve que la lematización automática aún presenta lagunas al clasificar unidades creadas a partir de sufijos que pueden ser productivos para diferentes categorías morfológicas.

8. Tampoco se obtienen resultados a través de la *Lista de palabras* restringiendo la consulta a la categoría morfológica *sustantivo* que finalice en *-izador*. Por el contrario, la búsqueda sí ofrece resultados del tipo *dinamizador* e *institucionalizador*, entre otros, que se han lematizado como sustantivos a pesar de que actúan como adjetivos en el contexto, hecho que se ha constatado tras la consulta de las concordancias.

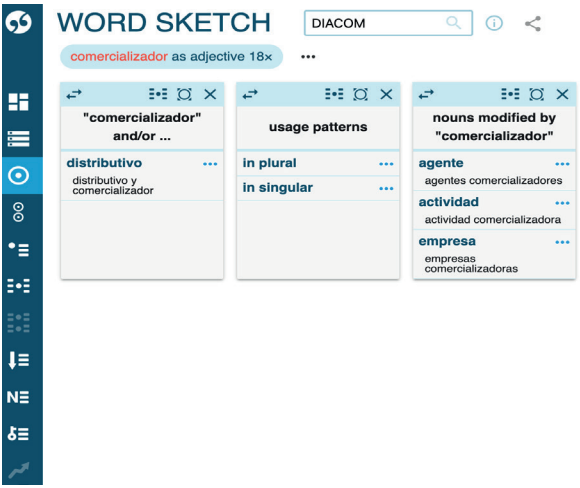


Figura 5. Resultados para *comercializador* como adjetivo.

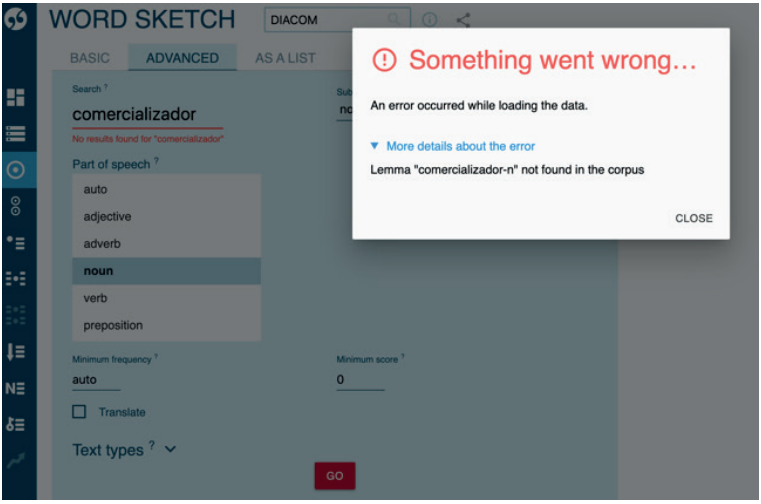


Figura 6. Resultados para *comercializador* como sustantivo.

Así pues, tras el análisis combinado con las diferentes funciones de Sketch Engine y el estudio de las concordancias, la familia léxica de *comercializar* representada en el corpus queda como se aprecia en la Tabla 1:

Lema	Categoría morfológica	Frecuencia (n.° / %)	Representación geográfica[9]
Comercialización	Sustantivo	191 / 53,35	AR, ES, MX
Comercializar	Verbo	135 / 37,71	AR, ES, MX
Comercializado	Adjetivo	13 / 3,63	AR, ES, MX
Comercializador	Sustantivo	13 / 3,63	AR, ES, MX
Comercializador	Adjetivo	5 / 1,4	ES, MX
Comercializable	Adjetivo	1 / 0,28	ES
Total		358 / 100	

Tabla 1. Familia léxica de *comercializar*.

3.1. Forma base

Como la mayoría de los verbos deadjetivales en *-izar* (Serrano-Dolader 1999: 4695) y como muestran los ejemplos 1-4, *comercializar* es transitivo y, por lo tanto, el núcleo terminológico al que acompaña (*productos*, *música*) actúa siempre como objeto directo.

(1) En cada división, los gerentes y los empleados diseñan, producen y *comercializan* sus propios productos. (MX07)
(2) Aquellas organizaciones que adecúan su producción a los estándares elaborados por Fairtrade International, tras una auditoría, acceden a dicha certificación y pueden *comercializar* sus productos dentro de este sistema. (AR07)
(3) Como ya sabemos, uno de los primeros productos que se *comercializó* con éxito por Internet fue el libro. (ES13)
(4) El sitio contiene información sobre las empresas de todos los continentes que *comercializan* música de forma legal. (AR06)

El verbo fraseológico *comercializar* ingresa en el diccionario académico por primera vez en 1970 como «[d]ar a un producto industrial,

9. La procedencia geográfica de los casos se mostrará siguiendo el código del país según la norma ISO 3166-1 alfa 2.

agrícola, etc., condiciones y organización comerciales para su venta» (*NTLLE*: *s. v. comercializar*), definición que se mantiene hasta la edición de 2001, en la que se incorpora una segunda acepción que atestigua la progresiva especialización de la unidad: «[d]ar a un producto condiciones y vías de distribución para su venta» y «[p]oner a la venta un producto» (*MDA* 2013: *s. v. comercializar*); esta última entrada se mantiene hasta hoy en día (*DLE* 2019: *s. v. comercializar*).

Tanto en el corpus de estudio como el corpus académico *CORPES XXI*[10] se confirma como un verbo muy frecuente que, incluso, entra en competición con el verbo denominal *comerciar* —«1. intr. Dedicarse a la compraventa o el intercambio de bienes o servicios» (*DLE* 2019: *s. v. comerciar*)— (ejemplo 4), al que transfiere su transitividad (ejemplos 5 y 6),[11] si bien suele usarse en este caso en estructuras de pasiva refleja:

(5) Los productos de consumo que tradicionalmente *se comercian* [a] través de marketing directo son los de mayor venta por Internet: libros, música, productos de informática o confección y calzado. (MX07)
(6) Por otra parte los productores y procesadores son igual de importantes al cumplir la función de proveer a las iniciativas de los productos que *se comercian* en ellas, esta parte es compleja al ser de muy diferentes bagajes culturales los diferentes productores y procesadores de las dos iniciativas, por lo tanto es muy interesante revisar su comportamiento e interacción. (MX04)

3.2. Nominalizaciones

En los lenguajes de especialidad la preferencia por la categoría nominal se justifica por su capacidad de abstracción y condensación de la

10. Dado el carácter sincrónico del estudio, se ha consultado el corpus de referencia *CORPES XXI* que recoge muestras textuales desde 2001 a 2020. En dicho corpus aparece el lema *comercializar* en 4711 ocasiones y en 3344 documentos.
11. El *Diccionario Panhispánico de Dudas* (*DPD* 2005: *s. v. comerciar*) recoge la voz *comerciar* definida como «[n]egociar comprando y vendiendo» y, además de su carácter intransitivo, remarca que «se construye con un complemento introducido por *en* o *con*». La Fundéu (2019) también dedica una de sus recomendaciones lingüísticas a los complementos del verbo *comerciar* e indica la frecuencia con la que se produce esta confusión. En el *CORPES XXI* se puede corroborar, asimismo, esta vacilación en las concordancias extraídas para el lema *comerciar*.

información. Por ello no es de extrañar que la manera más frecuente de insertar el verbo fraseológico en el contexto comunicativo sea a través de su nominalización, como muestran nuestros resultados (*vid.* Tabla 1). Los datos confirman, asimismo, la ya indicada productividad de la nominalización deverbal en español en general (Lang 1992: 184), señalada también en otros lenguajes de especialidad como el jurídico (por ejemplo, Alcaraz Varó y Hughes 2002: 30-31) o el de la genómica (Cabré y Estopà 2005: 78).

Los sufijos que encontramos dan lugar, por un lado, a un *nomen actionis* —*comercialización*— (ejemplos 7-10) y, por otro, de forma menos frecuente, a un *nomen agentis* —*comercializador*— (ejemplos 11-14).[12]

(7) En definitiva, cuando nos situamos en el campo del comercio electrónico, la protección al consumidor comprende todas las normas que regulan las relaciones de consumo, que nacen en la cadena de *comercialización* de bienes y servicios, es decir, en las siguientes áreas: manejo de datos personales, los defectos de calidad en los productos, los problemas en la prestación de un servicio, el incumplimiento de garantías, la información engañosa, el incumplimiento de los incentivos ofrecidos, los problemas del precio del producto o un servicio y la financiación de los mismos, etc. (ES12)

(8) La *comercialización* internacional, de productos normales o especiales, marcas, diseño de envases y empaques, precios, canales de distribución, promoción de ventas, publicidad, ventas, entre otros, se tendrá que adaptar a las operaciones en un mercado internacional, donde se tendrán que tomar en cuenta, aspectos, económicos, políticos y sociales, completamente diferentes a los de los mercados para los que fueron creados y diseñados originalmente, esto conlleva en ocasiones hasta cambios estructurales, por lo tanto, cuando una empresa decide incursionar en el camino de la *comercialización* internacional, debe estar dispuesta a

12. De hecho, la vitalidad de estos dos nombres deverbales se refleja de igual modo en repertorios lexicográficos especializados, por ejemplo, en el *Diccionario de Comercio Exterior* (2015) en el que ingresan como únicos representantes de la familia léxica *comercializar* y se definen como «[c]onjunto de acciones necesarias para hacer que un producto, ya sea industrial, industrial [*sic*] o agrícola alcance las condiciones y organización comercial necesarias para su venta» (*ibid. s. v. comercializador*) y «[p]ersona que se encarga de distribuir mercancías. Su actividad puede abarcar también la consolidación» (*ibid. s. v. comercializador*), para *comercialización* y *comercializador*, respectivamente.

realizar cambios incluso estructurales, que le permitan cumplir con los requerimientos internacionales. (MX02)
(9) Una estrategia nueva de distribución había nacido: un proceso transparente y eficiente de etiquetado abriría las puertas a la *comercialización* a gran escala. (AR07)
(10) Un importante eje de los negocios punto com estuvo centrado en los que aseguraban la "visibilidad" en el mercado y el desarrollo de nuevas formas de *comercialización* electrónica. (AR06)
(11) Los *comercializadores* [...] no concentran el proceso productivo completo, sino que compran, venden o procesan productos Fair Trade. (AR07)
(12) Tanto en Guatemala como en Honduras y Nicaragua, el café es un cultivo introducido y manejado por instituciones gubernamentales, las cuales fungieron durante mucho tiempo como intermediarias entre los pequeños productores y los *comercializadores*. (MX03)
(13) El bajo costo de las comunicaciones en Internet permite a las empresas con recursos financieros limitados, convertirse en *comercializadores* mundiales en las primeras etapas de su desarrollo. (MX07)
(14) Se instaló la Comisión Mixta para la Modernización de la Industria Micro, Pequeña y Mediana [...] que tenían [*sic*] como propósito promover la organización de estas empresas y coordinar acciones de apoyo con uniones de crédito y *comercializadoras*. (MX07)

El nombre de acción *comercialización* se actualiza a partir del sufijo *-ción*, afijo de gran vitalidad en el lenguaje económico y comercial (Lang 1992: 188; Gómez de Enterría 2003: 208; Gómez de Enterría y Martí 2009: 27; Cabré *et al*. 2010). La estructura de la nominalización sigue el esquema [*comercialización* + *de* + núcleo terminológico]; así, el objeto directo del verbo fraseológico pasa a ser sintagma preposicional precedido del nexo *de*.[13] No obstante, como muestran los ejemplos 9 y 10, el sustantivo deverbal también puede aparecer en el contexto sin núcleo terminológico, de forma que el término deviene más extenso[14] e, incluso, puede competir con *comercio*.

13. Esta estructura ha demostrado tener especial vitalidad en el ámbito económico para construir unidades terminológicas pluriverbales (Gómez de Enterría 2003: 213; Gómez de Enterría y Martí 2009: 30; Cabré *et al*. 2010).
14. Esto también se ha podido constatar en las obras lexicográficas académicas, pues *comercialización* entra a la par que el verbo del que deriva en 1970 con la acepción «[a]cción y efecto de comercializar» (*NTLLE*: *s. v. comercialización*). Esta

Por su parte, el nombre de agente *comercializador* se forma a partir del sufijo *-dor*, que se ha señalado como altamente productivo para la creación de *nomina agentis* (Lacuesta y Bustos Gisbert 1999: 4545), también en los lenguajes técnicos (Lang 1992: 189). Los derivados formados a partir de *-dor* en el corpus de estudio designan profesiones y pueden parafrasearse con una oración de relativo [que profesionalmente hace X] (Lacuesta y Bustos Gisbert 1999: 4543).[15] Desde el punto de vista terminológico es especialmente interesante porque esta nominalización se inserta siempre en el discurso sin núcleo terminológico (**comercializador de productos*), de forma que se condensa toda la carga semántica en el nombre agentivo.

Como se aprecia en los ejemplos, el sustantivo *comercializador* puede designar bien a una persona (ejemplos 11 y 12) bien a una entidad (ejemplos 13 y 14). En el primer caso, este nombre deverbal designa una profesión y, por lo tanto, podría concurrir con el término *comerciante*;[16] quizá la competencia entre ambas unidades explique la baja frecuencia de *comercializador* (13 ocurrencias) frente a *comerciante* (105 entradas) en el corpus de estudio.[17] Aún así, la formación de la unidad *comercializador* es síntoma de la evolución del campo del comercio y las consiguientes transformación y especialización de las profesiones. En el segundo caso, al referir a una entidad, se aprecia una alternancia entre

definición se mantiene hasta la edición de 1989, en la que se añade una segunda acepción con la marca diatécnica de economía: «[t]odas las actividades de una empresa relacionadas con la venta y distribución de productos» (*ibid.*); sin embargo, esta última se elimina en la siguiente edición de 1992 y no vuelve a aparecer (*MDA* 2013; *DLE* 2019: *s. v. comercialización*).

15. Este resultado contrasta con los datos mostrados para la economía actual en Gómez de Enterría (2003: 209) y Gómez de Enterría y Martí (2009: 28), donde se destaca el empleo del sufijo *-dor* para la formación de *nomina instrumenti* del tipo *navegador web*, *prestador de servicios*, *suministrador*, etc. Si bien este valor del afijo *-dor* también se constata en nuestro corpus de estudio, como exponemos en el caso que nos ocupa no constituye la única posibilidad.

16. Como se indica en Santiago Lacuesta y Bustos Gisbert (1999: 4544-4545) el sufijo *-dor* alterna con sustantivos agentivos en *-ante*.

17. El *CORPES XXI* también atestigua esta distancia entre las frecuencias de las dos unidades nominales: 524 casos para *comercializador* y 6739 para *comerciante*.

los géneros masculino y femenino a la hora de construir el nombre de agente; sin embargo la variante *comercializadora*, que se explicaría por el sustantivo de base para la formación (*empresa*) (cfr. Lacuesta y Bustos Gisbert 1999: 4544), es menos frecuente y solo aparece en tres ocasiones, dos en un texto de México y una en uno de España.

Por último, un dato que cabe destacar es que, a pesar de que los contextos del *nomen agentis* proceden de textos de España, Argentina y México, la frecuencia del término en estos dos últimos países es mayor, con un 53,85 % y 38,46 % de aparición respectivamente para Argentina y México. No obstante, el reducido número de ocurrencias no nos permite extraer conclusiones definitivas respecto a una posible preferencia diatópica por estas unidades en el discurso comercial.[18]

3.3. Adjetivación

La predominancia de la secuencia [nombre + adjetivo] en la formación de unidades pluriverbales en dominios de especialidad ha sido destacada en diversos estudios (Gómez de Enterría 2003: 212; Gómez de Enterría y Martí 2009: 29; Cabré *et al*. 2010). Por lo tanto, no es de extrañar que, en el caso de *comercializar*, la adjetivación deverbal también se revele como productiva y tenga lugar a través de 3 sufijos diferentes: *-dor*, *-ble* y, la forma de participio, *-(a)do.* Siguiendo la clasificación de Rainer (1999), el primero pertenecería a la clase de los adjetivos activos puros, mientras que los dos restantes constituirían adjetivos pasivos, incluyéndose *-(a)do* en el subconjunto participial y *-ble*, por contexto, en el subconjunto potencial.

3.3.1. El sufijo **-dor**

Este sufijo deverbal es uno de los más productivos en español para formar adjetivos deverbales puros (Rainer 1999: 4600) y aparece representado en el corpus de estudio en 5 ocasiones (Tabla 1).[19]

18. La consulta del *CORPES XXI* también ha puesto de manifiesto una mayor presencia de las dos variantes morfosintácticas (*comercializador* y *comercializadora*) en América.
19. En Gómez de Enterría (2003: 209), no obstante, se resalta especialmente la productividad del sufijo *-dor* para la formación de adjetivos disposicionales.

(15) El registro de Empresas de Comercio Exterior (ECEX) es un instrumento de promoción a las exportaciones, mediante el cual las *empresas comercializadoras* podrán ingresar en los mercados internacionales con facilidades administrativas y apoyos financieros de la banca de desarrollo, con el propósito de reunir la oferta nacional de productos en el extranjero, de acuerdo a la demanda internacional. (MX02)
(16) Entre los elementos considerados para optar por una u otra estrategia podemos citar: la naturaleza del producto, el ciclo de vida internacional del producto, los requerimientos legales, la competencia, las exigencias de los *agentes comercializadores*, el entorno físico, el tipo y condiciones del mercado y, lógicamente, la relación costo/beneficio. (ES14)
(17) En la exportación directa la actividad de exportación es desarrollada por la propia empresa fabricante que realizará las adaptaciones de su *actividad comercializadora* que resulten oportunas para adaptar su *marketing-mix* a los clientes internacionales. (ES14)

Como se aprecia en los ejemplos 15 y 16, los adjetivos formados a partir de *-dor* muestran la realización concreta de la acción verbal, de manera que sustituyen a una oración de relativo [empresa/agente que comercializa] concentrando la carga terminológica en el adjetivo. En el ejemplo 17, sin embargo, se aprecia que el núcleo al que acompaña no encajaría en la paráfrasis de relativo anterior [*actividad que comercializa]. A este respecto dice Rainer (1999: 4602) que «estamos asistiendo al nacimiento de un nuevo uso relacional de *-dor/a*». Así, lo que activa el adjetivo es un contenido relacional [actividad de comercialización], opción por la que quizá se haya optado para evitar la reproducción de la estructura [actividad + *de* + sustantivo deverbal en *-ción*] dentro de la frase.

3.3.2. El sufijo -(a)do

El adjetivo deverbal formado a partir de la forma participial en *-(a)do* se impone como construcción adjetiva más frecuente, con 13 ocurrencias en textos de los tres países (Tabla 1), lo que pone de manifiesto su productividad en el ámbito de especialidad.

(18) Entiende que es equiparable vender a través de una pequeña tienda especializada o del supermercado por el carácter de materia prima de los *productos*

comercializados en el CJ [comercio justo], como el café o el algodón, lo que hace, según él, que el beneficio del sistema recaiga en los países productores del Sur. (AR04)
(19) En este sentido, no tiene el control sobre el *marketing* internacional de los *productos comercializados* en otros mercados algo que no suele ser deseable, al menos a largo plazo. (MX07)
(20) Si no se suele aplicar ese margen, por ser un *producto o servicio no comercializado* habitualmente; se le aplica el margen habitual que aplican personas o entidades independientes. (ES14)

La forma de participio *comercializado* hereda las propiedades esenciales del elemento verbal (*vid.* Bevilacqua 2004: 206-212) y admite la paráfrasis «que {ha sido/está/es} PP» (Rainer 1999: 4597).

Asimismo, como se observa en el ejemplo 20, la negación de las formas de participio se realiza anteponiendo el adverbio *no*, que niega el significado del verbo considerado positivo: *comercializado ≠ no comercializado*.

3.3.3. El sufijo -ble

El sufijo *-ble* es uno de los más productivos e importantes en español y presenta una gran capacidad de combinación con verbos, en general, transitivos (Lang 1992: 204-205; Rainer 1999: 4609). Por ello lo incluimos en nuestro análisis, a pesar de que en el único ejemplo encontrado en el corpus aparezca tras un verbo copulativo (ejemplo 21). Semánticamente, el adjetivo deverbal podría parafrasearse a través de «que puede {ser PP/Vse}» (Rainer, 1999: 4597), a saber, [que puede ser comercializado/comercializarse] en cualquier contexto.

(21) Esto no significa necesariamente que el comercio internacional de servicios haya sido importante, y durante mucho tiempo se ha considerado que la mayor parte de los *servicios* no eran *comercializables* internacionalmente ya que requerían el contacto directo entre la empresa y el cliente, lo que ha condicionado la poca atención que se le ha dedicado en la literatura económica y en las estadísticas de comercio internacional. (ES14)

De especial interés al respecto es también la ausencia de la unidad *comerciable* en el corpus; así, este adjetivo deverbal, que sí se verifica en

la obra lexicográfica académica (*DLE* 2019: *s v. comerciable*), parece no ser usual en el dominio de especialidad estudiado. Dado que la falta de contextos puede deberse a lo restringido del corpus de estudio, a modo de control, hemos efectuado una consulta de la unidad *comerciable* en el *CORPES XXI*, con unos resultados de 32 ocurrencias en 29 textos. Estos datos confirman que, incluso en un corpus tan grande como el académico, la frecuencia actual de *comerciable* es anecdótica. Este hecho podría deberse a su competencia directa con la variante *comercializable*, mucho más frecuente en *CORPES XXI* (79 ocurrencias en 66 documentos), pues ambos adjetivos aparecen acompañando a idénticos sustantivos: *producto, bien, servicio*. Al mismo tiempo, también podría explicar la ausencia de *comerciable* en el corpus de estudio. No obstante, será necesario comprobar estos datos en un corpus especializado de mayor tamaño como el que se está construyendo en la Universidad de Verona (*vid*. nota 4).

4. Conclusiones

A través del presente estudio hemos mostrado la relevancia que tiene la fraseología a la hora de desarrollar la terminología de un ámbito de especialidad dinámico y en evolución. Para ello, hemos estudiado la familia léxica del verbo *comercializar* en un corpus especializado de textos comerciales. Los resultados han puesto de manifiesto la productividad de *comercializar* en el contexto actual. Así, la evolución del campo del comercio ha desembocado en una progresiva especialización de los procesos (desde la producción, hasta el *marketing* o la logística), la aparición de nuevas figuras y, con ello, la necesidad de dar respuesta a la comunicación especializada. Ante tal necesidad, las transformaciones paradigmáticas se confirman como uno de los recursos más provechosos para formar y enriquecer la terminología comercial y, de hecho, es la forma preferida a la hora de plasmar el verbo estudiado en el contexto comunicativo.

Especialmente la nominalización, pero también la adjetivación, se han constatado como recursos habituales y productivos a la hora de

incorporar el verbo en el discurso. Al mismo tiempo, las transformaciones estudiadas se insertan en estructuras del tipo [sustantivo deverbal + *de* + núcleo terminológico] y [núcleo terminológico + adjetivo deverbal], muy frecuentes en el discurso económico y comercial, que también revelan un mayor grado de lexicalización de la unidad al manifestarse en patrones morfosintácticos estables, que pueden derivar en su fosilización como términos.

Por último, quisiéramos poner de relieve la importancia de los corpus para descubrir tendencias terminológicas en los discursos de especialidad. Un estudio de tipo inductivo como el que hemos realizado nos ha permitido descubrir inflexiones dentro del paradigma verbal que sí son relevantes, por sus consecuencias conceptuales, para la formación de unidades terminológicas como, por ejemplo, en el caso de *comercializador* y *comercializadora*.

Referencias bibliográficas

Alcaraz Varó, Enrique; Hughes, Brian (2002): *El español jurídico*, Barcelona, Ariel.

Arias Rodríguez, Iván *et al.* (2019): *Taller sobre herramientas de análisis textual: La herramienta Sketch Engine*. En línea, <https://biblioteca.ulpgc.es/files/repositorio_de_docum152/noticias/taller_de_sketch_engine_18_02_2019_0.pdf> [Consulta: 15/10/2020].

Bevilacqua, Cleci Regina (2004): *Unidades fraseológicas especializadas eventivas: descripción y reglas de formación en el ámbito de la energía solar*, Barcelona, Universitat Pompeu Fabra, Tesis doctoral dirigida por María Teresa Cabré.

Cabré, M.ª Teresa *et al.* (2010): «Comparación de algunas características lingüísticas del discurso especializado frente al discurso general: el caso del discurso económico», en María Rosario Caballero Rodríguez, María Jesús Pinar Sanz (coords.), *Modos y formas de la comunicación humana*, Cuenca, Servicio de Publicaciones de la Universidad de Castilla-La Mancha, vol. 2, pp. 453-460.

Cabré, M.ª Teresa; Estopà, Rosa (2005): «Unidades de conocimiento especializado: caracterización y tipología», en María Teresa Cabré, Carme Bach Martorell (eds.), *Coneixement, llenguatge i discurs especialitzat*, Barcelona, Institut Universitari de Lingüística Aplicada, pp. 69-94.

CORPES XXI = Real Academia Española (en línea): *Corpus del español del siglo xxi*, <https://webfrl.rae.es/CORPES/> [Consulta: 15/10/2020].

Diccionario de Comercio Exterior = Núñez Hernández, Marco Antonio (2015): *Diccionario de Comercio Exterior*. En línea, <https://app.vlex.com/#WW/sources/12800> [Consulta: 09/10/2020].

Diéguez Morales, M. Isabel (2004): «El anglicismo léxico en el discurso económico de divulgación científica del español de Chile», *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, vol. 10, pp. 117-141.

Diéguez Morales, M. Isabel (2005): «Análisis contrastivo del anglicismo léxico en el discurso económico semiespecializado y de divulgación científica del español de Chile», *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, vol. 12, pp. 129-156.

De Beni, Matteo; Hourani-Martín, Dunia (2020): «El discurso especializado del comercio en español: diseño de un corpus para su estudio diacrónico», *Lexis*, vol. 44, n. 1, pp. 75-112.

DLE = Real Academia Española; Asociación de Academias de la Lengua Española (2019): *Diccionario de la Lengua Española*, versión electrónica 23.3. En línea, <http://www.rae.es> [Consulta: 08/10/2020].

DPD = Real Academia Española; Asociación de Academias de la Lengua Española (2005)*: Diccionario Panhispánico de Dudas*, versión electrónica 23.3. En línea, <https://www.rae.es/dpd/> [Consulta: 08/10/2020].

Fundéu = Fundación del Español Urgente (2019): «Comerciar con *algo*, *no* comerciar *algo*». En línea, <https://www.fundeu.es/recomendacion/comerciar-con-algo-no-comerciar-algo/> [Consulta: 08/10/2020].

Gómez de Enterría, Josefa (1992): «Neología y préstamo en el vocabulario de la economía», *Anuario de estudios filológicos*, vol. 15, pp. 97-106.

Gómez de Enterría, Josefa (2003): «Creación léxica en torno a la "Nueva Economía"», en Fernando Sánchez Miret (ed.), *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, Tübingen, Max Niemeyer, vol. 3, pp. 203-216.

Gómez de Enterría, Josefa; Gallardo, Natividad (2004): «Reflexiones sobre neología especializada en vocabularios procedentes de diversos ámbitos de carácter sociológico», en Luis González, Pollux Hernúñez (coord.), *Las palabras del traductor: actas del II Congreso "El español, lengua de traducción"*, Bruxelles, ESLETRA, pp. 411-425.

Gómez de Enterría, Josefa; Martí, Manuel (2009): «El vocabulario de la economía actual (1998-2008)», en Natividad Gallardo San Salvador, Josefa Gómez de Enterría (eds.), *Estudios de léxico especializado*, Granada, Atrio (*Traducción en el atrio*, n. 17), pp. 15-133.

Hourani Martín, Dunia (2017): *Unidades fraseológicas especializadas en un corpus de derecho ambiental sobre la protección frente al cambio climático (alemán-español)*, Granada, Universidad de Granada, Tesis doctoral dirigida por Esteban Tomás Montoro del Arco y Encarnación Tabares Plasencia.

Hourani-Martín, Dunia (2021): «El uso del corpus DIACOM para la enseñanza de terminología comercial: aplicaciones y posibilidades», en Elisa Sartor (ed.), *Los corpus especializados en la lingüística aplicada: traducción y enseñanza*, Mantova, Universitas Studiorum, Colección Pliegos Hispánicos, pp. 185-203.

Kjær, Anne Lise (2007): «Phrasemes in legal texts», en Harald Burger *et al.* (eds.), *Phraseologie: ein internationales Handbuch zeitgenössischer Forschung/Phraseology: An International Handbook of Contemporary Research*, Berlin, De Gruyter, pp. 506-516.

Lang, Mervyn F. (1992): *Formación de palabras en español. Morfología derivativa productiva en el léxico moderno*, Madrid, Cátedra.

López Zurita, Paloma (2018): «El uso de anglicismos en el contexto del vocabulario académico de Marketing», *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile*, vol. 39, pp. 114-139.

Lorente Casafont, Mercè (2002): «Verbos y discurso especializado», *Estudios de Lingüística del Español*, vol. 16. En línea, <http://elies.rediris.es/elies16/Lorente.html> [Consulta: 15/10/2020].

Mateo Martínez, José (2007): «El lenguaje de las ciencias económicas», en Enrique Alcaraz Varó, José Mateo Martínez, Francisco Yus Ramos (eds.), *Las lenguas profesionales y académicas*, Barcelona, Ariel, pp. 191-203.

MDA = Real Academia Española (2013): *Mapa de diccionarios*. En línea, <http://web.frl.es/ntllet> [Consulta: 08/10/2020].

NTLLE = Real Academia Española: *Nuevo tesoro lexicográfico de la lengua española*. En línea, <http://buscon.rae.es/ntlle/SrvltGUILoginNtlle> [Consulta: 08/10/2020].

Pavel, Silvia (1993): «Neology and phraseology as terminology-in-the-making», en Helmi B. Sonneveld, Kurt L. Loening (eds.), *Terminology: Applications in Interdisciplinary Communication*, Amsterdam/Philadelphia, John Benjamins, pp. 21-34.

Rainer, Franz (1999): «La derivación adjetival», en Ignacio Bosque, Violeta Demonte (dirs.), *Gramática Descriptiva de la Lengua Española*, Madrid, Espasa Calpe, t. III, pp. 4595-4643.

Serrano-Dolader, David (1999): «La derivación verbal y la parasíntesis», en Ignacio Bosque, Violeta Demonte (dirs.), *Gramática Descriptiva de la Lengua Española*, Madrid, Espasa Calpe, t. III, pp. 4683-4755.

Tabares Plasencia, Encarnación (2012): «Analyse und Abgrenzung juristischer phraseologischer Einheiten im Spanischen und Deutschen und ihre Bedeutung für die Übersetzung», *Lebende Sprachen*, vol. 57, n. 2, pp. 314-328.

Tabares Plasencia, Encarnación; Ivanova, Vessela (2020): «Die juristische Fachphraseologie und ihre Transformationen als Übersetzungsproblem im Sprachenpaar Deutsch-Spanisch», en Carsten Sinner, Christine Paasch-Kaiser, Johannes Härtel (eds.), *Translation 4.0. Dolmetschen und Übersetzen im Zeitalter der Digitalisierung*, Berlin, Peter Lang, pp. 289-309.

Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*, Amsterdam/Philadelphia, John Benjamins (*Studies in Corpus Linguistics*, n. 6).

Vázquez Amador, María (2017): «La influencia del inglés en la prensa comercial de los siglos xviii y xix», *Odisea: Revista de estudios ingleses*, vol. 18, pp. 133-148.

Vélez Barreiro, Marco (2003): *Anglicismos en la prensa económica española*, A Coruña, Universidade da Coruña, Tesis doctoral dirigida por Antonio Raúl de Toro Santos.

Los corpus como herramientas para la traducción español-inglés en el sector chacinero en el siglo xxi*

María Teresa Ortego Antón
Universidad de Valladolid

1. La traducción al inglés en la industria chacinera española

La industria agroalimentaria constituye uno de los sectores económicos más importantes en España (MAPA 2020) y la industria cárnica es la primera en lo que respecta a facturación y a empleos directos. Dentro de esta industria, las empresas chacineras se presentan como la opción más prometedora a medio y largo plazo en amplias regiones de tradición agrícola y ganadera que, debido al éxodo poblacional, se encuentran en riesgo de una alarmante despoblación, como ocurre con las áreas despobladas del sur de Europa (SSPA por sus siglas en inglés). Entre los factores que pueden sustentar esta esperanza de revertir la mencionada más que preocupante tendencia a la que se están viendo abocadas muchas regiones de Europa creemos que destaca, especialmente, la urgente necesidad de diseñar de forma sistemática la visualización internacional de su actividad socioeconómica en la web del conocimiento para que los productos de las PYME de este sector puedan comercializarse en mercados anglosajones. Constatamos que el mencionado sector requiere de servicios de redacción y de traducción del español al inglés para adaptar los datos de sus productos (páginas web, fichas descriptivas de producto, fichas técnicas, etiquetado, información nutricional, etc.) al mercado internacional y, así, atraer nuevos consumidores y fidelizar a los ya existentes.

* El presente trabajo ha sido realizado en el marco del proyecto «PROBITEXT 3.0.» (Ref. PID2020-1140644B-I00), «VIP II» (Ref. PID2020-112818GB-I00) y «PROFETA: El entorno PROFEsional del TrAductor: fiscalidad y herramientas informáticas para el ejercicio de la profesión» (ref. PIE 19-33, Tipo B, Universidad de Málaga).

En este contexto, el diseño y desarrollo de herramientas lingüísticas capaces de automatizar dichas actividades constituyen una prioridad, puesto que la traducción del español al inglés en este campo de especialidad define una parcela de conocimiento dentro de los Estudios de Traducción que todavía necesita abordarse con mayor profundidad para satisfacer la creciente demanda de las empresas chacineras que desean internacionalizarse y para las que, dado su perfil y dimensiones, nunca podrían acometer la adaptación al mercado exterior sin el apoyo de una tecnología que agilice y simplifique los procesos de traducción y de redacción multilingüe.

Conscientes de que hace falta profundizar en el estudio de este campo desde una perspectiva interlingüística, el grupo de investigación interuniversitario Análisis Contrastivo y Traducción Inglés-Español (ACTRES),[1] dirigido por la Dr.ª Rabadán (Universidad de León), ha publicado trabajos contrastivos español/inglés basados en corpus que analizan las recetas de cocina (Rabadán *et al.* 2016), los textos promocionales del queso (Labrador y Ramón 2015, Ramón y Labrador 2018), las notas de cata del vino (López Arroyo y Roberts 2014, 2015, 2016, 2017a, 2017b, entre otros) o las fichas descriptivas de embutidos y chacinas (Ortego Antón 2019, 2020, Fernández Nistal 2020, Ortego Antón y Fernández Nistal 2020). Los resultados derivados de estos estudios están permitiendo desarrollar e implementar aplicaciones que pueden integrarse en herramientas lingüísticas basadas en el procesamiento del lenguaje natural, tales como redactores semiautomáticos (BiTextCook, GDQ, FITEVI, GDEGA, PROMOCIONA-Té o GEFEM), que ayudarán a internacionalizar y dotar de visibilidad a las industrias del sector agroalimentario en la web del conocimiento.

Puesto que los productos procedentes del sector chacinero son muy demandados fuera de nuestras fronteras, en este trabajo pretendemos abordar las dificultades que plantea la traducción de la forma que adquiere un determinado embutido tomando como ejemplo el chorizo (sarta, vela,

1. ACTRES = Grupo Interuniversitario ACTRES: Análisis Contrastivo y Traducción Inglés-Español, en línea, <http://actres.unileon.es/?lang=es> [Consulta: 24/11/2020].

cular, etc.) y basando la metodología de análisis en el diseño, la compilación y la explotación de un corpus paralelo (P-GEFEM) y de un corpus comparable (C-GEFEM). Los datos de los corpus nos permitirán detectar si existen patrones de comportamiento típicos de la traducción y si dichos patrones también se reproducen en los textos originalmente redactados en lengua inglesa para, en posteriores etapas, aplicar los resultados al diseño de herramientas basadas en el procesamiento del lenguaje natural, tales como redactores semiautomáticos y programas de traducción asistida por ordenador para su empleo en el mencionado sector, además de incluirse en la formación de futuros profesionales de la traducción.

2. Los corpus virtuales como herramientas al servicio de la actividad traductora

El empleo de una metodología basada en el diseño, compilación y explotación de corpus virtuales,[2] definidos estos como «a corpus in which there are not many texts but that the few texts included are suited to the field of knowledge, genre and textual variety» (Corpas Pastor 2008: 91), está ampliamente extendida en los Estudios de Traducción (Bowker 2002, Laviosa 2002, Corpas Pastor 2008, Beeby *et al.* 2009, Kruger *et al*. 2011, Zanettin 2012, Sánchez Nieto 2015, Corpas Pastor y Seghiri 2017, entre otros):

> Some of the principal advantages of using them [corpora] are their objectivity, their reusability and multiple usage as a single resource. In addition, they are user-friendly and allow access and management to huge quantities of information in almost no time. (Corpas Pastor y Seghiri 2009: 77)

A las ventajas ya expuestas se añaden el acceso y la disponibilidad de textos escritos en una gran variedad de lenguas, características im-

2. Este tipo de corpus también ha recibido otras denominaciones, por ejemplo, *corpus ad hoc* (Corpas Pastor 2001: 164), *disposable corpus*, *do-it-yourself (DIY) corpus* (Zanettin 2002), *domain-specific corpus* (Corpas Pastor 2004: 226), *web corpus* (Fletcher 2004), *electronic corpus* (Corpas Pastor 2001, Varantola 2003), *ephemeral corpus* (Corpas Pastor 2004: 226), *precision corpus* (Varantola 1997) y *special purpose corpus* (Pearson 1998, Sánchez-Gijón 2003a, 2003b).

pensables hace unas décadas, que se han visto propiciadas por el desarrollo de la sociedad del conocimiento y por la globalización de la economía en entornos digitales. Ciertamente hemos sido testigos de un aumento del interés por emplear corpus bilingües y multilingües en los Estudios de Traducción, hasta el extremo de que su utilización se ha convertido en una herramienta casi imprescindible para la práctica de la traducción.

Aunque han surgido infinidad de taxonomías para clasificar los corpus atendiendo a sus características y a su aplicabilidad según el contexto traductológico (Laviosa 1997, Torruella y Llisterri 1999, Corpas Pastor 2001, Granger 2003, Zanettin 2012, entre otros), en este trabajo nos limitaremos a diferenciar y definir qué es un corpus paralelo y un corpus comparable. En este sentido, Rabadán y Fernández Nistal (2002: 51-52) definen los corpus paralelos como «un conjunto de textos fuente (en una lengua de partida) y sus traducciones (a otra/s, la/las de llegada), razón por la que, en ocasiones, se les conoce también con la denominación de corpus de traducciones». Dentro de los corpus paralelos, Olohan (2004: 24) distingue dos tipos: los corpus monodireccionales, formados por textos originales en la lengua A con sus traducciones correspondientes en la lengua B, y los corpus bidireccionales, que no solo recogen textos originales en la lengua A con sus respectivas traducciones en la lengua B, sino que también incorporan textos originales en lengua B con sus respectivas traducciones en lengua A. Además, Rabadán y Fernández Nistal (2002: 51-52) señalan que la principal ventaja de los corpus paralelos se corresponde con el hecho de que cualquier noción expresada en el texto origen debería encontrarse en el texto meta, dado que los textos y sus traducciones están íntimamente relacionados.

Por lo que respecta a los corpus comparables, EAGLES (1996) los define como «one which selects similar texts in more than one language or variety». Entre las características de este tipo de corpus, Rabadán y Fernández Nistal (2002: 53) subrayan que los textos incluidos deben funcionar de forma similar en el plano de la situación

comunicativa; es decir, que recojan contenidos similares, que estén redactados en fechas cercanas y que desempeñen una función semejante desde el punto de vista discursivo. Estas autoras destacan que este tipo de corpus presenta como ventaja el hecho de que la lengua de llegada no está influenciada por la lengua de partida, lo que es muy frecuente en el caso de las traducciones, por lo que son muy útiles para el estudio de la terminología y de la fraseología especializada y, además, constituyen una fuente muy valiosa para los estudios contrastivos en general y del léxico especializado en particular.

No obstante, el uso de corpus comparables no está exento de problemas:

> En primer lugar, desde el punto de vista semántico, los textos que forman parte de un corpus comparable no están íntimamente relacionados, como sucede en el caso de los corpus paralelos, por lo que no tienen por qué expresar las mismas nociones. En consecuencia, el lexicógrafo o el traductor no siempre encontrará allí los equivalentes que busque. (Ortego Antón 2019: 53)

A pesar de estas dos restricciones, los corpus comparables, como señala Sánchez Trigo (2005: 138), son de gran utilidad durante la actividad traductora:

> [...] constituyen una herramienta interesante para solucionar problemas de diferente naturaleza (temáticos, terminológicos, textuales, estilísticos, etc.) [...] ya que permiten compilar una documentación fiable y específica de manera económica (en tiempo y coste) y muy eficaz. (Sánchez Trigo 2005: 138)

Corpas Pastor (2012: 11) también subraya los beneficios de utilizar corpus comparables en traducción, puesto que proporcionan a los usuarios un inventario amplio de formas convencionales de expresar unidades de sentido y funciones específicas en un determinado registro y forma textual en la lengua de origen y en la lengua meta, que se pueden considerar equivalentes (en diverso grado). Igualmente, el uso de corpus comparables proporciona modelos en la producción del texto meta, ayuda a resolver problemas, a tomar decisiones y a validar las soluciones adoptadas en el proceso de la traducción.

Asimismo, en los Estudios de Traducción son varios los trabajos que utilizan una metodología basada en corpus en la que contrastar los datos procedentes de los corpus paralelos en corpus comparables (Calzada Pérez *et al*. 2006, Bernardini 2007, Frankenberg-García 2017, Rabadán 2019, Marco 2019, Ortego Antón y Fernández Nistal 2020, entre otros). Así pues, una vez expuestas las ventajas del uso de corpus virtuales paralelos y comparables como marco metodológico aplicable en los Estudios de Traducción, procedemos a describir el diseño, compilación y explotación del corpus virtual paralelo, P-GEFEM, y del corpus virtual comparable, C-GEFEM, para obtener una aproximación a los patrones empleados para trasvasar del español al inglés la forma que adquiere el chorizo.

3. Diseño y compilación de P-GEFEM y de C-GEFEM

3.1. P-GEFEM

P-GEFEM es un corpus virtual paralelo, bilingüe (español e inglés) y monodireccional que se ha diseñado y compilado siguiendo el protocolo descrito por Seghiri (2017). Dicho protocolo se compone de cuatro fases: búsqueda, descarga, formato y almacenamiento. En primer lugar, hemos localizado los textos que formarán nuestro corpus a partir de la búsqueda de fichas de chorizo en páginas web de industrias agroalimentarias españolas, por ejemplo, Beher,[3] Incarlopsa,[4] Ibéricos Montellano,[5] etc., cuyo contenido está disponible en las lenguas española e inglesa. Los textos están redactados originalmente en español y en lengua inglesa se ofrecen las traducciones de dichos textos. A continuación, hemos descargado cada ficha de producto de forma manual y la hemos guardado como archivo en texto plano (.txt ANSI) dentro de una carpeta denominada P-GEFEM que, a su vez, se organiza en dos subcarpetas: español (ES) e

3. BEHER, en línea, <https://beher.com/> [Consulta: 26/11/2020].
4. Incarlopsa, en línea, <http://www.incarlopsa.es/> [Consulta: 26/11/2020].
5. Ibéricos Montellano, en línea, <http://www.ibericosmontellano.com/> [Consulta: 26/11/2020].

inglés (EN). Por lo que respecta a la denominación que conferimos a los archivos, estos se guardan correlativamente con el número (101, 102, 103, etc.), el género textual (DM = *dried meats* en lengua inglesa), el origen (WS = web), dos iniciales con la denominación de la empresa (IM = Ibéricos Montellano), el tipo textual, según sea descripción de producto (PD = *product description*) o texto publicitario (AT = *advertisement text*), la fecha (AAMMDD), el campo (*Foodie*) y la lengua (ES o EN), por ejemplo, 146DMwsIMPD171219FoodieES.txt. En lo relativo al tamaño, P-GEFEM está compuesto de 100 textos en español (18.043 palabras o *tokens*) y de 100 textos en inglés (16.482 palabras o *tokens*). Todos los textos se corresponden con fichas de chorizo.

Además, P-GEFEM reúne dos características. Por un lado, es equilibrado, es decir, «the proportions of different kinds of text it contains should correspond with informed and intuitive judgements» (Sinclair 2005: 8). El equilibrio de P-GEFEM se consigue combinando dos factores, a saber, la fecha, dado que los textos se publicaron en 2017 y 2018, y la procedencia, ya que se han incluido más de 30 empresas. La segunda de las características se corresponde con la representatividad, que puede ser de dos tipos: cualitativa y cuantitativa. Respecto a la representatividad cualitativa, P-GEFEM la cumple por el riguroso proceso de compilación seguido atendiendo a los postulados de Seghiri (2017): el corpus se ha compilado seleccionando textos de un mismo género textual con sus respectivas traducciones procedentes de diversas empresas con la información disponible en la web y atendiendo a la diversificación tanto en el género textual (descripción de producto y texto publicitario) como al número de textos procedentes de cada empresa, con un máximo de tres textos por empresa. Para comprobar si P-GEFEM es representativo cuantitativamente utilizamos el programa ReCor, diseñado por Seghiri (2006 y 2015) y Corpas Pastor y Seghiri (2009 y 2010). En este sentido, Seghiri (2017: 50) argumenta que «para determinar la representatividad cuantitativa de la muestra, es decir, si se ha cu-

bierto la terminología básica empleada en este género (...)», ReCor analiza cada subcorpus y genera dos gráficos de representatividad. En el primer gráfico se ofrece el número de documentos necesarios para conseguir dicha representatividad y, en el segundo, se indica el número de palabras o *tokens* que debe tener dicho corpus.

Los gráficos resultantes de analizar la representatividad cuantitativa de P-GEFEM con ReCor se muestran en la Figura 1 para el subcorpus en español y en la Figura 2 para el subcorpus en inglés.

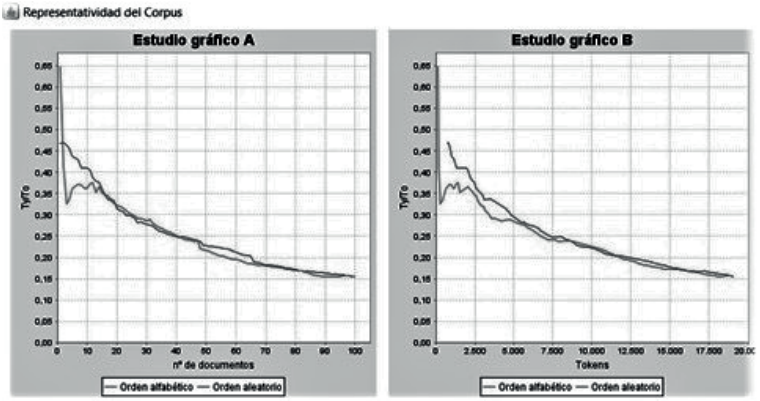


Figura 1. Representatividad cuantitativa del subcorpus en español de P-GEFEM.

En la Figura 1 se aprecia que en el subcorpus en español la representatividad se alcanza a partir de 70 textos y 15.000 palabras o *tokens*. Por su parte, en el subcorpus en inglés (Figura 2), dicha representatividad se consigue con un menor número de textos y de palabras, a partir de 50 documentos y 11.000 palabras.

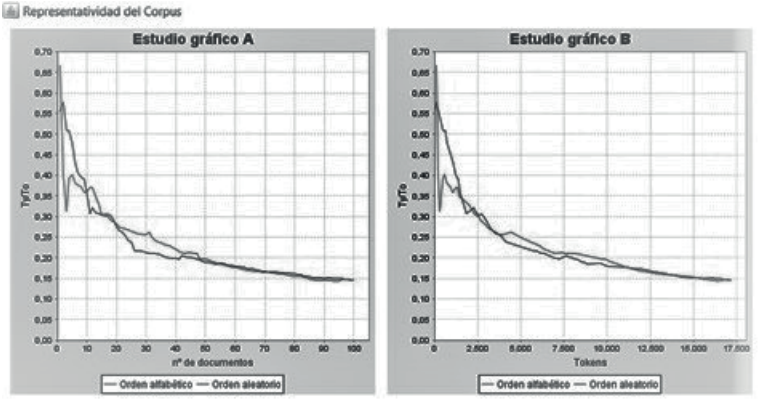


Figura 2. Representatividad cuantitativa del subcorpus en inglés de P-GEFEM.

Una vez asegurada la representatividad cualitativa y cuantitativa podemos afirmar que estamos ante un corpus representativo, fiable y de calidad en el que será posible detectar los patrones de comportamiento de la terminología y de su fraseología.

3.2. C-GEFEM

C-GEFEM (Ortego Antón y Fernández Nistal 2020) es un corpus virtual comparable español-inglés de fichas descriptivas de embutidos que forma parte del catálogo de corpus del grupo interuniversitario ACTRES. Se compiló con la finalidad de analizar y contrastar la estructura retórica y el comportamiento fraseológico del léxico de este género.[6] En su diseño, la selección de textos que componen cada uno de los subcorpus se efectuó teniendo en cuenta los siguientes criterios: la disponibilidad, el tamaño, la representatividad, el equilibrio y el perfil de usuario.

6. Para profundizar en los detalles de su compilación y obtener más información sobre los textos y la naturaleza de estos puede consultarse Ortego Antón (2019: 62-66).

Por lo que respecta a la disponibilidad, los textos se han extraído de páginas web de reconocidas empresas del sector (más de 40 en inglés y 30 en español), redactados originariamente en español o en inglés, lo que garantiza su autenticidad. C-GEFEM está compuesto por 100 textos en cada lengua, con un total de 24.604 palabras o *tokens* en inglés y 13.358 palabras o *tokens* en español. Aunque se puede llegar a pensar que el corpus no es equilibrado en lo relativo al tamaño, la variación en el número de palabras o *tokens* se debe a que las fichas de embutidos en inglés son mucho más complejas y extensas que en español (Ortego Antón 2020: 180). La representatividad cualitativa se ha conseguido combinando dos factores, la fecha, pues todos los textos datan de 2016, y la procedencia, mediante la inclusión de textos de diferentes zonas geográficas: Reino Unido, Estados Unidos, Canadá, Irlanda y Australia para la lengua inglesa y español de Europa para el subcorpus en español. La representatividad cuantitativa se determinó con ReCor siguiendo un procedimiento similar al llevado a cabo con P-GEFEM.

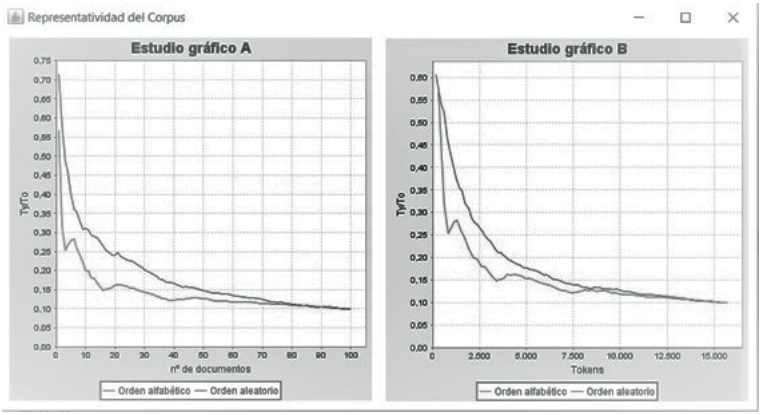


Figura 3. Representatividad cuantitativa del subcorpus en español de C-GEFEM.

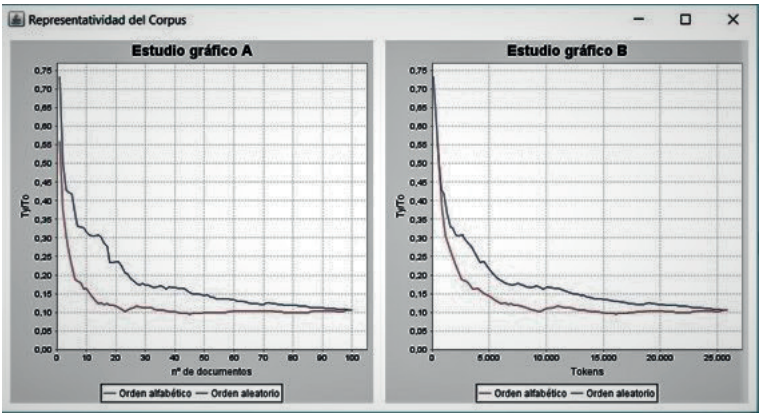


Figura 4. Representatividad cuantitativa del subcorpus en inglés de C-GEFEM.

4. Metodología de análisis

Para detectar los patrones de traducción que los traductores profesionales siguen en el trasvase interlingüístico de la forma del chorizo del español al inglés, en primer lugar, tenemos que delimitar la muestra de análisis. Para ello, nos asistimos del analizador de concordancias gratuito AntConc 3.7.8. (Anthony 2018). Cargamos el subcorpus en español de P-GEFEM y con la función «Clusters/N-GRAMS» obtenemos las colocaciones[7] más frecuentes de chorizo, como se muestra en la Figura 5.

7. En este trabajo emplearemos *colocación* como coaparición de palabras (Sinclair 2001: 170).

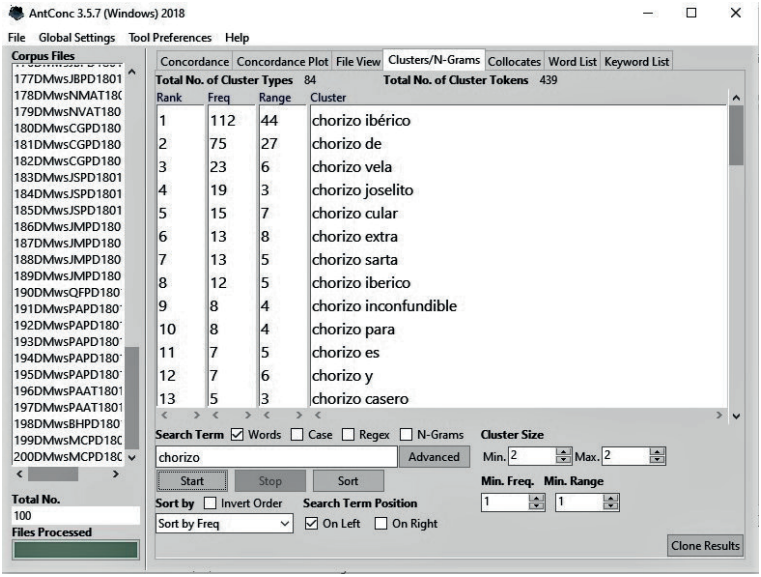


Figura 5. Colocaciones más frecuentes de chorizo en el subcorpus en español de P-GEFEM.

Manualmente hemos revisado las colocaciones y hemos extraído únicamente aquellas colocaciones que hacen referencia a las distintas formas que pueden adquirir los embutidos. Los resultados se muestran en la Tabla 1.

Colocación
~ atado
~ cular
~ herradura
~ sarta
~ vela
ristra de ~

Tabla 1. Muestra de análisis.

A continuación, seleccionamos la herramienta que nos permita alinear P-GEFEM, que se corresponde con ParaConc (Moropa 2007), uno de los principales programas de concordancias diseñado para realizar análisis contrastivos. Entre sus principales funciones destaca la alineación de los textos paralelos en cada una de las lenguas (hasta cuatro subcorpus), la búsqueda de las unidades léxicas en contexto (*KWIC*) y la obtención de un listado de las unidades léxicas ordenadas por frecuencia, así como de las colocaciones de un término.

Cargaremos P-GEFEM en ParaConc (Moropa 2007) y alinearemos los dos subcorpus utilizando la opción *New Line Delimiter,* como se muestra en la Figura 6.

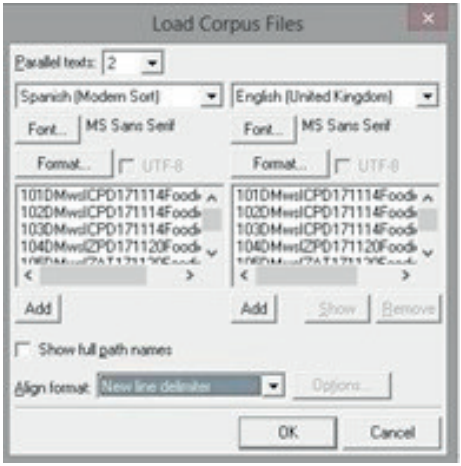


Figura 6. Alineación del corpus con ParaConc (Moropa 2007).

Cotejamos los equivalentes para las unidades objeto de estudio relativas a la forma que adquiere el chorizo y la interfaz nos muestra la unidad léxica en contexto (*KWIC*) y la traducción en la parte inferior de la pantalla, como se puede observar en la Figura 7.

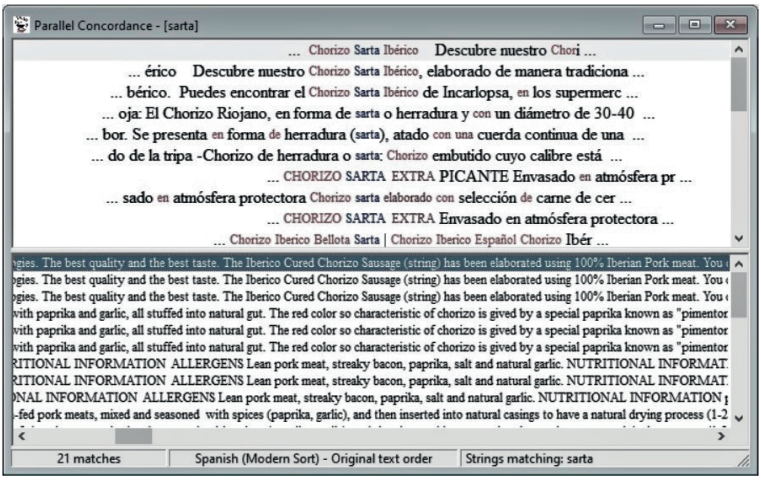


Figura 7. Concordancias de sarta y sus equivalentes en P-GEFEM.

Los resultados que se desprendan de las búsquedas se ordenarán por frecuencia acompañados de sus equivalentes, junto al número de ocurrencias y el porcentaje que dichos equivalentes representan sobre el total.

En una segunda fase, compararemos los datos resultantes de P-GEFEM con los que se desprendan de C-GEFEM, que explotaremos con AntConc 3.5.7. (Anthony 2018) utilizando una metodología similar a la empleada por Ortego Antón (2019: 77-90). Primeramente, verificaremos en el subcorpus en inglés de C-GEFEM si se emplean los equivalentes propuestos en P-GEFEM, para así confirmar si los equivalentes obtenidos son típicos de las traducciones o, por el contrario, también se emplean en los textos redactados originalmente en lengua inglesa. A continuación, con la función «Clusters/N-GRAMS» de AntConc 3.5.7. (Anthony 2018) obtendremos las colocaciones más frecuentes del chorizo en el subcorpus en inglés de C-GEFEM, revisaremos manualmente dicho listado y extraeremos las colocaciones que hacen referencia a la forma del chorizo para vislumbrar qué patrones se utilizan en los textos redactados originalmente en lengua inglesa.

5. El trasvase del español al inglés de la forma del chorizo

Siguiendo la metodología descrita, hemos introducido cada unidad léxica objeto de análisis en ParaConc (Moropa 2007) y esta herramienta nos ofrece los equivalentes que se emplean en las traducciones al inglés de las fichas de chorizo que constituyen P-GEFEM. Los resultados de comprobar cómo se traducen del español al inglés las unidades léxicas que constituyen la muestra de análisis se ofrece en la Tabla 2. En dicha tabla presentamos las unidades léxicas en español, el número de ocurrencias en el subcorpus en español de P-GEFEM, la frecuencia normalizada sobre 100, los diversos equivalentes empleados por los traductores, el número de ocurrencias de dichos equivalentes en el subcorpus en inglés de P-GEFEM, la frecuencia normalizada sobre 100, el porcentaje que dicha ocurrencia representa sobre el total de equivalentes y la técnica de traducción empleada siguiendo la categorización propuesta por Molina y Hurtado Albir (2002).

ES			**EN**				**Técnica de traducción**
Unidad léxica	Ocur.	Frec.	Equivalente	Ocur.	%	Frec.	
cular	32	0,18	cular	23	71,43	0,14	préstamo
			-	7	21,43	0,04	omisión
			thick-casing	2	7,14	0,01	equivalente establecido
vela	27	0,15	vela	27	100	0,16	préstamo
herradura	23	0,13	horseshoe	15	65,22	0,09	traducción literal
			herradura	5	21,74	0,03	préstamo
			-	3	13,04	0,02	omisión
sarta	21	0,12	u-shaped	7	33,33	0,04	equivalente establecido
			sarta	5	23,81	0,03	préstamo
			-	4	19,05	0,02	omisión
			string	3	14,29	0,02	traducción literal
			strung	2	9,52	0,01	traducción literal
atado	5	0,03	tied	2	40,00	0,01	equivalente establecido
			atado	1	20,00	0,01	préstamo
			bound	1	20,00	0,01	traducción literal
			stapled	1	20,00	0,01	traducción errónea
ristra	3	0,02	string	2	66,67	0,01	traducción literal
			ristra	1	33,33	0,01	préstamo

Tabla 2. Equivalentes de traducción de las formas que adquieren los embutidos en P-GEFEM.

De los resultados expuestos en la Tabla 2 se desprende la diversidad de equivalentes empleados por los traductores a la hora de trasvasar la forma que adquiere el chorizo a la lengua inglesa, puesto que constatamos la utilización de hasta cinco equivalentes distintos en el caso de 'sarta'. Dentro de la diversidad de opciones empleadas por los traductores, advertimos que en todos los casos se usa la técnica del préstamo, es decir, el término empleado en español también se utiliza en lengua inglesa, aspecto que habrá que contrastar con el empleo que de estas unidades léxicas se haga por parte de los usuarios que redactan textos originalmente en lengua inglesa con el análisis de los resultados del subcorpus en inglés de C-GEFEM.

De forma similar, detectamos que en múltiples ocasiones se tiende a omitir la descripción de la forma que adquiere el chorizo, probablemente porque es un obstáculo para traductores y redactores bilingües, así que optan por emplear la omisión como técnica de traducción en lugar de abordar el problema de traducción, como se muestra en los casos de 'cular', 'herradura' y 'sarta'.

Otra tendencia que observamos se corresponde con el empleo de la técnica de la traducción literal, es decir, se utiliza un equivalente que tiene el mismo significado con connotaciones diferentes y, en consecuencia, no se usa dicho equivalente en el discurso de los embutidos redactado por usuarios nativos de lengua inglesa, por ejemplo, 'herradura' por *horseshoe*, 'sarta' por *string/strung* y 'atado' por *bound*.

Una vez obtenidos los patrones de comportamiento en los textos traducidos, hemos procedido a triangular los datos y corroborar los patrones de traducción en C-GEFEM.

5.1. C-GEFEM

Con la ayuda del analizador de concordancias AntConc 3.5.7. (Anthony 2018), hemos verificado si los equivalentes propuestos en P-GEFEM se utilizan en el subcorpus en inglés de C-GEFEM.

Nos llama la atención que la mayoría de los equivalentes de traducción utilizados para describir la forma que adquieren los embutidos

no se emplean en los textos redactados originalmente en lengua inglesa. En este sentido, no hemos encontrado ninguna ocurrencia para 'cular', *thick-casing*, 'herradura', *u-shaped*, *string/strung*, *tied*, 'atado', *bound*, *stapled*, *string* o 'ristra'. De hecho, solo hemos obtenido dos ocurrencias de 'vela', una ocurrencia de *horseshoe* acompañando a *shaped* y siete ocurrencias de 'sarta', siempre empleando esta unidad léxica en la denominación del producto.

Con estos resultados, podemos afirmar que los traductores han utilizado equivalentes de traducción que no son funcionales en la cultura anglosajona, dado que el contraste de resultados en el subcorpus en inglés de C-GEFEM pone de manifiesto que dichos equivalentes no se emplean en lengua inglesa.

Ante este panorama, cargamos en AntConc 3.5.7. (Anthony 2018) el subcorpus en inglés de C-GEFEM y con la función «Clusters/N-GRAMS» hemos realizado una búsqueda para obtener el listado de las colocaciones más frecuentes de *chorizo*. Manualmente hemos examinado los resultados y observamos que *chorizo ring* tiene 19 ocurrencias y *chorizo slice,* seis. Así pues, estos resultados ponen de manifiesto que para designar la forma que adquiere el embutido en lengua inglesa se emplean estas dos unidades léxicas: *ring* y *slices*, según se comercialice el producto entero en una sola pieza o en lonchas.

A la vista de la triangulación de los resultados obtenidos tras la explotación de un corpus paralelo (P-GEFEM) con los que se desprenden del análisis del corpus comparable (C-GEFEM) durante el trasvase interlingüístico de las unidades léxicas empleadas para designar la forma que adquieren los embutidos, podemos afirmar que existen patrones de comportamiento típicos de los textos traducidos. En este sentido, detectamos que las distintas formas del chorizo se tienden a trasvasar a la lengua inglesa por medio de omisiones, de préstamos o de traducciones literales, en tanto que en los textos redactados originalmente en lengua inglesa se suele optar por dos tipos de formas: *ring* para denominar al embutido entero o *sliced* si dicho embutido se comercializa en lonchas. En consecuencia, a la hora de trasvasar la

forma del producto habrá que tener muy en cuenta la normativa de comercialización, puesto que los textos redactados originalmente en inglés muestran un uso mucho más reducido de términos que los textos traducidos.

6. Conclusiones

Las traducciones al inglés de los textos pertenecientes al género textual de las fichas de producto revelan, cuando menos, la escasa atención prestada a las traducciones de las páginas web de las empresas españolas y, más específicamente, a la descripción de los productos, como constatan Ortego Antón (2019) y Fernández Nistal (2020). La disparidad en el uso de equivalentes para hacer referencia a la forma del chorizo entre el inglés traducido y el uso efectivo en la lengua inglesa plantea, por ejemplo, la urgencia de concienciar a las empresas respecto al deterioro de la imagen corporativa que puede implicar para estas el empleo de una traducción que no se ajuste a las convenciones de la lengua inglesa. Así pues, consideramos necesario colaborar con las empresas del sector agroalimentario para que ajusten la calidad lingüística de sus páginas web y de la información que acompaña a sus productos al nivel de excelencia de los mencionados productos, a los que pretenden dotar de visibilidad internacional en la web de conocimiento.

Asimismo, la utilización de los corpus virtuales bilingües, tanto paralelos como comparables, permite demostrar a los traductores profesionales y a los traductores en formación hasta qué punto su consulta contribuye, entre otros aspectos, a tomar conciencia de las dificultades de equivalencia cultural en este sector y a observar y analizar las decisiones sobre normas y uso terminológico adoptadas por otros profesionales para aprender a implementar una práctica más rigurosa y apropiada para los proyectos de traducción.

También nos gustaría destacar la importancia del uso de los corpus para la creación de glosarios especializados que, a su vez, sirvan para la práctica profesional, pues es evidente que el traductor profesional no

dispone, en muchos casos, de suficiente tiempo y de medios para compilar un corpus virtual y observar las ocurrencias y las colocaciones.

Por lo demás, la práctica en el aula de este tipo de estudios (Bestué y Rodríguez Inés 2020: 130; Seghiri 2020) en la fase documental de los proyectos de traducción y exige la utilización de las herramientas tecnológicas disponibles en la actualidad que desembocarán en la mayor rapidez y precisión en sus decisiones como traductores. La combinación adecuada de las diferentes técnicas que se incorporan en esta aproximación metodológica en el estudio y manejo del léxico especializado en la formación de traductores resultará de vital importancia no solo para la traducción en el sector que hemos esbozado en este breve estudio de caso, sino en la mayoría de los sectores económicos.

En definitiva, los resultados obtenidos en el presente estudio de caso ponen en valor el empleo este tipo de técnicas de análisis y de metodología en la traducción profesional, así como en la formación de futuros traductores en el diseño y compilación de corpus virtuales paralelos. Asimismo, ponen de manifiesto la necesidad de contrastar los resultados obtenidos en corpus comparables para que puedan alimentar las bases de datos terminológicas de uso en los entornos de traducción asistida por ordenador (TAO).

Referencias bibliográficas

Anthony, Laurence (2018): *AntConc (Version 3.5.7)* Tokyo, Waseda University [programa informático]. En línea, <http://www.laurenceanthony.net/software> [Consulta: 17/03/2018].

Beeby, Allison; Rodríguez Inés, Patricia; Sánchez-Gijón, Pilar (2009): *Corpus Use and Translating,* Amsterdam/Philadelphia, John Benjamins, pp. 75-107.

Bernardini, Silvia (2007): «Collocations in Translated Language: Combining Parallel, Comparable and Reference Corpora», en Matthew Davies, Paul Rayson, Susan Hunston, Pernilla Danielsson (eds.), *Proceedings of the Corpus Linguistics Conference CL2007 (27-30 July 2007),* Birmingham, University of Birmingham, pp. 1-16.

Bestué, Carmen; Rodríguez Inés, Patricia (2020): «Metodología de corpus en la enseñanza de la traducción de cuentas anuales (inglés-español/catalán): un activo intangible», en Miriam Seghiri (ed.), *La lingüística de corpus aplicada al desa-*

rrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas, Berlin, Peter Lang, pp. 107-132.

Bowker, Lynne (2002): *Computer-Aided Translation Technology: A Practical Introduction*, Ottawa, University of Ottawa Press.

Calzada Pérez, María *et al.* (2006): «ECPC (European Parliamentary Comparable and Parallel Corpora)», *Procesamiento del Lenguaje Natural,* vol. 37, pp. 349-350.

Corpas Pastor, Gloria (2001): «Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada», *TRANS. Revista de Traductología*, vol. 5, pp. 155-184.

Corpas Pastor, Gloria (2004): «Localización de recursos y compilación de corpus vía Internet: Aplicaciones para la didáctica de la traducción médica especializada», en Consuelo Gonzalo García, Valentín García Yebra (eds.), *Manual de documentación y terminología para la traducción especializada*, Madrid, Arco Libros, pp. 223-257.

Corpas Pastor, Gloria (2008): *Investigar con corpus en traducción: los retos de un nuevo paradigma,* Frankfurt am Main, Peter Lang.

Corpas Pastor, Gloria (2012): «Corpus, Tecnología y Traducción», en María García Antuña (ed.), *XII Jornadas de Lingüística*, Cádiz, Servicio de Publicaciones de la Universidad de Cádiz, pp. 2-21.

Corpas Pastor, Gloria; Seghiri, Miriam (2009): «Virtual Corpora as Documentation Resources: Translating Travel Insurance Documents (English-Spanish)», en Allison Beeby, Patricia Rodríguez Inés, Pilar Sánchez-Gijón (eds.), *Corpus Use and Translating*, Amsterdam/Philadelphia, John Benjamins, pp. 75-107.

Corpas Pastor, Gloria; Seghiri, Miriam (2010): «Size Matters: A Quantitative Approach to Corpus Representativeness», en Rosa Rabadán *et al.* (eds.), *Lengua, traducción, recepción: en honor de Julio César Santoyo*, León, Universidad de León, Área de Publicaciones, pp. 111-145.

Corpas Pastor, Gloria; Seghiri, Miriam (2017): *Corpus-based Approaches to Translation and Interpreting. From Theory to Applications*, Wien, Peter Lang.

EAGLES (1996): *Preliminary Recommendations on Corpus Typology* [documento técnico EAGLES EAG-TCWG-CTYP/P]. En línea, <http://www.ilc.cnr.it/EAGLES/corpustyp/corpustyp.html> [Consulta: 17/03/2019].

Fernández Nistal, Purificación (2020): «Los corpus como herramienta de traducción para los traductores e intérpretes del siglo XXI: el caso del chorizo ibérico de bellota», en Susana Álvarez Álvarez, María Teresa Ortego Antón (eds.), *Perfiles estratégicos de traductores e intérpretes en la transmisión de la información experta multilingüe en la sociedad del conocimiento del siglo XXI*, Granada, Comares, pp. 143-160.

Fletcher, William H. (2004): «Facilitating the Compilation and Dissemination of Ad-Hoc Web Corpora», en Guy Aston, Silvia Bernardini, Dominic Steward

(eds.), *Corpora and Language Learners,* Amsterdam/Philadelphia, John Benjamins, pp. 273-300 [DOI: https://doi.org/10.1075/scl.17.21fle].

Frankenberg-García, Ana (2017): «A Corpus Study of Loans in Translated and Non-Translated Texts», en Gloria Corpas Pastor, Miriam Seghiri (eds.), *Corpus-based Approaches to Translation and Interpreting. From Theory to Applications*, Wien, Peter Lang, pp. 19-42.

Granger, Sylviane (2003): «The Corpus Approach: A Common Way Forward for Contrastive Linguistics and Translation Studies?», en Sylviane Granger, Jacques Lerot, Stephanie Petch-Tyson (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, Amsterdam/New York, Editions Rodopi, pp. 17-30.

Kruger, Alet *et al.* (2011): *Corpus-Based Translation Studies. Research and Application*, London, Continuum.

Labrador, Belén; Ramón, Noelia (2015): «"Perfectly smooth creamy and full flavoured": Online cheese descriptions», *Procedia: Social and Behavioural Sciences*, n. 198, pp. 226-232.

Laviosa, Sara (1997): «How Comparable can "Comparable Corpora" Be?», *Target*, vol. 9, n. 2, pp. 289-319.

Laviosa, Sara (2002): *Corpus-based Translation Studies. Theory, Findings, Applications*, Amsterdam/New York, Rodopi.

López Arroyo, Belén; Roberts, Roda P. (2014): «English and Spanish Descriptors in Wine Tasting Terminology», *Terminology*, vol. 20, n. 1, pp. 25-49 [DOI: 10.1075/term.20.1.02lop].

López Arroyo, Belén; Roberts, Roda P. (2015): «Unusual Sentence Structure in Wine Tasting Notes: A Contrastive Corpus-Based Study», *Language in Contrast*, vol. 15, n. 2, pp. 162-180.

López Arroyo, Belén; Roberts, Roda P. (2016): «Differences in Wine Tasting Notes in English and Spanish», *Babel,* vol. 62, n. 3, pp. 370-401.

López Arroyo, Belén; Roberts, Roda P. (2017a): «El lenguaje metafórico en las fichas de cata de vino en inglés y en español», *Hermēneus: Revista de Traducción e Interpretación,* n. 19, pp. 139-163.

López Arroyo, Belén; Roberts, Roda P. (2017b): «Genre and Register in Comparable Corpora: An English / Spanish Contrastive Analysis», *Meta*, vol. 62, n. 1, pp. 114-136.

Marco, Josep (2019): «Living with Parallel Corpora. The Potentials and Limitations of Their Use in Translation Research», en Irene Doval, M. Teresa Sánchez Nieto (eds.), *Parallel Corpora for Contrastive and Translation Studies. New resources and applications*, Amsterdam/Philadelphia, John Benjamins, pp. 39-56.

MAPA (2020): *Marco Estratégico para la Industria de Alimentación y Bebidas. Alimentamos futuro, 2020*, Madrid, Ministerio de Agricultura, Pesca y Alimentación. En línea, <https://www.mapa.gob.es/es/alimentacion/temas/industria-agroa-

limentaria/informemarcoestrategicoparalaiab_tcm30-87439.PDF> [Consulta: 24/02/2020].

Molina, Lucía; Hurtado Albir, Amparo (2002): «Translation Techniques Revisited: A Dynamic and Functionalist Approach», *Meta*, vol. 47, n. 4, pp. 498-512 [DOI: https://doi.org/10.7202/008033ar].

Moropa, Koliswa (2007): «Analysing the English-Xhosa Parallel Corpus of Technical Texts with ParaConc: A Case Study of Term Formation Processes», *Southern African Linguistics and Applied Language Studies*, vol. 25, n. 2, pp. 183-205.

Olohan, Maeve (2004): *Introducing Corpora in Translation Studies*, New York, Routledge.

Ortego Antón, María Teresa (2019): *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*, Berlin, Peter Lang.

Ortego Antón, María Teresa (2020): «Las fichas descriptivas de embutidos en español y en inglés: Un análisis contrastivo de la estructura retórica basado en corpus», *Revista Signos. Estudios de Lingüística*, vol. 53, n. 102, pp. 170-194 [DOI: 10.4067/S0718-09342020000100170].

Ortego Antón, María Teresa; Fernández Nistal, Purificación (2020): «Estudio contrastivo de la terminología de embutidos en inglés y en español con ParaConc y tlCorpus a partir del corpus paralelo P-GEFEM y del comparable C-GEFEM», en Miriam Seghiri (ed.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*, Bern, Peter Lang, pp. 23-48.

Pearson, Jennifer (1998): *Terms in Context. Studies in Corpus Linguistics*, Amsterdam/Philadelphia, John Benjamins.

Rabadán, Rosa (2019): «Working with Parallel Corpora. Usefulness and Usability», en Irene Doval, M.ª Teresa Sánchez Nieto (eds.), *Parallel Corpora for Contrastive and Translation Studies. New Resources and Applications*, Amsterdam/Philadelphia, John Benjamins, pp. 57-78.

Rabadán, Rosa; Fernández Nistal, Purificación (2002): *La traducción inglés-español: fundamentos, herramientas, aplicaciones*, León/Valladolid, Universidad de León/ITBYTE.

Rabadán, Rosa *et al.* (2016): «Bi-Texting Your Food: Helping the Gastro Industry Reach the Global Market», en Antonio Moreno Ortiz, Chantal Pérez Hernández (eds.), *CILC. 8th International Conference on Corpus Linguistics*, pp. 361-371 [DOI: 10.29007/4xtp].

Ramón, Noelia; Labrador, Belén (2018): «Selling Cheese Online», *Terminology*, vol. 24, n. 2, pp. 210-235.

Sánchez-Gijón, Pilar (2003a): «És la web pública la nova biblioteca del traductor?», *Tradumàtica: Traducció i tecnologies de la informació i la comunicación*, n. 2. En línea, <https://ddd.uab.cat/record/4873> [Consulta: 15/01/2020].

Sánchez-Gijón, Pilar (2003b): *Els documents digitals especialitzats: utilització de la lingüística de corpus com a front de recursos per a la traducción*, Barcelona, Universidad Autónoma de Barcelona, Tesis doctoral dirigida por Mariana Orozco Jutorán.

Sánchez Nieto, María Teresa (2015): *Corpus-based Translation and Interpreting Studies: From Description to Application / Estudios traductológicos basados en corpus: de la descripción a la aplicación*, Berlin, Frank und Timme.

Sánchez Trigo, Elena (2005): «Investigación traductológica en la traducción científica y técnica», *TRANS: Revista de Traductología*, n. 9, pp. 131-150.

Seghiri, Miriam (2006): *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*, Málaga, Universidad de Málaga, Tesis doctoral dirigida por Gloria Corpas Pastor. En línea, <http://hdl.handle.net/10630/2715> [Consulta: 15/01/2020].

Seghiri, Miriam (2015): «Determinación de la representatividad cuantitativa de un corpus *ad hoc* bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos», en María Teresa Sánchez Nieto (ed.), *Corpus-based Translation and Interpreting Studies: From Description to Application / Estudios traductológicos basados en corpus: de la descripción a la aplicación*, Berlin, Frank und Timme, pp. 125-146.

Seghiri, Miriam (2017): «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores», *Babel*, vol. 63, n. 1, pp. 43-64 [DOI: 10.1075/babel.63.1.04seg].

Seghiri, Miriam (2020): *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas*, Berlin, Peter Lang.

Sinclair, John (2001): *Corpus, Concordance, Collocation*, Cambridge, Cambridge University Press.

Sinclair, John (2005): «Corpus and Text-Basic Principles», en Martin Wynne (ed.), *Developing Linguistic Corpora: A Guide to Good Practice,* Oxford, Oxbow Books, pp. 1-16.

Torruella, Joan; Llisterri, Joaquim (1999): «Diseño de corpus textuales y orales», en José Manuel Blecua, Gloria Clavería, Carlos Sánchez, Joan Torruella (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Barcelona, Milenio/Universitat Autonoma de Barcelona, pp. 45-77.

Varantola, Krista (1997): «Translators, Dictionaries and Text Corpora», en Silvia Bernardini, Federico Zanettin (eds.), *I corpora nella didattica della traduzione,* Bologna, CLUEB, pp. 117-133.

Varantola, Krista (2003): «Translators and Disposable Corpora», en Federico Zanettin, Silvia Bernardini, Dominic Stewart (eds.), *Corpora in Translator Education*, London/New York, Routledge, pp. 55-70.

Zanettin, Federico (2002): «DIY Corpora: The WWW and the Translator», en Belinda Maia, Johan Haller, Margherita Ulyrck (eds.), *Training the Language Services Provider for the new Millennium*, Porto, Facultade de Letras, Universidade do Porto, pp. 239-248.

Zanettin, Federico (2012): *Translation Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*, Manchester/Kinderhook, St. Jerome Publishing.

La traducción médica (inglés-español): una propuesta de clasificación y evaluación de recursos lexicográficos en línea y en papel*

Lucía Escribano Meseguer
Universidad de Málaga

Miriam Seghiri
Universidad de Málaga

1. Introducción

Tal y como apunta Sánchez Ramos (2017: 179), «[l]a documentación, ya sea en formato papel o electrónico, se supone como uno de los pilares básicos de la formación del traductor e intérprete». Ahora bien, siguiendo a Seghiri (2006: 35), «[l]a multimalla mundial supone, así pues, en la actualidad, no sólo la mayor biblioteca mundial que ha existido jamás, sino la que abre las puertas a la información de forma más rápida y cómoda. Sin embargo, esta que podría ser [...] la Alejandría ideal soñada por cualquier traductor presenta un gran número de inconvenientes que pueden llegar a disminuir su utilidad e, incluso, a obstaculizar la práctica traductora».

Por ello, en este trabajo pretendemos presentar los principales recursos lexicográficos disponibles y de calidad, tanto en red como en papel, para la traducción de las especialidades médicas más demanda-

El presente trabajo se enmarca en el seno de las redes «T3: Terminología, Técnica y Traducción», «INTEC» y «TACTRAD» (Ref. 719/2018), coordinadas por la Universidad de Málaga (España), así como en el marco de los proyectos «INMOCOR» (Ref. P20-00109), «VIP II» (Ref. PID2020-112818GB-I00), «Training network on language technologies for interpreters» (Ref. EUIN2017-87746), «Aplicación de técnicas avanzadas de PLN al ámbito de las Tecnologías de la Traducción y la Interpretación» (Ref. EQC2018-004572-P), «MI4ALL» (AndalucíaTech), «Introducción al mercado profesional de la traducción especializada» (Acción 731, UMA 2020), «POSTrad IV» (UVa), «TRIAJE» (Ref. UMA18-FEDERJA-067) y «PROFETA» (Ref. PIE19-033, UMA).

das en el ámbito sanitario español y que suponen el mayor volumen de trabajo de los traductores que, como justificaremos más adelante, son la obstetricia y ginecología, la cirugía ortopédica y traumatología, la cardiología, la oncología, la cirugía y la farmacología (y dentro de esta última, la botánica y la fitoterapia). Estos recursos pueden ser de gran utilidad no solo para la práctica traductora, sino también para la interpretación o, cómo no, como base para le creación de un corpus especializado en el ámbito médico y sus especialidades.

2. La medicina y la traducción médica

El interés por la salud ha supuesto siempre una de las primeras preocupaciones del ser humano de todos los tiempos. A lo largo de la historia de la humanidad, ha sido una actividad fundamental orientada a la mejora de las condiciones de vida de nuestros semejantes. Incluso, el desarrollo de las ciencias médicas ha tenido en ocasiones mucho que ver con lo mágico o lo maravilloso, y en otras ocasiones se han desarrollado a través de la ciencia empírica. Así, estamos antes una actividad humana que no ha dejado de evolucionar, que tiene una proyección de futuro sin límites y que, a su vez, arranca de lo más remoto de la propia reflexión humana sobre el ser y la finitud del hombre.

Esta importancia, *per se*, de la medicina como preocupación fundamental del ser humano, como saber universal, ha propiciado que los textos científicos pertenecientes a esta disciplina hayan sido traducidos de una lengua a otras en esa necesidad de extender el saber. La traducción de textos de medicina ha generado una gran producción de documentos y, por lo tanto, se hace necesario disponer de unos instrumentos de traducción que permitan a los profesionales de la medicina el acercamiento a un conocimiento exacto, objetivo, universal de la materia, a través de textos que de manera efectiva y real participen de esa exactitud y objetividad que son la base fundamental de todas las ciencias.

Esta necesidad de traducción fiable ha sido también una actividad tradicional dentro de los estudios de medicina; de hecho, la traducción es una de las actividades más antiguas (Hurtado Albir 2001: 99).

A través de traducciones fiables es como se desarrolla la propia ciencia: los progresos obtenidos por unos científicos les permiten a los demás avanzar en la investigación. Aquí el papel de la correcta traducción de los textos científicos se revela como fundamental. De esta misma opinión es Muñoz-Miquel (2016: 236) cuando sostiene que:

> [a]lgunos de los ejemplos que nos brinda la historia los encontramos en la actividad traductora desarrollada en la Casa de la Sabiduría de Bagdad, gracias a la cual los árabes asimilaron el legado médico griego y produjeron obras originales, o en el papel que tuvieron los traductores clásicos y del Renacimiento para transmitir y conservar las grandes obras de la medicina.

Enmarcada dentro de la traducción científico-técnica, la denominación más extendida para referirnos a este tipo de traducción es *traducción médica* si bien, debido a la amplitud de campos que se tratan, existen otras denominaciones para referirnos a esta misma labor, como pueden ser: *traducción biomédica* (Vázquez y del Árbol 2006), *traducción biosanitaria* (García Esteban y Varela Salinas 2009) o *traducción médico-sanitaria* (Montalt Resurrecció y Shuttleworth 2012), por citar solo algunas de las más frecuentes.

En la actualidad, la traducción médica es uno de los campos de traducción más productivos. Además, de acuerdo con Muñoz-Miquel, existe una amplia oferta de trabajo para traductores médicos, lo que demuestra «cuán necesarios son estos profesionales» (Muñoz-Miquel 2016: 242).

De hecho, y de acuerdo con el estudio realizado por Navas Romero y Palomares Perraut (2002), la traducción médica se encuentra dentro de los sectores económicos en los que más se demanda el servicio del traductor. En estudios más recientes, Rico Pérez y García Aragón (2016) exponen que el 70 % de las empresas de traducción encuestadas sobre su actividad en los años 2014 y 2015 dice recibir demanda de traducción médica. En 2017, la Comisión Europea publicó el informe *2017 Language Industry Survey – Expectations and Concerns of the European Language Industry,* en el que se muestra cómo la traducción de textos médicos está entre los sectores de mayor demanda (European Commission 2017: 6).

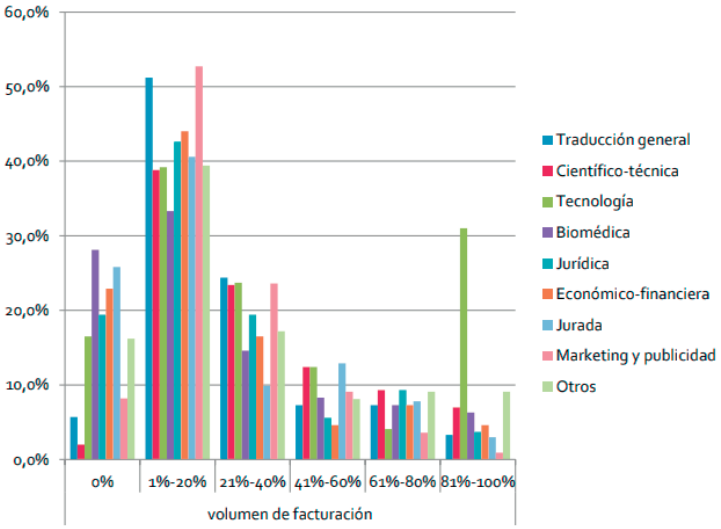


Figura 1: Especialidades de traducción más demandadas. Porcentaje con respecto a la facturación anual (Rico Pérez y García Aragón 2016: 37).

El campo de la traducción médica es tan vasto como el de la propia medicina. Siguiendo la clasificación que realiza la UNESCO (1988) dentro del campo de las ciencias médicas existen 14 disciplinas (ciencias clínicas, epidemiología, medicina forense, medicina del trabajo, medicina interna, ciencias de la nutrición, patología, farmacodinámica, farmacología, psiquiatría, salud pública, cirugía y toxicología). A su vez, estas disciplinas están divididas en subdisciplinas, las cuales alcanzan un número de 95. Además, la UNESCO propone como disciplinas afines las de bioquímica, biología humana, fisiología humana y medicina tradicional. Como es de esperar, en este trabajo no podemos tratar de manera pormenorizada todas y cada una de las subdisciplinas, es por ello por lo que debemos centrar nuestros esfuerzos en un grupo reducido de ellas.

Para seleccionar dicho grupo, por un lado, nos fijaremos en la elección de especialidades que han hecho los estudiantes del examen MIR

tras el examen de este año 2020, lo que arrojará un poco de luz sobre cuál es el futuro de la medicina en España, qué especialidades son las más demandadas. Esto significa que serán las especialidades más trabajadas y estudiadas y, por ende, las que probablemente más volumen de traducciones producirán en un futuro. En lo que respecta a la oferta, las especialidades con un mayor número de plazas disponibles eran, según Cortés (2020), a saber:

> medicina familiar y comunitaria, con 2.224 plazas, seguida por pediatría con 481, anestesiología y reanimación con 387, medicina interna con 360, psiquiatría con 272, obstetricia y ginecología con 271, cirugía ortopédica y traumatología con 267, radiodiagnóstico con 257, cirugía general y del aparato digestivo con 212 y cardiología con 180.

De entre toda la oferta de plazas, las primeras especialidades que se agotaron y, por tanto, las que más interesan a los futuros médicos fueron cirugía plástica, dermatología, cirugía oral y maxilofacial y cardiología (cfr. Negrete 2020). Además, el informe *Estimación de la oferta y demanda de médicos especialistas. España 2018-2030* realizado en 2018 y revisado en 2019 muestra que las diez especialidades con una mayor oferta de empleo público por cada 100.000 habitantes son obstetricia y ginecología, geriatría, anestesiología y reanimación, medicina interna, cirugía ortopédica y traumatología, psiquiatría, radiodiagnóstico, cirugía oral y del aparato digestivo, oftalmología y cardiología (Barber Pérez y González López-Valcárcel 2019: 60).

Por otro lado, nos interesa saber cuál ha sido la tendencia hasta ahora en cuanto a las especialidades médicas más traducidas. Así, García-Izquierdo y Conde, en su trabajo «Necesidades documentales del traductor médico en España», ponen de manifiesto que «[d]os de cada tres sujetos dicen haber traducido textos del ámbito de la farmacología. Después de esta, las categorías más mencionadas han sido medicina general, traumatología, cirugía, instrumentación y oncología» (García-Izquierdo y Conde 2014: 6), seguidas por cardiología y obstetricia y ginecología. Por su parte, Corpas Pastor y Roldán Juárez (2014) recogen las especialidades más frecuentes en la traducción mé-

dica, donde destacan la oncología, la cardiología, la medicina familiar y comunitaria, la medicina preventiva, la neurología y la pediatría (Corpas Pastor y Roldán Juárez 2014: 182).

Como hemos comentado anteriormente, un campo de las ciencias que está relacionado con las ciencias de la salud es el de la medicina tradicional, también llamado *medicinas complementarias* (Cochrane 2020) o *terapias alternativas* (NCCIH 2018), entre otras denominaciones. En cualquier caso, la medicina tradicional puede definirse, de acuerdo con la OMS, como:

> [...] todo el conjunto de conocimientos, aptitudes y prácticas basados en teorías, creencias y experiencias indígenas de las diferentes culturas, sean o no explicables, usados para el mantenimiento de la salud, así como para la prevención, el diagnóstico, la mejora o el tratamiento de enfermedades físicas o mentales. (Zhang 2010).

Si bien la situación actual de la medicina tradicional en España es algo convulsa, en muchos países europeos está en auge. De hecho, la Organización Mundial de la Salud en su informe *Estrategia de la OMS sobre medicina tradicional 2002-2005* (2002: 2) expone que «[e]n muchos países desarrollados el popular uso de la MCA [medicina complementaria y alternativa] está propulsado por la preocupación sobre los efectos adversos de los fármacos químicos hechos por el hombre». Es por esto por lo que la propia OMS lleva tiempo trabajando en unas estrategias implementadas en las que se promueve la reglamentación e investigación de la medicina tradicional.

En el *Informe sobre Terapias Naturales* (MSPSI 2011) llevado a cabo por Ministerio de Sanidad, Política Social e Igualdad español en 2011, se presenta una clasificación basada en la propuesta del *National Center for Complementary and Alternative Medicine* (NCCAM) donde se destacan cinco grandes áreas: sistemas integrales o completos, prácticas biológicas, prácticas de manipulación y basadas en el cuerpo, técnicas de la mente y el cuerpo y técnicas sobre la base de la energía. De entre todas estas, nos fijaremos en las prácticas biológicas que son

aquellas que «emplean sustancias que se encuentran en la naturaleza, usadas por su sabor, aroma o posibles propiedades terapéuticas, como hierbas, flores, hojas, cortezas de árboles, frutas, semillas, tallos y raíces, alimentos y vitaminas». Dentro de las prácticas biológicas encontramos: aromaterapia, apiterapia, homeopatía, naturopatía, dietética, terapia floral y fitoterapia. Las prácticas biológicas son las que están más cercanas a la medicina moderna puesto que forman parte de la base de especialidades como la farmacología (relacionada, en concreto, con la fitoterapia). Es por eso por lo que en nuestro estudio incluiremos, como parte de la documentación del campo de la farmacología, recursos sobre fitoterapia y botánica, pues son tres campos íntimamente relacionados.

Finalmente, nos gustaría mencionar la razón por la que hemos escogido realizar este estudio para el par de lenguas inglés-español, y es debido a que son las dos lenguas que más trabajo aportan a los traductores médicos y sanitarios españoles «siendo el inglés la lengua de partida más común (81,25 %) y el español, la lengua meta (62,5 %)» (Corpas Pastor y Roldán Juárez 2014: 181).

En cuanto al país de edición de los recursos aquí recogidos cabe comentar que se han tenido en cuenta tanto los recursos editados en España como en diferentes países de Latinoamérica, así como recursos editados en Europa, Reino Unido y Estados Unidos, si bien es cierto que no todos se comentan en este trabajo dado el amplísimo número de recursos existentes en el escopo de búsqueda en el que nos movemos (desde el siglo XIX hasta la fecha).

De modo que, una vez expuesto todo lo anterior, encontramos algunos puntos comunes entre las especialidades que más se trabajan en el ámbito sanitario en nuestro país en la actualidad, las preferencias de los futuros médicos y la tendencia de traducción en los últimos años. Dada la extensión de la que disponemos y los motivos supraexpuestos, en este artículo nos centraremos en las siguientes especialidades: obstetricia y ginecología, cirugía ortopédica y traumatología, cardiología, oncología, cirugía, farmacología, botánica y fitoterapia.

3. Evaluación de los recursos en línea y en papel

Para poder analizar y evaluar los recursos existentes para estas especialidades, se ha creado una plantilla de análisis y evaluación puntuable electrónica.[1] Se trata de una plantilla diseñada para que el usuario pueda completar los campos y realizar la evaluación del recurso según sus necesidades específicas. La plantilla incluye hasta 15 indicadores para la evaluación de los recursos, algunos de estos indicadores no se aplicarían dependiendo de la especialidad que se esté evaluando. El usuario evaluador asigna a cada uno de los parámetros una puntuación entre 0 y 3 (siendo 0 la menor puntuación y 3 la más alta) dependiendo de si el recurso dispone o no de dicha característica o de la importancia o interés que tiene para el usuario que el recurso disponga de la misma. Así, se establecen unos intervalos según los cuales aquel recurso que obtenga entre el 0 % y el 60 % de puntuación se considerará de interés o fiabilidad bajos, el recurso que esté entre el 61 % y el 80 % se considerará de interés o fiabilidad medios y todo aquel recurso que tenga una puntuación igual o superior al 81 %, será considerado de interés o fiabilidad altos. Los recursos recogidos más abajo han sido evaluados siguiendo dicha plantilla y todos han obtenido una puntuación superior al 60 % de la puntuación total, es decir, se trata de recursos de calidad media o alta.

Una vez explicado de manera breve cómo se utiliza la plantilla de análisis y evaluación se procede a establecer un supuesto de traducción que permita comprobar su validez, en este caso se evaluarán los recursos desde la óptica de un traductor profesional que se tiene que enfrentar por primera vez a la traducción de un texto actual que contiene un gran número de términos especializados. Al ser un traductor profesional ya tiene experiencia en el campo de la traducción, pero no en el campo de la especialidad en concreto, por lo que necesita que las entradas del diccionario sean lo más completas posibles.

1. Para más información acerca de los campos incluidos en la plantilla y cómo utilizarla consultar Escribano Meseguer (2021 en prensa).

NOMBRE DEL RECURSO			
ASPECTOS GENERALES		Observaciones	PUNTUACIÓN
N.º de lenguas			
Lenguas			
N.º de páginas			
N.º de entradas			
Año de publicación			
Última actualización			
Nivel de léxico			
Soporte			
MACROESTRUCTURA		Observaciones	PUNTUACIÓN
Indicaciones			
Mapas			
Hipertexto			
Orden de las entradas			
MICROESTRUCTURA		Observaciones	PUNTUACIÓN
Definiciones			
Información gramatical			
Denominación			
Sinónimos y antónimos			
Marcas			
Ejemplos			
Ilustraciones			
		PUNTUACIÓN TOTAL	0
MÁS INFORMACIÓN			
Nombre del autor			
Profesión/organismo/institución			
Ciudad			
Editorial			
Org. lógica de la información			
Presentación o navegación			

Figura 2. Plantilla de análisis y evaluación electrónica puntuable
(Escribano Meseguer 2021 en prensa).

4. Catálogo de fuentes terminológicas médicas por especialidad

En primer lugar, cabe mencionar que los traductores especializados en textos médicos hacen gran uso de recursos lexicográficos de la medicina general tan importantes como pueden ser el *Diccionario de términos médicos* (RANM 2011) de la Real Academia Nacional

de Medicina de España, disponible tanto en papel como en línea, el *Diccionario de dudas y dificultades de traducción del inglés médico* de Fernando A. Navarro (Navarro 2021) también disponible en papel y en línea en el portal *Cosnautas*, el recurso electrónico y multilingüe *Multilingual Glossary of technical and popular medical terms in nine European Languages* llevado a cabo por el *Heymans Institute of Pharmacology* en la Universidad de Gante (Van der Stichele 2020) o el *Diccionario de Ciencias Médicas Stedman bilingüe (Español-Inglés)* (Stedman y Beatty 1999) en papel. Sin embargo, en este estudio nos centraremos en aquellos recursos que profundizan en las diversas especialidades propuestas y que complementan o amplían a los recursos generales aquí mencionados.

Organizaremos este estudio de los recursos para la traducción relativos a la ciencia médica en sus distintas especialidades de la siguiente manera: en primer lugar, estudiaremos los recursos monolingües en español y, a continuación, los recursos monolingües en inglés. Después seguirán los recursos bilingües y, si los hubiera, aquellos multilingües. Dentro de cada categoría de las antedichas, los ordenaremos atendiendo a la importancia de los autores y de las autoridades que los avalan, de manera que los primeros serán los publicados por autoridades médicas de reconocido prestigio. Y finalmente, también dentro de esas categorías, primero los recursos en línea, por ser, por lo general, los más actuales, más accesibles y los más consultados por los traductores y, después, los publicados exclusivamente en papel ya que suelen ser anteriores a los recursos en línea, menos utilizados hoy en día y menos accesibles.

4.1. Obstetricia y ginecología

Dentro de los recursos avalados por instituciones de reconocido prestigio encontramos, como monolingües en español (cfr. Tabla 1), el *Diccionario Ginecológico* (AA. VV. 2016a) en línea que nos proporciona el Hospital Gregorio Marañón. Son más abundantes los recursos, también en línea, que nos ofrecen entidades de carácter más restringi-

do, como pueden ser asociaciones de profesionales o incluso personas de carácter individual. De ámbito más restringido, estos glosarios se ofrecen a través de blogs de carácter médico y tienen un carácter de *apunte* personal; el responsable es la persona que gestiona el blog o página web. Así, no se pone en duda la veracidad y la exactitud de esos glosarios, pero no aparecen avalados por un grupo colegiado de profesionales, ni por una entidad de reconocido prestigio médico. Un ejemplo es el *Diccionario de Ginecología* del doctor Francisco Carmona (Carmona 2019). Como perteneciente a asociaciones de profesionales de carácter no internacional tenemos el *Glosario Ginecológico* del Centro Clínico Betanzos 60 (AA. VV. 2020), en el que es evidente el ámbito más restringido. En cuanto a los recursos en papel, también en español, encontramos que dos de los recursos principales los firman Cruz y Hermida y Becerro de Bengoa, *Diccionario de Términos Obstétrico-Ginecológico* (Becerro de Bengoa Callau y Cruz y Hermida 2017b) y *Diccionario de obstetricia y ginecología ilustrado* (Becerro de Bengoa Callau y Cruz y Hermida 2017a) uno de ellos ilustrado y el otro no.

Figura 3. *Diccionario Ginecológico* del Hospital Gregorio Marañón.

En lo referente a los recursos monolingües en inglés encontramos que todos aquellos que están disponibles en línea tienen una calidad o fiabilidad baja (cfr. Tabla 1). En lo relativo a los textos en papel, la editorial De Gruyter publica *Dictionary of Obstetrics and Gynecology* (Zink 1988). Y hay que añadir también otro diccionario de varios autores, *Melloni's Illustrated Dictionary of Obstetrics and Gynecology* (Dox, Melloni y Sheld 2000) en el que destacan los dibujos y esquemas explicativos, así como la traducción fonética de las entradas.

En cuanto a los recursos bilingües (inglés y español) encontramos solamente en línea, y todos ellos, de nuevo, con una calidad o fiabilidad baja (cfr. Tabla 1).

Obstetricia y Ginecología	
Monolingües español	
Recurso	**Índice de calidad**
Martín, Concha: *Diccionario de términos ginecológicos*, en línea, <http://www.conchamartin.es/web/diccionario-de-terminos-ginecologicos/> [Consulta: 22/10/2020].	Bajo
AA. VV. (2020): *Glosario Ginecológico del Centro Clínico Betanzos 60*, en línea, <https://centroclinicobetanzos60.es/medicos-especialistas/glosario-2/> [Consulta: 22/10/2020].	Medio
Todopapás. Diccionario médico de Obstetricia, en línea, <https://www.todopapas.com/diccionario/obstetricia/1> [Consulta: 22/10/2020].	Bajo
Carmona, Francisco: *Diccionario de Ginecología*, en línea, <https://www.drfcarmona.com/glosario-de-ginecologia/> [Consulta: 22/10/2020].	Medio
Terminología de gineco y obstetricia, Apuntes de Ginecología, en línea, <https://www.docsity.com/es/terminologia-de-gineco-y-obstetricia/642850/> [Consulta: 22/10/2020].	Bajo
AA. VV. (2016a): *Diccionario Ginecológico del Hospital Gregorio Marañón*, en línea, <https://booksmedicos.org/diccionario-ginecologico/> [Consulta: 22/10/2020].	Medio
Instituto Bernabeu. Diccionario Ginecológico, en línea, <https://www.institutobernabeu.com/es/diccionario-ginecologico/> [Consulta: 22/10/2020].	Bajo
Instituto Nacional de Cancerología-ESE Colombia. Glosario Clínica de Ginecología, en línea, <https://cancer.gov.co/content/glosario-cl%C3%ADnica-de-cl%C3%ADnica-de-ginecolog%C3%ADa> [Consulta: 22/10/2020].	Bajo
Becerro de Bengoa Callau, Claudio; Cruz y Hermida, Julio (2017a): *Diccionario de Obstetricia y Ginecología ilustrado*, España, Marbán.	Alto
Becerro de Bengoa Callau, Claudio; Cruz y Hermida, Julio (2017b): *Diccionario de Términos Obstétrico-Ginecológico*, España, Marbán.	Medio
Zink, Christoph (1988a): *Diccionario Pschyrembel de Ginecología y Obstetricia*, Berlín, Walter de Gruyter.	Bajo

MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
NHS. Saint Mary's Hospital. Glossary of Obstetric Terminology, en línea, <https://mft.nhs.uk/saint-marys/services/maternity-services-obstetrics/glossary-of-obstetric-terminology/> [Consulta: 22/10/2020].	Bajo
RCOG. Royal College of Obstetricians and Gynaecologists. Medical terms explained, en línea, <https://www.rcog.org.uk/en/patients/medical-terms/> [Consulta: 22/10/2020].	Bajo
RCOG. Royal College of Obstetricians and Gynaecologists. Glossary of terms, en línea, <https://www.rcog.org.uk/en/careers-training/resources--support-for-trainees/useful-resources-for-trainees/glossary-of-terms/> [Consulta: 22/10/2020].	Bajo
Women&Babies. Medical terms, en línea, <http://www.womenandbabies.co.uk/medical-terms.html> [Consulta: 22/10/2020].	Bajo
Instituto Bernabeu. Dictionary of Gynaecology, en línea, <https://www.institutobernabeu.com/en/dictionary-gynecological/> [Consulta: 22/10/2020].	Bajo
Instituto Nacional de Cancerología-ESE Colombia. Glossary of Gynecology Clinic, en línea, < https://cancer.gov.co/content/glosario-cl%C3%ADnica-de-cl%C3%ADnica-de-ginecolog%C3%ADa> [Consulta: 22/10/2020]. *Johns Hopkins Medicine. Abbreviations Commonly Used in Obstetrics and Gynecology*, en línea, <https://www.hopkinsmedicine.org/gynecology_obstetrics/education/medstudents/_docs/learning_mtls/abbreviations.pdf> [Consulta: 22/10/2020].	Bajo Bajo
ZINK, Christoph (1988b): *Dictionary of Obstetrics and Gynecology*, Berlín, Walter de Gruyter.	Medio
DOX, Ida; MELLONI, June; SHELD, Harrison (2000): *Melloni's Illustrated Dictionary of Obstetrics and Gynecology*, CRC Press.	Medio
BILINGÜES	
Recurso	**Índice de calidad**
CHACÓN ESTRADA, Númida María; LEYVA PÉREZ, María Margarita (2003): Miniglosario en inglés y español de la terminología más útil en Ginecología y Obstetricia, en línea, <http://eprints.rclis.org/5110/1/aci05403.pdf> [Consulta: 22/10/2020].	Bajo
Transpanish. Gynecology – Obstetrics Glossary, en línea, <https://transpanish.biz/gynecology-glossary/> [Consulta: 22/10/2020].	Bajo
COPELAND, Robert: *ProZ. Translation glossary: Robert Copeland-Gynecology and Obstetric Terms*, en línea, <https://www.proz.com/personal-glossaries/28798-robert-copeland-gynecology-and-obstetric-terms> [Consulta: 22/10/2020].	Bajo

Tabla 1. Recursos de la especialidad de obstetricia y ginecología.

4.2. Cirugía ortopédica y traumatología

Traumatología es la especialidad de la que menos recursos encontramos a todos los niveles y en ambos idiomas. Sin embargo, esto no quiere decir que los recursos encontrados, si bien escasos, no sean de una calidad aceptable para el usuario.

En lo que respecta a recursos monolingües en español (cfr. Tabla 2) encontramos uno en línea y uno en papel. Un trabajo particular es el del español Miguel Vázquez, MIR en cirugía ortopédica y traumatología quien, en su web sanitaria acreditada Traumatopedia publica un *Diccionario de traumatología* (Vázquez 2021) el cual incluye amplias explicaciones, así como imágenes ilustrativas. Por último, faltaría mencionar que en formato papel disponemos del *Diccionario Enciclopédico ilustrado de traumatología* (Bastos Mora 1999) publicado en la editorial médica JIMS, SL.

Figura 4. *Diccionario de traumatología* (Vázquez 2021) de la web Traumatopedia.

Hemos localizado tres recursos monolingües en inglés, disponibles en línea y de Estados Unidos, avalados por universidades e instituciones de prestigio, aunque todos ellos han obtenido una calificación de calidad o fiabilidad baja (cfr. Tabla 2).

Para cerrar este capítulo, entre las obras que publica la editorial Díaz de Santos encontramos *Terminología inglés-español para la práctica médica* (Puente Fonseca 1997), en cuyo capítulo 12.5 aparece un glosario de vocabulario frecuente en ortopedia y traumatología.

CIRUGÍA ORTOPÉDICA Y TRAUMATOLOGÍA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
VÁZQUEZ, Miguel: *Traumatopedia. Diccionario traumatología*, en línea, <https://traumatopedia.com/diccionario-traumatologia/> [Consulta: 22/10/2020].	Medio
Traumatología -glosario, Apuntes de Traumatología, en línea, <https://www.docsity.com/es/traumatologia-glosario/5074841/> [Consulta: 22/10/2020].	Bajo
BASTOS MORA, Felipe (1999): *Diccionario Enciclopédico ilustrado de traumatología*, Barcelona, Editorial Médica JIMS.	Alto
MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
University of Rochester Medical Center. Kessler Burn & Trauma Center. Glossary of Terms Related to Burn and Trauma Treatment, en línea, <https://www.urmc.rochester.edu/burn-trauma/terms.aspx> [Consulta: 22/10/2020].	Bajo
National Center for Biotechnology Information. Major Trauma: Assessment and Initial Management, en línea, <https://www.ncbi.nlm.nih.gov/books/NBK368126/> [Consulta: 22/10/2020].	Bajo
Institute for Digital Health & Innovation. Trauma Rehabilitation Resources Program. Glossary of Medical Terms, en línea, <https://atrp.ar.gov/resources/glossary/> [Consulta: 22/10/2020].	Bajo
BILINGÜES	
Recurso	**Índice de calidad**
PUENTE FONSECA, Claudio Julio (1997): *Terminología inglés-español para la práctica médica*, Madrid, Ediciones Díaz de Santos.	Medio

Tabla 2. Recursos de la especialidad de cirugía ortopédica y traumatología.

4.3. Cardiología

En primer lugar, llama la atención que los recursos en inglés (18) son más numerosos que los recursos en español (10) y que los recursos bilingües (7); sin embargo, no todos tienen una calidad o fiabilidad de nivel medio o alto.

Entre los creados en español (cfr. Tabla 3) encontramos que son escasos los publicados por instituciones de prestigio: *Gestión de Pacientes Ambulatorios en Cardiología*, que incluye glosario de abreviaturas (AA. VV. 2001), que lo encontramos en la página web de la Sociedad Española de Cardiología (en donde el glosario no es el contenido principal). Perteneciente a particulares, encontramos el *Diccionario Temático de Medicina. Especialidad: Cardiología* (AA. VV. 2021d),

publicado en Portales Médicos, con la colaboración de diferentes profesionales de la salud, de naturaleza más heterogénea que las obras concebidas con un único plan. Y como recurso en papel destacamos el *Diccionario LID Cardiología* (Anguita Sánchez 2009), de Manuel P. Anguita.

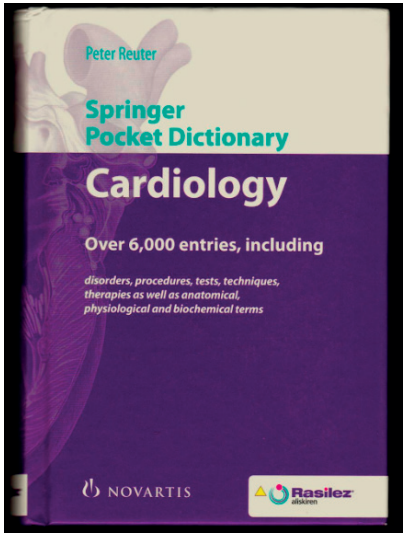


Figura 5. *Springer Pocket Dictionary. Cardiology* (Reuter 2010) de Peter Reuter.

En lo referente a los recursos en inglés en línea, todos los encontrados tienen una fiabilidad o calidad baja (cfr. Tabla 3). Los recursos en papel en inglés también son más abundantes que los publicados en español. La mayoría de estos recursos en papel son pertenecientes a autores particulares: Sleight (1960), Anderson *et. al* (1992), Rhodes, Dorland y David (2001a), Hayes y Asirvatham (2007), Reuter (2010).

En cuanto a recursos bilingües de calidad media o alta solo podemos mencionar el *Glosario de términos y abreviaturas en cardiología inglés-español* (Alegría Ezquerra y Teresa Galván 2005), si bien existen muchos más recursos de una calidad algo inferior (cfr. Tabla 3).

CARDIOLOGÍA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
ARAÚJO, Olga: *Glosario de términos cardiovasculares*, en línea, <https://medicointernista.es/glosario-de-terminos-cardiovasculares/> [Consulta: 22/10/2020].	Bajo
AA. VV. (2001): *Gestión de Pacientes Ambulatorios en Cardiología*, en línea, <http://www.comsor.es/pdf/formacion/Seccion%2001/gestion_pacientes_ambulatorios_en_cardiologia.pdf> [Consulta: 22/10/2020].	Medio
Cardio Quirón. Glosario, en línea, <https://cardioquiron.com/paciente/preguntas-frecuentes/glosario/> [Consulta: 22/10/2020].	Bajo
AA. VV. (2021d): *Diccionario Temático de Medicina. Especialidad: Cardiología*, en línea, <https://www.portalesmedicos.com/diccionario_medico/index.php?title=Cardiologia&no_redirect=true> [Consulta: 22/10/2020].	Medio
Fundación Española del Corazón. Glosario de Salud, en línea, <https://fundaciondelcorazon.com/informacion-para-pacientes/mas-salud/glosario-salud.html> [Consulta: 22/10/2020].	Bajo
Texas Heart Institute. Glosario de terminología cardiovascular, en línea, <https://www.texasheart.org/heart-health/heart-information-center/topics/glosario-de-terminologia-cardiovascular/> [Consulta: 22/10/2020].	Bajo
ANGUITA SÁNCHEZ, Manuel (dir.) (2009): *Diccionario LID Cardiología*, Madrid, LID Editorial Empresarial S.L.	Medio
ANDERSON, Robert; DAWSON, Rex; OLDERSHAW, Paul (1992): *Diccionario de cardiología ilustrado para el médico clínico*, Barcelona, Science Press.	Bajo
SMITH, Anthony (1992*): Diccionario de Cardiología*, s. l., Harofarma.	Bajo
MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
BRITISH HEART FOUNDATION: *The heart - technical terms explained*, en línea, <https://www.lhch.nhs.uk/media/1150/his18_heart_technicalterms_booklet.pdf> [Consulta: 22/10/2020].	Bajo
SHARP. Scottish Heart and Arterial Risk disease Prevention. What is a cardiovascular disease? Commonly used terms, en línea, <https://www.sharpscotland.org.uk/information/what-is-cardiovascular-disease/> [Consulta: 22/10/2020].	Bajo
NORELL, Michael: *A glossary of terms used in interventional cardiology: part 2*, en línea, <https://www.chss.org.uk/heart-information-and-support/resources-to-help/heart-medical-terms-glossary/> [Consulta: 22/10/2020].	Bajo
Chest Heart & Stroke Scotland. Heart conditions glossary, en línea, <https://bjcardio.co.uk/2010/07/a-glossary-of-terms-used-in-interventional-cardiology-part-2/> [Consulta: 22/10/2020].	Bajo
Cambridge Heart Clinic. Glossary of terms associated with cardiology conditions, en línea, <https://cambridgeheartclinic.co.uk/patient-information/glossary/> [Consulta: 22/10/2020].	Bajo
The British Cardiac Patients Association. Glossary Of Terms, en línea, <http://bcpa.uk/> [Consulta: 22/10/2020].	Bajo
RAKHIT, Roby: *Glossary of Cardiology Terms and Tests*, en línea, <https://www.robyrakhit.co.uk/glossary> [Consulta: 22/10/2020].	Bajo
NICOR. National Institute for Cardiovascular Outcomes Research. Glossary, en línea, <https://www.nicor.org.uk/glossary/> [Consulta: 22/10/2020].	Bajo
ESC. European Society of Cardiology. Glossary, en línea, <https://www.escardio.org/Education/Glossary> [Consulta: 22/10/2020].	Bajo

NHS. Greater Glasgow and Clyde. Heart, Stroke and Diabetes. Glossary of Terms, en línea, <http://live.nhsggc.org.uk/media/223784/Glossary%20of%20Terms.pdf> [Consulta: 22/10/2020].	Bajo
Texas Heart Institute. Cardiovascular Glossary, en línea, <https://www.texasheart.org/heart-health/heart-information-center/topics/a-z/> [Consulta: 22/10/2020].	Bajo
ANDERSON, Robert *et al.* (1991): *Clinician's Illustrated Dictionary of Cardiology*, 2ª ed., Londres, Science Press.	Alto
SLEIGHT, Peter (1960): *International Dictionary of Cardiology*, Londres, Radcliffe Publishing Ltd.	Medio
LENEHAM, Anne (1995): *Stedman's Dictionary of Cardiology from Bayer*, Lippincott Williams and Wilkins.	Medio
REUTER, Peter (2010): *Springer Pocket Dictionary of Cardiology*, s. l., JP Medical Publishers.	Bajo
HAYES, David; ASIRVATHAM, Samuel (2007): *Dictionary of Cardiac Pacing, Defibrillation, Resynchronization, and Arrhythmias*, 2ª ed., Mineápolis, Cardiotext Publishing.	Medio
RHODES, Sharon (edit.); DORLAND STAFF; DAVID, Mary (2001a): *Dorland's Cardiology Word Book for Medical Transcriptionist*, s. l., Saunders.	Medio
BILINGÜES	
Recurso	**Índice de calidad**
LATORRE VILALLONGA, Jordi; COSTA MER, Ana María (2005): *Diccionario de cardiología y abreviaturas de la especialidad*, Barcelona, Nexus Médica.	Bajo
Revista Española de Cardiología. Abbreviations, en línea, <https://www.revespcardiol.org/en-abreviaturas-articulo-13116902?redirect=true> [Consulta: 22/10/2020].	Bajo
MedicalSpanish.com. The Heart, en línea, <https://www.medicalspanish.com/medical-vocabulary/heart.html> [Consulta: 22/10/2020].	Bajo
Intérpretes de Conferencias. Glosario de Cardiología General, en línea, <http://www.interpretesdeconferencias.mx/edocman/28%20Glosario%20de%20Cardiologa%20%20%20Gabriela%20Durazo.pdf> [Consulta: 22/10/2020].	Bajo
ALEGRÍA EZQUERRA, Eduardo; DE TERESA GALVÁN, Eduardo (2005): *Glosario de términos y abreviaturas en cardiología inglés-español*, Madrid, Acción Médica.	Medio

Tabla 3. Recursos de la especialidad de cardiología.

4.4. Oncología

Los recursos lexicográficos que encontramos para la especialidad médica de oncología son muy numerosos (aunque esto no quiere decir que todos sean de buena calidad), y los hay disponibles tanto en línea como en papel (cfr. Tabla 4).

Analizaremos, para empezar, los recursos monolingües en español disponibles en línea. En primer lugar, recogeremos aquellos recursos que son respaldados por hospitales e institutos de investigación como puede ser el *Diccionario del cáncer del NCI*, del Instituto Nacional del Cáncer, Estados Unidos (AA. VV. 2021b) y algunos otros recursos, aunque de menor calidad (cfr. Tabla 4). También cabe destacar la Asociación Española Contra el Cáncer, la cual goza de reconoci-

miento a nivel nacional y que pone a nuestra disposición un glosario redactado por varios profesionales relacionados con la oncología en formato PDF, *Glosario del Cáncer* (AECC 2007a), y otro en formato html disponible en su página web que contiene un número menor de entradas, *Glosario* (AECC 2007b). Y, finalmente, los recursos que encontramos a título más individual o personal de profesionales de la salud o investigadores médicos resultan tener una calidad o fiabilidad baja (cfr. Tabla 4).

En cuanto a los recursos monolingües publicados en papel, cabe desatacar que ninguno de ellos está recogido en una editorial especializada en el ámbito médico o científico; sin embargo, encontramos algunos títulos interesantes: *Diccionario LID Oncología* (González de Francisco y Lacal 2009) donde se muestran las definiciones de las entradas en español así como la traducción al inglés de los términos y también incluye un diccionario inverso del inglés al español; y *Diccionario de términos oncológicos para los no oncólogos* (Benlloch Rodríguez *et al.* 2018), cuya coordinadora es Raquel Benlloch, en Didot Edición.

Figura 6. *Diccionario del cáncer del NCI* (AA. VV. 2021b) del Instituto Nacional del Cáncer de Estados Unidos.

Si buscamos recursos lexicográficos monolingües en inglés en línea (cfr. Tabla 4), encontramos a título personal, aunque enmarcado dentro de la Universidad de Newcastle upon Tyne, los dos glosarios que firma Sim Cotterill, uno de terminología oncológica y otro de abreviaturas y acrónimos: *Cancer Glossary* (Cotterill 2002), albergado en CancerIndex, y *Medical Terminology and Cancer. Abbreviations and Acronyms for Oncology* (Cotterill 2000).

En lo que a recursos monolingües en inglés se refiere hay que señalar que, a diferencia de las publicaciones en español, las editoriales que albergan estos recursos en sus colecciones sí son médicas o científicas en su mayoría. Entre estas publicaciones destacamos las dos siguientes: *Cancer – A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References* (Parker y Parker 2004) de la editorial ICON Health Publications, una valiosísima publicación que, además de incluir una amplia lista de la literatura existente en Internet sobre la especialidad, también contiene un diccionario de términos; y *Dorland's Radiology/Oncology Word Book for Medical Transcriptionists* (Rhodes, Dorland y Pugh 2001b) publicado en Saunders, una amplia fuente de términos relacionados con la oncología y la radiología. Además, gracias al reconocido servicio de publicaciones de la Universidad de Cambridge, disponemos de la guía M*odern Oncology: An A-Z of Key Topics* (Munro 2001) que proporciona una amplia discusión en la gran mayoría de los términos para una mejor comprensión de los mismos, así como numerosas referencias cruzadas. Una publicación similar sería *Cancer from A to Z: A Dictionary of Cancer-Related Terms* (Laughlin 2008), pero esta vez autopublicada a título personal por Edward Laughlin, quien además del glosario incluye una sección detallada sobre los diferentes tipos de cáncer.

Para finalizar la búsqueda de recursos sobre oncología hemos de mencionar un muy buen recurso en línea bilingüe (cfr. Tabla 4) hallados avalado por la Universidad de Granada, que alberga en su página web la base de datos terminológicos *OncoTerm. Sistema Bilingüe de Información y Recursos Oncológicos* (AA. VV. 2002b) en el que la lista de términos parece en inglés y dentro de la ficha terminológica encontramos su traducción al español.

ONCOLOGÍA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
AECC (2007a): *Glosario del Cáncer*, en línea, <https://www.aecc.es/sites/default/files/migration/actualidad/publicaciones/documentos/glosario.pdf> [Consulta: 22/10/2020].	Medio
AECC (2007b): *Asociación Española Contra el Cáncer. Glosario*, en línea, <https://www.aecc.es/es/todo-sobre-cancer/glosario> [Consulta: 22/10/2020].	Bajo
Oncodudas. Glosario de términos, en línea, <http://oncodudas.es/diccionario/> [Consulta: 22/10/2020].	Bajo
Conviveconelcancer. Diccionario, en línea, <https://conviveconelcancer.com/diccionario/> [Consulta: 22/10/2020].	Bajo
Hospital Universitario del Vinalopó. Guía Paciente Oncológico, en línea, <https://www.vinaloposalud.com/wp-content/uploads/2018/10/Gu%C3%ADaPacienteOncol%C3%B3gico_Castellano_VS.pdf> [Consulta: 22/10/2020].	Bajo
GARCÍA GARCÍA, María del Carmen (2015): «Glosario de términos relacionados con la Enfermería oncológica», *Portales Médicos*, en línea, <https://www.revista-portalesmedicos.com/revista-medica/glosario-de-terminos-enfermeria-oncologica/> [Consulta: 22/10/2020].	Bajo
Initia. Centro Oncológico Integral. Glosario del Cáncer, en línea, <https://initiaoncologia.com/glosario-page/> [Consulta: 22/10/2020].	Bajo
MD Anderson Cancer Center Madrid. Glosario. Definición de términos médicos usados en la web, en línea, <https://mdanderson.es/elcancer/glosario> [Consulta: 22/10/2020].	Bajo
AA. VV. (2021b): *Diccionario del cáncer del NCI*, en línea, <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/?redirect=true> [Consulta: 22/10/2020].	Medio
MANDAL, Ananya: *Glosario del cáncer*, en línea, <https://www.news-medical.net/health/Cancer-Glossary-(Spanish).aspx> [Consulta: 22/10/2020].	Bajo
VILADIU QUEMADA, Pau. (1993): *Diccionario oncológico: Abreviaturas, siglas y acrónimos*, Barcelona, JIMS S. A.	Bajo
GONZALEZ DE FRANCISCO, Tomás; LACAL, Juan Carlos (2009): *Diccionario LID Oncología*, España, LID Editorial Empresarial S. L.	Medio
BENLLOCH RODRÍGUEZ, Raquel *et al.* (2018): *Diccionario de términos oncológicos para los no oncólogos*, s. l., Edición Punto Didot.	Medio
CEBALLOS ATIENZA, Rafael (2018): *Glosario de términos de oncología y cuidados paliativos (Ciencias sanitarias)*, s. l., Formación Alcalá.	Bajo
DUQUE AMUSCO, Antonio; ORDÓÑEZ GALLEGO, Amalio (1994): *DICCIONARIO ONCOLÓGICO GRAMATICAL (Con Apéndice De Términos Médicos Inusuales)*, Madrid, Libro del Año Editorial SL.	Bajo
MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
NCRAS. National Cancer Registration and Analysis Service. Glossary of Cancer Terms and Vocabulary, en línea, <http://www.ncin.org.uk/search/glossary> [Consulta: 22/10/2020].	Bajo
Irish Cancer Society. Glossary of cancer terms, en línea, <https://www.cancer.ie/cancer-information-and-support/cancer-information/about-cancer/glossary> [Consulta: 22/10/2020].	Bajo
Cambridge Oncology Partners. Glossary of terms, en línea, <http://www.cambridgeoncology.co.uk/conditions-treatments/glossary/> [Consulta: 22/10/2020].	Bajo
COTTERILL, Simon (2002): *CancerIndex. Cancer glossary*, en línea, <http://www.cancerindex.org/glossary.htm> [Consulta: 22/10/2020].	Medio

COTTERILL, Simon (2000): *Medical Terminology and Cancer. Abbreviations and Acronyms for Oncology*, en línea, <https://www.staff.ncl.ac.uk/s.j.cotterill/medtm15a.htm> [Consulta: 22/10/2020].	Bajo
NHS. Health Careers. Glossary, en línea, <https://www.healthcareers.nhs.uk/glossary#Oncology> [Consulta: 22/10/2020].	Bajo
Nottinghamshire Head & Neck Cancer Service. Glossary, en línea, <http://nottshncs.nhs.uk/glossary> [Consulta: 22/10/2020].	Bajo
NHS. Health in Wales. Glossary of medical terms, en línea, <https://cutt.ly/phlWA7S> [Consulta: 22/10/2020].	Bajo
NIH. National Cancer Institute. NCI Dictionary of Cancer Terms, en línea, <https://www.cancer.gov/publications/dictionaries/cancer-terms> [Consulta: 22/10/2020].	Bajo
MUNRO, Alastair (2001): *Modern Oncology: An A-Z of Key Topics*, Cambridge, Cambridge University Press.	Medio
PARKER, James; PARKER, Phillip (eds.) (2004): *Cancer – A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References*, s. l., ICON Health Publications.	Medio
GIBSON, John; POTPARIC, Oliverira (1994): *A Dictionary of Human Oncology: A Concise Guide to Tumors,* Londres, Taylor & Francis Ltd.	Bajo
RHODES, Sharon (edit.); DORLAND STAFF; PUGH, Janet (2001b): *Dorland's Radiology/ Oncology Word Book for Medical Transcriptionists*, s. l., Saunders.	Medio
LAUGHLIN, Edward (2008): *CANCER from A to Z: A Dictionary of Cancer-Related Terms*, Bloomington, AuthorHouse.	Medio
BILINGÜES	
Recurso	**Índice de calidad**
AA. VV. (2002b): *OncoTerm. Sistema Bilingüe de Información y Recursos Oncológicos*, en línea, <https://www.ugr.es/~oncoterm/index.html> [Consulta: 22/10/2020].	Alto
Instituto Catalán de Oncología. Diccionari del cancer, en línea, <http://ico.gencat.cat/es/el-cancer/diccionari_del_cancer/> [Consulta: 22/10/2020].	Bajo
American Cancer Society. Glosario sobre cáncer, en línea, <https://www.cancer.org/es/cancer/glossary.html> [Consulta: 22/10/2020].	Bajo

Tabla 4. Recursos de la especialidad de oncología.

4.5. Cirugía

La mayoría de los recursos lexicográficos que existen para la especialidad médica de cirugía pertenecen a algún subtipo (cirugía plástica, cirugía estética, cirugía bucal, etc.). En este estudio hemos recogido algunos recursos disponibles sobre cirugía general. Asimismo, cabe destacar que, a diferencia de los recursos de especialidades anteriores, en este caso muchos de los encontrados tienen más de cien años de antigüedad.

En cuanto a recursos monolingües se refiere destacamos cuatro recursos en línea. De una parte, el *Diccionario Quirúrgico Ilustrado Complutense* (García Barreno y Aller Reyero 2016) redactado por el De-

partamento de Cirugía de la Facultad de Medicina de la Universidad Complutense de Madrid, en el que se incluye la traducción al inglés de las entradas, así como amplias explicaciones e ilustraciones que ayudan a la comprensión de los términos; y *El cuerpo herido. Un diccionario filosófico de la cirugía* (Pera 1998), llevado a cabo por Cristóbal Pera con el respaldo de la Universidad de Barcelona, en el que se tratan los términos de la especialidad desde un prisma más histórico, incluyendo citas de autores, la etimología de la palabra y desde cuando se conoce o se utiliza. Y, por otro lado, dos recursos publicados en el siglo XIX y, por lo tanto, no demasiado actualizados dado el ritmo al que avanza la ciencia, son el *Vocabulario médico-quirúrgico o diccionario de medicina y cirugía, que comprende la etimología y definición de todos los términos usados en estas dos ciencias por los autores antiguos y modernos* (Hurtado de Mendoza 1840) y el *Diccionario de medicina y cirugía ó Biblioteca manual médico quirúrgica* (Ballano 1807), ambos disponibles para consulta en línea en la Biblioteca Virtual de la Filología Española.

Figura 7. *Diccionario Quirúrgico Ilustrado Complutense* (García Barreno y Aller Reyero 2016).

Los recursos editados en inglés que aquí recogemos han sido todos publicados en papel (cfr. Tabla 5). Al igual que, como hemos visto, sucede con algunos de los recursos en español, encontramos tres recursos que fueron publicados en el siglo XIX y que, aunque son muy concisos, han quedado desactualizados: *Dictionary of Practical Surgery* (Heath 1886), *A Dictionary of Domestic Medicine and Household Surgery* (Thomson y Steele 1883) y *A Dictionary of Practical Surgery: Containing a Complete Exhibition of the Present State of the Principles and Practice of Surgery, Collected from the Best and Most Original Sources of Information, and Illustrated by Critical Remarks* (Cooper 1810). Más actual es *Surgery – A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References* (Parker y Parker 2003) publicado en ICON Health Publications que, como ya hemos mencionado en el apartado anterior sobre la publicación de oncología de la misma editorial, supone una fuente con un gran potencial.

Finalmente, podemos destacar un recurso bilingüe (cfr. Tabla 5) recogido en el capítulo 12.2 de la publicación *Terminología inglés-español para la práctica médica* (Puente Fonseca 1997), publicado por la editorial científico-técnica Díaz de Santos.

CIRUGÍA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
HURTADO DE MENDOZA, Manuel (1840): *Vocabulario médico-quirúrgico o diccionario de medicina y cirugía, que comprende la etimología y definición de todos los términos usados en estas dos ciencias por los autores antiguos y modernos*, Madrid, Boix, en línea, <http://bdh-rd.bne.es/viewer.vm?id=0000099758&page=1> [Consulta: 22/10/2020].	Medio
GARCÍA BARRENO, Pedro (dir.); ALLER REYERO, María Angeles (coord.) (2016): *Diccionario Quirúrgico Ilustrado Complutense*, Madrid, Universidad Complutense de Madrid, en línea, <https://www.ucm.es/cirugia/diccionario-quirurgico-ilustrado> [Consulta: 22/10/2020].	Medio
PERA, Cristóbal (1998): *El cuerpo herido. Un diccionario filosófico de la cirugía*, Barcelona, Edicions Universitat de Barcelona (*Estudi General*, n. 3).	Medio
BALLANO, Antonio (1807): *Diccionario de medicina y cirugía ó Biblioteca manual médico quirúrgica*, Madrid, Imprenta de la calle de la Greda, en línea, <https://www.bvfe.es/es/directorio-bibliografico-diccionarios-vocabularios-glosarios-tratados-y-obras-lexicografia/15265-diccionario-de-medicina-y-cirugia-o-biblioteca-manual-medico-quirurgica.html> [Consulta: 22/10/2020].	Medio
MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
HEATH, Christopher (1886): *Dictionary of Practical Surgery*, Londres, Smith Elder & Co.	Medio
THOMSON, Spencer; STEELE, J. C.; REID, Geo (1883): *A Dictionary of Domestic Medicine and Household Surgery,* London, Charler Griffin & Co.	Medio
GIBSON, John; POTPARIC, Oliverira (1992): *A Dictionary of Medical and Surgical Syndromes*, London, Taylor & Francis Ltd.	Bajo
BALLANTYNE, John William (2010*): Green's Encyclopedia and Dictionary of Medicine and Surgery V3 (1907)*, Whitefish, Kessinger Publishing	Bajo
PARKER, James; PARKER, Phillip (eds.) (2003): *Surgery – A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References*, s. l., ICON Health	Medio
COOPER, Samuel (1810): *A Dictionary of Practical Surgery: Containing a Complete Exhibition of the Present State of the Principles and Practice of Surgery, Collected from the Best and Most Original Sources of Information, and Illustrated by Critical Remarks. With Notes and Additions, by John Syng Dorsey, M. D.*, Philadelphia, B. & T. Kite	Medio
BILINGÜES	
Recurso	**Índice de calidad**
Glosarios especializados de Ciencias, Artes, Técnicas y Sociedad. Abreviaturas y siglas en cirugía (inglés-español), en línea, <https://glosarios.servidor-alicante.com/abreviaturas-siglas-cirugia_en-es> [Consulta: 22/10/2020].	Bajo
PUENTE FONSECA, Claudio Julio (1997): *Terminología inglés-español para la práctica médica*, Madrid, Ediciones Díaz de Santos.	Medio

Tabla 5. Recursos de la especialidad de cirugía.

4.6. Farmacología

En lo que respecta a recursos monolingües en español (cfr. Tabla 6) de la especialidad médica de farmacología tenemos disponibles varios recursos interesantes y que se encuentran en línea. Por ejemplo, la Sociedad Española de Farmacia Hospitalaria pone a nuestra disposición un amplio catálogo de *Terminología Farmacéutico-médica* (Casasín Edo 2020) en formato PDF que incluye abreviaturas y símbolos utilizados en la especialidad, así como ilustraciones en blanco y negro para aclarar conceptos. Queremos mencionar también el *Glosario de términos farmacológicos* llevado a cabo por el Centro Nacional de Información de Ciencias Médicas de Cuba. En formato papel, la misma Fundación Dr. Antonio Esteve nombrada anteriormente publica, de Laurence *et al*., un *Diccionario de farmacología y temas afines* (Carpenter y Laurence 2000) y *Un glosario para farmacólogos* (Laurence y Shaw 1987), donde se trata la terminología de la especialidad de farmacología desde el punto de vista del lenguaje real que utilizan los farmacólogos. Por último, mencionar, de la editorial Panamericana, el libro *Farmacología. Texto y Atlas* (Hein *et. al* 2010) que contiene un glosario de terminología del campo.

En cuanto a recursos lexicográficos monolingües en lengua inglesa (cfr. Tabla 6), la *Boston University School of Medicine* presenta en su web el Gl*ossary of Terms and Symbols Used in Pharmacology* (Pelikan 2004). El grueso de recursos sobre farmacología en inglés se encuentra en papel. Destacamos en un primer momento *A Dictionary of Pharmacology and Allied Topics* (Carpenter 1998) publicado por la editorial médica de conocido prestigio Elsevier junto a autoridades científicas de este campo como John Richard Carpenter. Debemos anotar la versión en inglés de la obra ya mencionada de Desmond Roger Laurence junto a John Richard Carpenter, *A Dictionary Of Pharmacology And Clinical Drug Evaluation* (Laurence y Carpenter 1994), publicada por CRC Press, la misma editorial que tiene disponible un *Dictionary of Pharmacological Agents* (Ganellin y Triggle 1996). Finalmente, otra de renombre es Springer, entre cuyas obras encon-

tramos *Concise Dictionary of Pharmacological Agents: Properties and Synonyms* y *Dictionary of Pharmaceutical Medicine* (Hall y Morton 1999), el cual contiene una lista de fármacos y sus propiedades además de un glosario descriptivo de términos técnicos. Finalmente, la editorial Wiley-Blackwell pone a disposición del usuario el *Dictionary of Pharmacology* (Bowman *et al.* 1986) escrito por el farmacéutico inglés W. C. Bowman.

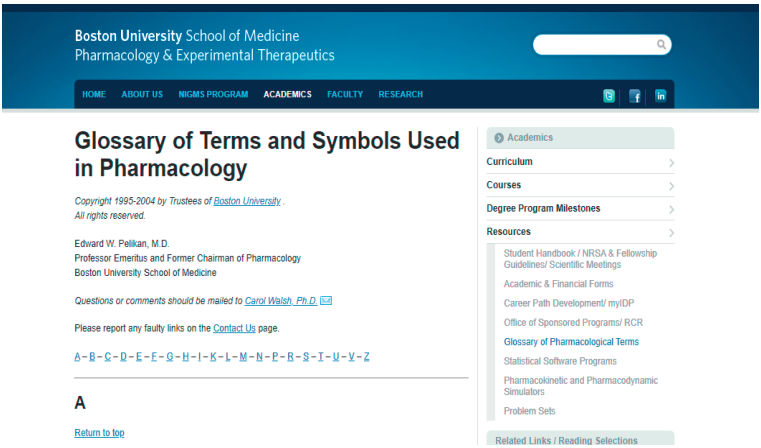


Figura 8. *Glossary of Terms and Symbols Used in Pharmacology* de la Boston University.

En lo que a recursos bilingües respecta (cfr. Tabla 6) la OMS colaboró con el Departamento de Salud de Austria para ofrecernos en formato PDF un *Glosario de términos farmacéuticos/Glossary of pharmaceutical terms* (WHO Collaborating Center for Pharmaceutical Pricing and Reimbursement Information 2012), donde la definición de los términos se encuentra en español. En formato papel disponemos de una obra enfocada a la traducción, el *Diccionario Terminológico de las Ciencias Farmacéuticas* (Alcaraz Varó *et al.* 2007), de la Real Academia Nacional de Farmacia y escrito por Alcaraz Varó *et al.* Esta obra reciente recopila unos 15.000 términos en inglés-español y 13.000 en español-inglés ordenados según 25 campos semánticos de la especialidad.

Para finalizar este apartado, debemos mencionar el recurso multilingüe que nos llega gracias a la prestigiosa editorial médica Elsevier (Sliosberg 1968-1980).

FARMACOLOGÍA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
GARCÍA GARCÍA, Antonio (coord..) (2003): «Diccionario de términos farmacológicos y médicos», *Actualidad en farmacología y terpáutica*, vol. 1, n. 3, p. 165, en línea, <http://www.socesfar.es/attachments/article/69/AFT-3-web.pdf> [Consulta: 22/10/2020].	Bajo
CASASÍN EDO, Tomás (2020): *Terminología Farmacéutico-médica*, en línea, <https://www.sefh.es/bibliotecavirtual/auxiliares/area3.pdf> [Consulta: 22/10/2020].	Medio
CIMA. Centro de Información online de Medicamentos de la AEMPS. Glosario de términos, en línea, <https://cima.aemps.es/cima/publico/glosario.html> [Consulta: 22/10/2020].	Bajo
DEL VILLAR RUIZ DE LA TORRE, José Alfonso.; MELO HERRÁIZ, Esther (2009): «Glosario de términos botánicos, farmacognósticos, farmacológicos, farmacéuticos y terapéuticos», *Guía de plantas medicinales del Magreb. Establecimiento de una conexión intercultural. Cuadernos de la Fundación Dr. Antonio Esteve*, n. 18, Barcelona: Fundación Dr. Antonio Esteve, en línea, <https://www.esteve.org/capitulos/glosario-de-terminos-botanicos-farmacognosticos-farmacologicos-farmaceuticos-y-terapeuticos/> [Consulta: 22/10/2020].	Bajo
RUIZ PACHECO, Marco; HURTADO MONTERO, Welington: *Glosario de Farmacología Terapéutica*, Universidad de Guayaquil, en línea, <https://es.scribd.com/document/363315190/Glosario-de-Farmacologia> [Consulta: 22/10/2020].	Bajo
AA. VV. (2014): *Glosario de términos farmacológicos*, en línea, <http://glosario.sld.cu/terminos-farmacologicos/> [Consulta: 22/10/2020].	Medio
LAURENCE, Desmond Roger; SHAW, Ian (1987): *Un glosario para farmacólogos,* Barcelona, Fundación Dr. Antonio Esteve	Medio
BOWMAN, William *et al.* (1988): *Diccionario de farmacología*, Barcelona, Medici.	Medio
CARPENTER, John Richard; LAURENCE, Desmond Roger (2000): *Diccionario de farmacología y temas afines*, Barcelona, Fundación Dr. Antonio Esteve.	Medio
HEIN, Lutz; LÜLLMANN, Heinz; MOHR, Klaus (2010): *Farmacología. Texto y Atlas*, 6ª ed., Madrid, Editorial Médica Panamericana SA.	Alto
MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
Stratech. Pharmacological Glossary, en línea, <https://www.stratech.co.uk/selleck_chemicals/pharmacological-glossary/> [Consulta: 22/10/2020].	Bajo
DICKENSON, John *et al.* (2012): «Glossary», en *Molecular Pharmacology: From DNA to Drug Discovery*, s. l., John Wiley & Sons, Ltd., pp. 381-402, en línea, <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118451908> [Consulta: 22/10/2020].	Bajo
PELIKAN, Edward (2004): *Glossary of Terms and Symbols Used in Pharmacology*, Boston University, en línea, <https://www.bumc.bu.edu/busm-pm/academics/resources/glossary/> [Consulta: 22/10/2020].	Medio

GANELLIN, Charon Robin; TRIGGLE, David (1996): *Dictionary of Pharmacological Agents*, Londres, Champan and Hall/CRC Press.	Medio
HEIN, Lutz; LÜLLMANN, Heinz; MOHR, Klaus (2017): *Color Atlas of Pharmacology*, 5ª ed., s. l., Thieme.	Bajo
CARPENTER, John Richard (1998): *A dictionary of pharmacology and allied topics*, Amsterdam, Elsevier.	Medio
HALL, Judith; MORTON, Ian (1999): *Concise Dictionary of Pharmacological Agents: Properties and Synonyms*, s. l., Springer Science+Business Media Dordrecht.	Alto
NAHLER, Gerhard *et al.* (2017): *Dictionary of Pharmaceutical Medicine*, 4ª ed., s. l., Springer International Publishing.	Medio
BOWMAN, William *et al.* (1986): *Dictionary of Pharmacology*, Oxford, Wiley-Blackwell.	Medio
LAURENCE, Desmond Roger; CARPENTER, John (1994): *A Dictionary Of Pharmacology And Clinical Drug Evaluation*, s. l., CRC Press.	Medio
BILINGÜES	
Recurso	**Índice de calidad**
WHO COLLABORATING CENTER FOR PHARMACEUTICAL PRICING AND REIMBURSEMENT INFORMATION (2012): *Glosario de términos farmacéuticos/Glossary of pharmaceutical terms*, en línea, <https://ppri.goeg.at/sites/ppri.goeg.at/files/inline-files/Spanish%20Glossary%20final.pdf> [Consulta: 22/10/2020].	Medio
SOTOLONGO MONTERO, Mercedes *et al.* (2007): «PHARMACOLOGICAL GLOSSARY», *Revista Habanera de Ciencias Médicas*, vol. 6, n. 3, pp.1-14, en línea, <https://www.redalyc.org/articulo.oa?id=180417641015> [Consulta: 22/10/2020].	Bajo
BENÍTEZ GENER, Aurora *et al.* (2007): «Glosario de términos farmacológicos. Glossary of pharmacological terms», *Revista de Ciencias Médicas de Pinar del Río*, vol. 11, n. 5, pp. 119-128, en línea, <http://www.revcmpinar.sld.cu/index.php/publicaciones/article/view/391/839> [Consulta: 22/10/2020].	Bajo
ALCARAZ VARÓ, Enrique *et al.* (2007): *Diccionario Terminológico de Ciencias Farmacéuticas*, Barcelona, Ariel.	Alto
ZAYAS, Hilda (2005): *Spanish Chemical and Pharmaceutical Glossary*, Rockville, Schreiber Publishing.	Bajo
MULTILINGÜES	
Recurso	**Índice de calidad**
SLIOSBERG, Anatole (1968-1980): *Elsevier's dictionary of pharmaceutical science and techniques, in five languages: English, French, Italian, Spanish, German*, Amsterdam, London, New York, Elsevier.	Bajo

Tabla 6. Recursos de la especialidad de farmacología.

4.7. Botánica

Como hemos mencionado en la primera parte de este trabajo, dentro de la farmacología nos pueden resultar útiles recursos de ciencias afines como pueden ser la botánica (al que se dedica este subapartado) y la fitoterapia (siguiente subapartado).

En lo que respecta a recursos monolingües en español disponibles en línea para la botánica no hemos localizado ninguno de fiabilidad o calidad media o alta (cfr. Tabla 7). Recursos monolingües en papel en español encontramos algunos más. La editorial VOX nos propone dos obras diferentes en las que han trabajado diferentes autores de manera conjunta: *Lexis/22 VOX. Botánica* (AA. VV. 1980) y *Gran Vox. Botánica y Zoología* (Colàs Gil 2002), ambos comparten algunos términos, incluyen el nombre común y el nombre científico y muestran imágenes para algunas de las entradas, aunque el primero, al centrarse únicamente en botánica, incluye también mapas y un árbol filogenético del reino vegetal. A título personal encontramos una obra más: *Diccionario de Botánica* (Font Quer 1970), realizado por el prestigioso botánico, farmacéutico y químico español Pío Font Quer y donde se incluye información gramatical y etimológica, numerosos dibujos explicativos, así como términos de especialidades afines a la botánica.

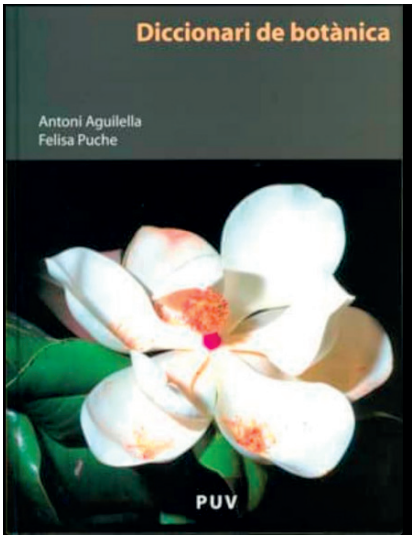


Figura 9. *Diccionari de Botànica* (Aguilella y Puche 2004) de la Universidad de Valencia.

Hemos recogido dos recursos monolingües en inglés en papel (cfr. Tabla 7), en concreto, un recurso del siglo XIX como es *The Encyclopaedia or the Treasury of Botany of the Vegetable Kingdom: With Which Is Incorporated a Glossary of Botanical Terms* (Lindley y Moore 1995[1873]) y *Botanica: The Illustrated A-Z of Over 10.000 Garden Plants and How to Cultivate Them* (Olds y Page 2004) de la editorial Könemann, que tiene un gran valor por el amplísimo número de fotografías a color que contiene, además de las que en las definiciones se incluye el nombre científico, el nombre común y sinónimos, así como un mapa con las zonas donde se puede encontrar cada planta.

Centrándonos en recursos bilingües, destacamos dos disponibles en línea y una obra en papel, los tres estadounidenses y avalados por alguna institución de prestigio o universidad, pero que no alcanzan el nivel de calidad medio o alto (cfr. Tabla 7).

Acabamos nuestro análisis de los recursos de botánica mencionando un recurso multilingüe avalado por la Universidad de Valencia: *Diccionari de Botánica* (Aguilella y Puche 2004), el cual contiene más de 5000 términos. Las entradas se encuentran ordenadas alfabéticamente en valenciano (misma lengua de las definiciones), con las respectivas equivalencias al español, inglés y francés.

BOTÁNICA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
Universidad Nacional de la Palma. Glosario de términos botánicos, en línea, <https://sistematicavegetal.weebly.com/uploads/8/0/5/2/8052174/glosario_de_terminos_botanicos_facagronomaunlapa.pdf> [Consulta: 22/10/2020].	Bajo
Botanical-online. Glosario de Botánica, en línea, <https://www.botanical-online.com/diccionario> [Consulta: 22/10/2020].	Bajo
AA. VV. (1980): *Lexis/22 VOX. Botánica*, Barcelona, Círculo de Lectores.	Medio
FONT QUER, Pío (1970): *Diccionario de Botánica,* Barcelona, Editorial Labor.	Medio
ÁLVAREZ SÁNCHEZ, Julio (1979): *Diccionarios Rioduero. Botánica*, Madrid, Ediciones Rioduero.	Bajo
COLÀS GIL, Jaume (coord..) (2002): *Gran Vox. Botánica y Zoología*, 8ª ed., Barcelona, SPES EDITORIAL.	Medio
TOSCO, Uberto (1973): *Diccionario de Botánica*, Barcelona, Editorial Teide.	Bajo

MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
Botanical-online. Glossary of Botany, en línea, <https://www.botanical-online.com/en/glossary> [Consulta: 22/10/2020].	Bajo
Plant Facts. Glossary, en línea, <https://plantfacts.osu.edu/wiki/index.php?title=Main_Page> [Consulta: 22/10/2020].	Bajo
University of California Museum of Paleontology. UCMP Glossary: Botany, en línea, <https://ucmp.berkeley.edu/glossary/gloss8botany.html> [Consulta: 22/10/2020].	Bajo
FloraBase, the Western Australian Flora. Glossary of Botanical Terms, en línea, <https://florabase.dpaw.wa.gov.au/help/glossary> [Consulta: 22/10/2020].	Bajo
LINDLEY, John; MOORE, Thomas (1995[1873]): *The Encyclopaedia or the Treasury of Botany of the Vegetable Kingdom: with which is incorporated a Glossary of Botanical Terms*, Nueva Delhi, Low Price Publications.	Medio
HICKEY, Michael; KING, Clive (2014): *The Cambridge Illustrated Glossary of Botanical Terms*, Cambridge, Cambridge University Press.	Bajo
OLDS, Margaret; PAGE, Susan (eds.) (2004): *Botanica: the illustrated a-z of over 10.000 garden plants and how to cultivate them*, Königswinter, KÖNEMANN.	Medio
BILINGÜES	
Recurso	**Índice de calidad**
MEDINA, Alvin Leroy (1988): *An English-Spanish glossary of terminology used in forestry, range, wildlife, fishery, foils and botany. Glosario en inglés-español de terminología usados en forestales, pastizales, fauna silvestre, pesquería, suelos y botánica*, Fort Collins, U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, en línea, <https://pclive.peacecorps.gov/pclive/index.php/pclive-resources/resource-library/224-r0091-glossary-of-terminology-spanish-english/file> [Consulta: 22/10/2020].	Bajo
Missouri Botanical Garden. Tropicos. Glossary, en línea, <http://legacy.tropicos.org/GlossarySearch.aspx> [Consulta: 22/10/2020].	Bajo
SUGDEN, Andrew (1997): *Diccionario ilustrado de la botánica*, León, EDITORIAL EVEREST.	Bajo
MULTILINGÜES	
Recurso / **Enlace**	**Índice de calidad**
AGUILELLA, Antoni; PUCHE, Felisa (2004): *Diccionari de botánica*, Valencia, PUV.	Medio
CAUWET-MARC, Anne-Marie; VALLES XIRAU, Joan; VIGO, Josep (2004): *Vocabulari de Botànica*, Castellón de la Plana, Xarxa d'universitats. Institut Joan Lluís Vives.	Bajo

Tabla 7. Recursos de la especialidad de botánica.

4.8. Fitoterapia

Dentro de los recursos monolingües en español disponibles en línea (cfr. Tabla 8) para la fitoterapia destacamos, en primer lugar, el *Manual de uso de hierbas medicinales del Paraguay* realizado por la Fundación Celestina Pérez de Almada de Paraguay con el apoyo de la UNESCO, que incluye una lista de plantas medicinales, sus propiedades y el nombre científico. En segundo lugar, otro recurso disponible

en línea, y avalado por la Universidad Central Marta Abreu de las Villas de Cuba, es el *Glosario etimológico de las plantas medicinales con uso antiasmático registradas en Cuba* (Rojas Fraguela 2013), con información etimológica. En papel destacamos cuatro recursos: *Fitoterapia, Vademécum de Prescripción* (Vanaclocha y Cañigueral 2019) que incluye un glosario además de información sobre las plantas medicinales, fotografías, indicaciones terapéuticas y fórmulas magistrales, redactado por Vanaclocha y Cañigueral de la editorial médica Elsevier; *Plantas Medicinales. El Dioscorides Renovado* (Font Quer 1990) por el reconocido autor Pío Font Quer y publicado en la editorial Labor donde se incluyen más de 11.000 términos, mapas e imágenes; *Gran Enciclopedia de las Plantas Medicinales. Terapia Natural Para el Tercer Milenio* (Berdonces i Serra 2002) de la editorial Tikal-Susaetta y *Guía Ilustrada de plantas medicinales* (Thompson 1980) publicada en Blume, que incluye nombre común y nombre científico, referencias diatópicas y dibujos a color de las entradas.

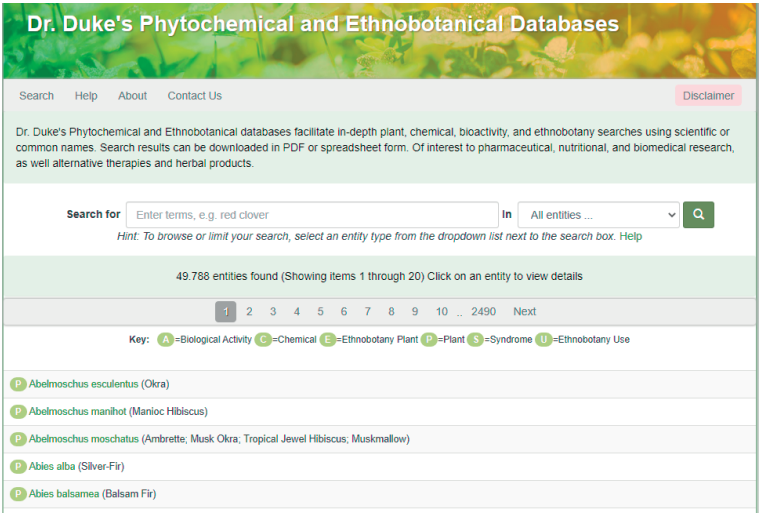


Figura 10. *Dr. Duke's Phytochemical and Ethnobotanical Databases* (AA. VV. 2016b).

Si pasamos a los recursos monolingües en inglés en línea (cfr. Tabla 8) sobresalen tres recursos avalados por universidades o entidades gubernamentales de Estados Unidos: *Dr. Duke's Phytochemical and Ethnobotanical Databases* del U.S. Department of Agriculture, Agricultural Research Service (AA. VV. 2016b), *Herbal Medicine: Expanded Commission E Monographs* del American Botanical Council (AA. VV. 2021c) y *An Herbal/Medical Dictionary* (Moore 1995) de la Southwest School of Botanical Medicine. Otra entidad menor como *The Complementary Medical Association* (Reino Unido) publica en su página web el recurso *A-Z Glossary of Terms Used in Herbal Medicine* (AA. VV. 2021a). En cuanto a recursos en inglés en papel citamos recursos de editoriales de ámbito más general que tienen publicaciones relativas a la fitoterapia: *Encyclopedia of Herbal Medicine: 550 Herbs and Remedies for Common Ailments* (Chevallier 2016), donde se incluye amplia información de cada una de las entradas e imágenes a todo color.

Para concluir, queremos mencionar un recurso multilingüe titulado *Plantas medicinales* (Vanaclocha i Vanaclocha y Cañigueral Folcarà 2020) mencionado anteriormente en papel y en español y que, con la traducción de los términos en varios idiomas, está disponible en línea en la web Fitoterapia.

FITOTERAPIA	
MONOLINGÜES ESPAÑOL	
Recurso	**Índice de calidad**
ROJAS FRAGUELA, Yadira (2013): *Glosario etimológico de las plantas medicinales con uso antiasmático registradas en Cuba*, Universidad Central "Marta Abreu" de las Villas, Trabajo de diploma, en línea, <https://dspace.uclv.edu.cu/bitstream/handle/123456789/504/Glosario%20etimol%C3%B3gico%20de%20las%20plantas%20....pdf?sequence=1&isAllowed=y> [Consulta: 22/10/2020].	Medio
AA. VV. (2002a): *Manual de uso de hierbas medicinales del Paraguay*, en línea, <http://portal.unesco.org/en/file_download.php/c9010dd7f603adeb359ff68830c3c978hierbasmedicinales.pdf> [Consulta: 22/10/2020].	Medio
Espacio el árbol. Glosario para Fitoterapia, en línea, <http://espacioelarbol.cl/glosario-para-fitoterapia/> [Consulta: 22/10/2020].	Bajo
BERDONCES I SERRA, Josep Lluís (2002): *Gran Enciclopedia de las Plantas Medicinales. Terapia Natural Para el Tercer Milenio*, Madrid, Tikal-Susaeta.	Alto
FONT QUER, Pío (1990): *Plantas Medicinales. El Dioscorides Renovado*, 14ª ed., Barcelona, Labor.	Medio

VANACLOCHA I VANACLOCHA, Bernat; CAÑIGUERAL FOLCARÀ, Salvador (eds.) (2019): *Fitoterapia, Vademécum de Prescripción*, 5ª ed., Barcelona, Elsevier.	Alto
THOMPSON, William (1980a): *Guía Ilustrada de plantas medicinales*, Barcelona, Blume.	Medio
MONOLINGÜES INGLÉS	
Recurso	**Índice de calidad**
The Complementary Medical Association. A-Z GLOSSARY OF TERMS USED IN HERBAL MEDICINE, en línea, <https://www.the-cma.org.uk/articles/az-glossary-of-terms-used-in-herbal-medicine-a-3325/> [Consulta: 22/10/2020].	Medio
AA. VV. (2021a): *Dr. Duke's Phytochemical and Ethnobotanical Databases*, en línea, <https://phytochem.nal.usda.gov/phytochem/help/index/about> [Consulta: 22/10/2020].	Medio
AA. VV. (2021c): *Herbal Medicine: Expanded Commission E Monographs*, en línea, <https://www.herbalgram.org/resources/expanded-commission-e/> [Consulta: 22/10/2020].	Medio
MOORE, Michael (1995): *An Herbal/Medical Dictionary. A Glossary of terms used in Herbalism, Medicine and Physiology. Descriptions, Explanations and Implications in Wholistic and Vitalist Therapy, Bisbee, Southwest School of Botanical Medicine*, en línea, <https://www.swsbm.com/ManualsMM/MedHerbGloss2.pdf> [Consulta: 22/10/2020].	Medio
Walsh Botanical's. Phytotherapeutics terminology, en línea, <https://www.walshbotanicals.net/phytotherapeutic-terminology.html> [Consulta: 22/10/2020].	Bajo
Integrative Lifestyle. Glossary of Phytomedicine, en línea, <https://integrativelifestyle.org/glossary-of-phytomedicine/> [Consulta: 22/10/2020].	Bajo
THOMPSON, William (1980b): *Healing Plants - A Modern Herbal*, Londres, Macmillan & Co.	Bajo
HARROP, Renny (1979): *Encyclopedia of Herbs*, Londres, Marshall Cavendish.	Bajo
CAPASSO, Francesco *et al.* (2003): *Phytotherapy: A Quick Reference to Herbal Medicine*, Berlín, Springer Science & Business Media.	Bajo
CHEVALLIER, Andrew (2016). *Encyclopedia of Herbal Medicine: 550 Herbs and Remedies for Common Ailments*, Londres, Dorling Kindersley Ltd.	Alto
BILINGÜES	
Recurso	**Índice de calidad**
BRITO LASERRA, Franklin *et al.* (2010): «Breve glosario bilingüe inglés-español de plantas medicinales. Brief bilingual English-Spanish glossary of herbs», *MEDISAN*, vol. 14, n. 8, Santiago de Cuba, en línea, <http://scielo.sld.cu/pdf/san/v14n8/san17810.pdf> [Consulta: 22/10/2020].	Bajo
MULTILINGÜES	
Recurso	**Índice de calidad**
VANACLOCHA I VANACLOCHA, Bernat; CAÑIGUERAL FOLCARÀ, Salvador: *Fitoterapia.net. Plantas medicinales*, en línea, <https://www.fitoterapia.net/vademecum/plantas/index.html?opt=abc&letra=A> [Consulta: 22/10/2020].	Alto

Tabla 8. Recursos de la especialidad de fitoterapia.

5. Discusión

Los datos ofrecidos en el apartado precedente se recogen cuantitativamente en las siguientes tablas (cfr. Tabla 9 y cfr. Tabla 10):

	ESPAÑOL		INGLÉS		BILINGÜE		MULTILINGÜE		TOTAL
ESPECIALIDAD	Online	Papel	Online	Papel	Online	Papel	Online	Papel	
Ginecología	8	3	7	2	3	0	0	0	23
Traumatología	2	1	3	0	1	0	0	0	7
Cardiología	6	3	11	7	4	1	0	0	32
Oncología	10	5	9	5	3	0	0	0	32
Cirugía	3	1	0	6	2	0	0	0	12
Farmacología	6	4	3	7	3	2	1	0	26
Botánica	2	5	4	3	2	1	0	2	19
Fitoterapia	3	4	6	4	1	0	1	0	19
SUBTOTAL	40	26	43	34	19	6	2	2	170
TOTAL	66		77		23		4		

Tabla 9. Extracción de datos cuantitativos de la totalidad de los recursos analizados en cada especialidad.

A tenor de los datos numéricos contenidos en la tabla precedente, conviene subrayar que llama la atención que existen, en general, menos recursos bilingües que monolingües (ya sean diccionarios o glosarios) en cualquiera de las especialidades médicas presentes en este estudio para un par de lenguas tan común como es el de inglés-español. El carácter del *lingua franca* del inglés con respecto a la ciencia se manifiesta en la abundancia de recursos monolingües en inglés, que supera con mucho a los recursos bilingües como a los recursos monolingües en español. En este aspecto, hay que señalar que los recursos multilingües se quedan a mucha distancia del resto de recursos.

	ESPAÑOL		INGLÉS		BILINGÜE		MULTILINGÜE		TOTAL
ESPECIALIDAD	Online	Papel	Online	Papel	Online	Papel	Online	Papel	
Ginecología	3	2	0	2	0	0	0	0	7
Traumatología	1	1	0	0	1	0	0	0	3
Cardiología	2	1	0	6	0	1	0	0	10
Oncología	2	2	1	4	1	0	0	0	10
Cirugía	3	1	0	4	1	0	0	0	9
Farmacología	2	4	1	6	1	1	0	0	15
Botánica	0	3	0	2	0	0	0	1	6
Fitoterapia	2	4	4	1	0	0	1	0	12
SUBTOTAL	15	17	6	25	4	2	1	1	
TOTAL	32		31		6		2		71

Tabla 10. Extracción de datos cuantitativos de los recursos con una fiabilidad media o alta en cada especialidad.

Asimismo, y pesar de que traumatología es una de las especialidades médicas que más volumen de trabajo supone para los traductores, encontramos una manifiesta escasez de recursos lexicográficos. Si bien hay que mencionar que en su gran mayoría estos recursos están avalados por instituciones y editoriales médicas o científicas de prestigio.

Las especialidades que más recursos presentan, tanto en papel como en línea, son cardiología y oncología. Además, instituciones de importancia reconocida o universidades los respaldan.

Se hace patente, sobre todo en la especialidad de cardiología, que la amplia mayoría de los recursos monolingües se encuentra en lengua inglesa, 18 recursos, frente a los 9 en lengua española, quizá debido a que el inglés es, como hemos apuntado, la *lingua franca* del conocimiento y la investigación científica.

Como ya hemos indicado, la mayoría de los recursos lexicográficos que encontramos en línea para la especialidad de cirugía son específicos de algún tipo de cirugía (oral, plástica, estética...). Por esto, se hace patente cierta escasez en los recursos de cirugía general. Todos los recursos tratados en este estudio sobre cirugía general monolingües en inglés se encuentran en papel. También hay que destacar que un gran número de los recursos lexicográficos de cirugía en papel mencionados en este estudio (tanto en español como en inglés) son del siglo XIX, por lo que estarían desactualizados. Sin embargo, resulta sorprendente que estas obras monolingües en español están escaneadas y disponibles en línea.

Finalmente, hay que destacar la gran cantidad de recursos bilingües en línea y en papel que se han encontrado sobre la especialidad médica de farmacología en comparación con otras especialidades, cuando se valora el número de recursos por especialidad en su totalidad (cfr. Figura 11).

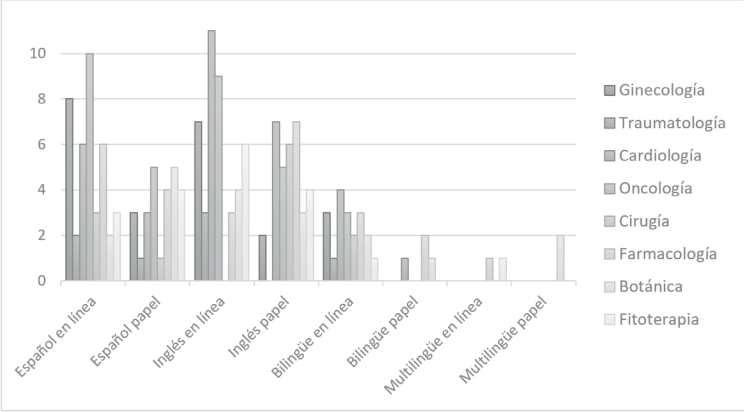


Figura 11. Número de recursos por especialidad
(fiabilidad o calidad alta, media y baja).

Igualmente ocurre cuando se analiza el de recursos por especialidad con fiabilidad o calidad alta y media, únicamente (cfr. Figura 12).

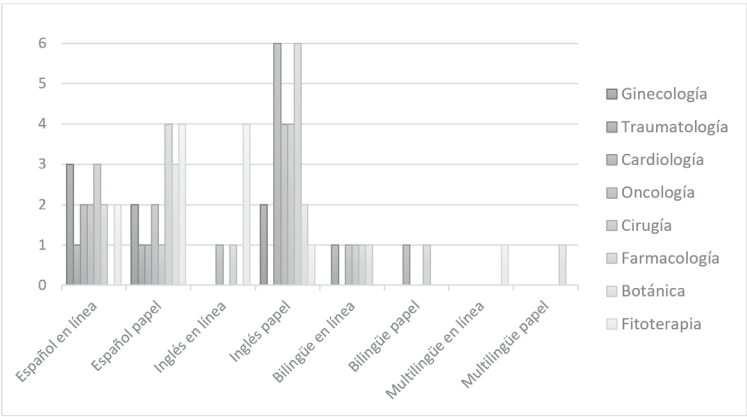


Figura 12: Número de recursos por especialidad
(fiabilidad o calidad alta y media, únicamente).

Todos estos recursos tienen detrás un organismo importante (universidades, Organización Mundial de la Salud) o son específicos para traductores, lo cual tiene mucho valor para nuestro estudio.

6. Conclusiones

En el presente trabajo se ha observado la escasez de recursos en estas áreas de la medicina con tan alta demanda, tanto por especialistas médicos, como por traductores. Quizá sería conveniente ampliar los recursos ya existentes o crear recursos propios, a fin de que sean lo más adecuados posible. Asimismo, se pone de manifiesto que, de los escasos recursos que existen, todavía son menos los que, de acuerdo con nuestra plantilla de evaluación (cfr. Escribano Meseguer 2021 en prensa), superan unos ciertos estándares de calidad. Sin embargo, aquellos que sí lo hacen conforman una muestra de recursos de los que se asegura la calidad, y que pueden resultar de gran utilidad para médicos y otro personal sanitario, traductores, intérpretes y para los especialistas en lingüística de corpus que deseen contar con recursos de calidad para compilar sus propios corpus.

Referencias bibliográficas

Barber Pérez, Patricia; González López-Valcárcel, Beatriz (2019): *Estimación de la oferta y demanda de médicos especialistas. España 2018-2030*, Las Palmas de Gran Canaria, Universidad de Las Palmas de Gran Canaria.

Cochrane (2020): *Cochrane. Complementary Medicine. Operational definition of complementary medicine*. En línea, <https://cam.cochrane.org/operational-definition-complementary-medicine> [Consulta: 22/10/2020].

Corpas Pastor, Gloria; Roldán Juárez, Marina (2014): «Análisis de necesidad documentales y terminológicas de médicos y traductores médicos como base para el diseño de un diccionario multilingüe de nueva generación», *MonTI*, n. 6, pp. 167-202.

Cortés, Nacho (2020): *MIR 2020: La oferta total de plazas por especialidades y comunidades autónomas*. En línea, <https://www.consalud.es/especial-mir/mir-2020-oferta-plazas-especialidades-comunidades-autonomas_81218_102.html> [Consulta: 22/10/2020].

Escribano Meseguer, Lucía (2021, en prensa): «Implementación de una plantilla electrónica y puntuable para el análisis y la evaluación de glosarios, léxicos y diccionarios de Botánica», en *Actas de las IV Jornadas Doctorales organizadas por*

el Programa de Doctorado en Lingüística, Literatura y Traducción, Málaga, Universidad de Málaga.

European Commission (2017): *2017 Language Industry Survey – Expectations and Concerns of the European Language Industry*. En línea, <http://www.digiling.eu/wp-content/uploads/2017/06/2017-Language-Industry-Survey-Report_6April2017.pdf> [Consulta: 22/10/2020].

García Esteban, Ana Isabel; Varela Salinas, María José (2009): «Propuesta de una unidad didáctica de traducción biosanitaria alemán-español», en María José Varela Salinas (ed.), *Panorama actual del estudio y la enseñanza de discursos especializados*, Bern, Peter Lang, pp. 123-144.

García-Izquierdo, Isabel; Conde, Tomás (2014): «Necesidades documentales del traductor médico en España», *TRANS. Revista de Traductología*, n. 18, pp. 141-162.

Hurtado Albir, Amparo (2001): *Traducción y traductología*, Madrid, Cátedra.

Montalt Resurrecció, Vicent; Shuttleworth, Mark (2012): «Research in Translation and Knowledge Mediation in Medical and Healthcare Settings», *Linguistica Antverpiensia*, n. 11, pp. 9-29.

MSPSI = Ministerio de Sanidad, Política Social e Igualdad (2011): *Análisis de situación de las terapias naturales*, Madrid, MSPSI.

Muñoz-Miquel, Ana (2016): «La traducción médica como especialidad académica: algunos rasgos definitorios», *Hermēneus*, n. 18, pp. 235-267.

Navas Romero, Cristina; Palomares Perraut, Rocío (2002): «Un estudio del mercado español de la traducción en la internet", *Translation Journal*, vol. 6, n. 2, s. p.

NCCIH (2018): «Complementary, Alternative, or Integrative Health: What's In a Name?». En línea, <https://www.nccih.nih.gov/health/complementary-alternative-or-integrative-health-whats-in-a-name> [Consulta: 22/10/2020].

Negrete, Borja (2020): «MIR 2020: todas las plazas agotadas en el número de orden 9.275». En línea, <https://www.redaccionmedica.com/secciones/formacion/mir-2020-ultimo-dia-plazas-agotadas-numero-de-orden-9-275-6616> [Consulta: 22/10/2020].

Organización Mundial de la Salud (2002): *Estrategia de la OMS sobre medicina tradicional 2002-2005*, Gèneve, Organización Mundial de la Salud.

Rico Pérez, Celia; García Aragón, Álvaro (2016): *Análisis del sector de la traducción en España (2014-2015)*, Campus de Villaviciosa de Odón, Universidad Europea.

Sánchez Ramos, María del Mar (2017): «Compilación y análisis de un corpus *ad hoc* como herramienta de documentación electrónica en Traducción e Interpretación en los Servicios Públicos (TISP)», *Estudios de Traducción*, vol. 7, pp. 177-190.

Seghiri, Miriam (2006): *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*, Málaga, SPICUM.

UNESCO (1988): *Proposed International Standard Nomenclature for Fields of Science and Technology*, Paris, UNESCO.

Vázquez y del Árbol, Esther (2006): «La redacción del discurso biomédico (inglés-español): rasgos principales», *Panace@*, vol. 7, n. 24, pp. 307-317.

Zhang, Xiaorui: «Medicina tradicional: definiciones». En línea, <https://www.who.int/topics/traditional_medicine/definitions/es/> [Consulta: 22/10/2020].

Corpus lexicográfico

AA. VV. (1980): *Lexis/22 VOX. Botánica*, Barcelona, Círculo de Lectores.

Aguilella, Antoni; Puche, Felisa (2004): *Diccionari de botánica*, Valencia, PUV.

Alcaraz Varó, Enrique *et al.* (2007): *Diccionario Terminológico de Ciencias Farmacéuticas*, Barcelona, Ariel.

Alegría Ezquerra, Eduardo; Teresa Galván, Eduardo de (2005): *Glosario de términos y abreviaturas en cardiología inglés-español*, Madrid, Acción Médica.

Anderson, Robert *et al.* (1992): *Diccionario de cardiología ilustrado para el médico clínico*, Barcelona, Science Press.

Anguita Sánchez, Manuel (dir.) (2009): *Diccionario LID Cardiología*, Madrid, LID Editorial Empresarial S.L.

Bastos Mora, Felipe (1999): *Diccionario Enciclopédico ilustrado de traumatología*, Barcelona, Editorial Médica JIMS.

Becerro de Bengoa Callau, Claudio; Cruz y Hermida, Julio (2017a): *Diccionario de Obstetricia y Ginecología ilustrado*, España, Marbán.

Becerro de Bengoa Callau, Claudio; Cruz y Hermida, Julio (2017b): *Diccionario de Términos Obstétrico-Ginecológico*, España, Marbán.

Benlloch Rodríguez, Raquel *et al.* (2018): *Diccionario de términos oncológicos para los no oncólogos*, s. l., Edición Punto Didot.

Berdonces i Serra, Josep Lluís (2002): *Gran Enciclopedia de las Plantas Medicinales. Terapia Natural Para el Tercer Milenio*, Madrid, Tikal-Susaeta.

Bowman, William *et al*. (1986): *Dictionary of Pharmacology*, Oxford, Wiley-Blackwell.

Carpenter, John Richard (1998): *A Dictionary of Pharmacology and Allied Topics*, Amsterdam, Elsevier.

Carpenter, John Richard; Laurence, Desmond Roger (1994): *A Dictionary of Pharmacology and Clinical Drug Evaluation*, London, UCL Press.

Carpenter, John Richard; Laurence, Desmond Roger (2000): *Diccionario de farmacología y temas afines*, Barcelona, Fundación Dr. Antonio Esteve.

Chevallier, Andrew (2016): *Encyclopedia of Herbal Medicine: 550 Herbs and Remedies for Common Ailments*, London, Dorling Kindersley Ltd.

Colàs Gil, Jaume (coord.) (2002): *Gran Vox. Botánica y Zoología*, 8.ª ed., Barcelona, Spes Editorial.

Cooper, Samuel (1810): *A Dictionary of Practical Surgery: Containing a Complete Exhibition of the Present State of the Principles and Practice of Surgery, Collected from the Best and Most Original Sources of Information, and Illustrated by Critical Remarks. With Notes and Additions, by John Syng Dorsey, M. D.*, Philadelphia, B. & T. Kite.

Dox, Ida; Melloni, June; Sheld, Harrison (2000): *Melloni's Illustrated Dictionary of Obstetrics and Gynecology*, s. l., CRC Press.

Font Quer, Pío (1970): *Diccionario de Botánica,* Barcelona, Editorial Labor.

Font Quer, Pío (1990): *Plantas Medicinales. El Dioscorides Renovado*, 14.ª ed., Barcelona, Labor.

Ganellin, Charon Robin; Triggle, David (1996): *Dictionary of Pharmacological Agents*, London, Champan and Hall/CRC Press.

González de Francisco, Tomás; Lacal, Juan Carlos (2009): *Diccionario LID Oncología*, s. l., LID Editorial Empresarial S. L.

Hall, Judith; Morton, Ian (1999): *Concise Dictionary of Pharmacological Agents: Properties and Synonyms*, s. l., Springer Science+Business Media Dordrecht.

Hayes, David; Asirvatham, Samuel (2007): *Dictionary of Cardiac Pacing, Defibrillation, Resynchronization, and Arrhythmias*, 2.ª ed., Minneapolis, Cardiotext Publishing.

Heath, Christopher (1886): *Dictionary of Practical Surgery*, London, Smith Elder & Co.

Hein, Lutz *et al.* (2010): *Farmacología. Texto y Atlas*, 6.ª ed., Madrid, Editorial Médica Panamericana SA.

Laughlin, Edward (2008): *CANCER from A to Z: A Dictionary of Cancer-Related Terms*, Bloomington, AuthorHouse.

Laurence, Desmond Roger; Carpenter, John (1994): *A Dictionary Of Pharmacology And Clinical Drug Evaluation*, s. l., CRC Press.

Laurence, Desmond Roger; Shaw, Ian C. (1987): *Un glosario para farmacólogos*, Barcelona, Fundación Dr. Antonio Esteve.

Lindley, John; Moore, Thomas (1995[1873]): *The Encyclopaedia or the Treasury of Botany of the Vegetable Kingdom: With Which is Incorporated a Glossary of Botanical Terms*, New Delhi, Low Price Publications.

Munro, Alastair (2001): *Modern Oncology: An A-Z of Key Topics*, Cambridge, Cambridge University Press.

Olds, Margaret; Page, Susan (eds.) (2004): *Botanica: The Illustrated A-Z of Over 10.000 Garden Plants and How to Cultivate Them*, Königswinter, Könemann.

Parker, James; Parker, Phillip (eds.) (2003): *Surgery – A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References*, s. l., ICON Health Publications

Parker, James; Parker, Phillip (eds.) (2004): *Cancer – A Medical Dictionary, Bibliography, and Annotated Research Guide to Internet References*, s. l., ICON Health Publications.

Puente Fonseca, Claudio Julio (1997): *Terminología inglés-español para la práctica médica*, Madrid, Ediciones Díaz de Santos, SA.

RANM = Real Academia Nacional de Medicina de España (2011): *Diccionario de términos médicos*, s. l. Editorial Médica Panamericana S.A.

Reuter, Peter (2010): *Springer Pocket Dictionary of Cardiology*, s. l., JP Medical Publishers.

Rhodes, Sharon; Dorland; David, Mary (2001a): *Dorland's Cardiology Word Book for Medical Transcriptionist*, Philadelphia, Saunders.

Rhodes, Sharon; Dorland; Pugh, Janet C. (2001b): *Dorland's Radiology/Oncology Word Book for Medical Transcriptionists*, Philadelphia, Saunders.

Sleight, Peter (1960): *International Dictionary of Cardiology*, London, Radcliffe Publishing Ltd.

Sliosberg, Anatole (1968-1980): *Elsevier's Dictionary of Pharmaceutical Science and Techniques, in Five Languages: English, French, Italian, Spanish, German*, Amsterdam, London, New York, Elsevier.

Stedman, Thomas; Beatty, William (1999): *Diccionario de Ciencias Médicas Stedman bilingüe (Español-Inglés)*, s. l., Editorial Médica Panamericana, S.A.

Thompson, William (1980): *Guía Ilustrada de plantas medicinales*, Barcelona, Blume.

Thomson, Spencer; Steele, J. C. (1883): *A Dictionary of Domestic Medicine and Household Surgery,* London, Charles Griffin & Co.

Vanaclocha, Bernat; Cañigueral, Salvador (eds.) (2019): *Fitoterapia, Vademécum de Prescripción*, 5.ª ed., Barcelona, Elsevier.

Zink, Christoph (1988): *Diccionario Pschyrembel de Ginecología y Obstetricia*, Berlin, Walter de Gruyter.

Corpus lexicográfico en línea

AA. VV. (2001): *Gestión de Pacientes Ambulatorios en Cardiología*. En línea, <http://www.comsor.es/pdf/formacion/Seccion%2001/gestion_pacientes_ambulatorios_en_cardiologia.pdf> [Consulta: 22/10/2020].

AA. VV. (2002a): *Manual de uso de hierbas medicinales del Paraguay*. En línea, <http://portal.unesco.org/en/file_download.php/c9010dd7f603adeb359ff68830c3c978hierbasmedicinales.pdf> [Consulta: 22/10/2020].

AA. VV. (2002b): *OncoTerm. Sistema Bilingüe de Información y Recursos Oncológicos*. En línea, <https://www.ugr.es/~oncoterm/index.html> [Consulta: 22/10/2020].

AA. VV. (2014): *Glosario de términos farmacológicos*. En línea, <http://glosario.sld.cu/terminos-farmacologicos/> [Consulta: 22/10/2020].

AA. VV. (2016a): *Diccionario Ginecológico del Hospital Gregorio Marañón*. En línea, <https://booksmedicos.org/diccionario-ginecologico/> [Consulta: 22/10/2020].

AA. VV. (2016b): *Dr. Duke's Phytochemical and Ethnobotanical Databases*. En línea, <https://phytochem.nal.usda.gov/phytochem/help/index/about> [Consulta: 22/10/2020].

AA. VV. (2020): *Glosario Ginecológico del Centro Clínico Betanzos 60*. En línea, <https://centroclinicobetanzos60.es/medicos-especialistas/glosario-2/> [Consulta: 22/10/2020].

AA. VV. (2021a): *A-Z Glossary of Terms Used in Herbal Medicine*. En línea, <https://www.the-cma.org.uk/articles/az-glossary-of-terms-used-in-herbal-medicine-a-3325/> [Consulta: 22/10/2020].

AA. VV. (2021b): *Diccionario del cáncer del NCI*. En línea, <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/?redirect=true> [Consulta: 22/10/2020].

AA. VV. (2021c): *Herbal Medicine: Expanded Commission E Monographs*. En línea, <https://www.herbalgram.org/resources/expanded-commission-e/> [Consulta: 22/10/2020].

AA. VV. (2021d): *Diccionario Temático de Medicina. Especialidad: Cardiología*. En línea, <https://www.portalesmedicos.com/diccionario_medico/index.php?title=Cardiologia&no_redirect=true> [Consulta: 22/10/2020].

AECC (2007a): *Glosario del Cáncer*. En línea, <https://www.aecc.es/sites/default/files/migration/actualidad/publicaciones/documentos/glosario.pdf> [Consulta: 22/10/2020].

AECC (2007b): *Glosario*. En línea, <https://www.aecc.es/es/todo-sobre-cancer/glosario> [Consulta: 22/10/2020].

Ballano, Antonio (1807): *Diccionario de medicina y cirugía ó Biblioteca manual médico quirúrgica*, Madrid, Imprenta de la calle de la Greda. En línea, <https://www.bvfe.es/es/directorio-bibliografico-diccionarios-vocabularios-glosarios-tratados-y-obras-lexicografia/15265-diccionario-de-medicina-y-cirugia-o-biblioteca-manual-medico-quirurgica.html> [Consulta: 22/10/2020].

Carmona, Francisco (2019): *Diccionario de Ginecología*. En línea, <https://www.drfcarmona.com/glosario-de-ginecologia/> [Consulta: 22/10/2020].

Casasín Edo, Tomás (2020): *Terminología Farmacéutico-médica*. En línea, <https://www.sefh.es/bibliotecavirtual/auxiliares/area3.pdf> [Consulta: 22/10/2020].

Cotterill, Simon (2000): *Medical Terminology and Cancer. Abbreviations and Acronyms for Oncology*. En línea, <https://www.staff.ncl.ac.uk/s.j.cotterill/medtm15a.htm> [Consulta: 22/10/2020].

Cotterill, Simon (2002): *CancerIndex. Cancer glossary*. En línea, <http://www.cancerindex.org/glossary.htm> [Consulta: 22/10/2020].

García Barreno, Pedro (dir.); Aller Reyero, María Angeles (coord.) (2016): *Diccionario Quirúrgico Ilustrado Complutense*, Madrid, Universidad Complutense de Madrid. En línea, <https://www.ucm.es/cirugia/diccionario-quirurgico-ilustrado> [Consulta: 22/10/2020].

Hurtado de Mendoza, Manuel (1840): *Vocabulario médico-quirúrgico o diccionario de medicina y cirugía, que comprende la etimología y definición de todos los términos usados en estas dos ciencias por los autores antiguos y modernos*, Madrid, Boix. En línea, <http://bdh-rd.bne.es/viewer.vm?id=0000099758&page=1> [Consulta: 22/10/2020].

Moore, Michael (1995): *An Herbal/Medical Dictionary. A Glossary of terms used in Herbalism, Medicine and Physiology. Descriptions, Explanations and Implications in Wholistic and Vitalist Therapy, Bisbee, Southwest School of Botanical Medicine*. En línea, <https://www.swsbm.com/ManualsMM/MedHerbGloss2.pdf> [Consulta: 22/10/2020].

Navarro, Fernando (2021): *Libro rojo. Diccionario de dudas y dificultades de traducción del inglés médico*, 3.ª ed. En línea, <https://www.cosnautas.com/es/catalogo/diccionario-medico-librorojo> [Consulta: 22/10/2020].

Pelikan, Edward (2004): *Glossary of Terms and Symbols Used in Pharmacology*, Boston University. En línea, <https://www.bumc.bu.edu/busm-pm/academics/resources/glossary/> [Consulta: 22/10/2020].

Pera, Cristóbal (1998): *El cuerpo herido. Un diccionario filosófico de la cirugía*, Barcelona, Edicions Universitat de Barcelona (*Estudi General*, n. 3).

Rojas Fraguela, Yadira (2013): *Glosario etimológico de las plantas medicinales con uso antiasmático registradas en Cuba*, Universidad Central "Marta Abreu" de las Villas, Trabajo de diploma. En línea, <https://dspace.uclv.edu.cu/bitstream/handle/123456789/504/Glosario%20etimol%C3%B3gico%20de%20las%20plantas%20....pdf?sequence=1&isAllowed=y> [Consulta: 22/10/2020].

Van der Stichele, Robert (2020): *Multilingual Glossary of technical and popular medical terms in nine European Languages*. En línea, <https://users.ugent.be/~rvdstich/eugloss/welcome.html> [Consulta: 22/10/2020].

Vanaclocha i Vanaclocha, Bernat; Cañigueral Folcarà, Salvador (2020): *Plantas medicinales*. En línea, <https://www.fitoterapia.net/vademecum/plantas/index.html?opt=abc&letra=A> [Consulta: 22/10/2020].

Vázquez, Miguel (2021): *Traumatopedia. Diccionario traumatología*. En línea, <https://traumatopedia.com/diccionario-traumatologia/> [Consulta: 22/10/2020].

WHO Collaborating Center for Pharmaceutical Pricing and Reimbursement Information (2012): *Glosario de términos farmacéuticos/Glossary of pharmaceutical terms*. En línea, <https://ppri.goeg.at/sites/ppri.goeg.at/files/inline-files/Spanish%20Glossary%20final.pdf> [Consulta: 22/10/2020].

El corpus paralelo especializado como herramienta de extracción terminológica: ejemplo de uso y aplicación en la Traducción e Interpretación en los Servicios Públicos (TISP)

Rebeca Busto Acedo
Universidad de Alcalá

María del Mar Sánchez Ramos
Universidad de Alcalá

1. Introducción

En los últimos años, España ha sido país receptor de (in)migrantes, situación que desencadena una gran variedad de escenarios comunicativos y barreras tanto lingüísticas como culturales (Lázaro Gutiérrez 2010; Macipe Costa y Gimeno Feliu 2011). Bajo este contexto, sucede una nueva realidad multilingüe y multicultural con importantes efectos económicos y sociales (Ortega Herráez 2013: 10). Es aquí donde la figura del traductor e intérprete es esencial y necesaria en múltiples contextos, como los servicios públicos, donde estos profesionales deben actuar de puente entre los proveedores de servicios y los usuarios. Con el fin de reducir estas barreras lingüísticas y culturales surge la llamada Traducción e Interpretación en los Servicios Públicos (TISP). La TISP se conforma como ese «puente» necesario para hacer que los servicios administrativos, colegios, hospitales, organizaciones, etc. sean accesibles a los ciudadanos sin importar su país de origen o su lengua materna. Por tanto, podría afirmarse que la TISP permite que estos ciudadanos puedan ejercer sus derechos y además sean conscientes de sus deberes en el país de llegada.

Uno de los pilares claves en torno a la evolución y profesionalización de la TISP es la formación de traductores e intérpretes en los servicios públicos. Es por ello por lo que el número de programas formativos ha ido creciendo en los últimos años (Valero Garcés 2019). En general, un traductor y/o intérprete debe poseer un amplio bagaje temático y

lingüístico, así como textual y cultural. La expansión de internet y la era tecnológica en la que vivimos han hecho posible que, junto con las herramientas electrónicas más tradicionales de documentación (diccionarios especializados, glosarios y bases de datos en línea), el uso de memorias de traducción, gestores de corpus y concordancias, así como herramientas de extracción terminológica formen parte del paisaje documental en las tareas de traducción e interpretación.

El presente capítulo se propone como principal objetivo ofrecer una metodología de corpus para la confección de recursos documentales que sirvan de ayuda en la fase inicial de traducción[1] de textos sanitarios. Se enmarca en la propuesta formativa de posgrado de la Universidad de Alcalá (Madrid, España), el Máster Universitario en Comunicación Intercultural, Traducción e Interpretación en los Servicios Públicos. Si bien se ejemplifica dicha metodología con un tema sanitario, nuestra intención es que la propuesta pueda adaptarse y ser útil en cualquier ámbito especializado. El marco teórico se recoge en la sección 2, en donde se contextualiza la propuesta dentro de las necesidades documentales en la TISP y se describen los principales conceptos relacionados con la lingüística de corpus. Por último, la sección 3 detalla el proceso de compilación de un corpus paralelo bilingüe (inglés-español) y la extracción terminológica para la elaboración de un glosario especializado como recurso documental previo a las tareas de traducción.

2. Documentación y corpus especializados para la TISP

Distintos trabajos han puesto de manifiesto la importancia documental para el traductor e intérprete especializado (Gallego Hernández y Tolosa Igualada 2012, Seghiri 2017). El desarrollo de los procesos documentales dentro del ámbito de la traducción y la interpretación va de la mano del uso generalizado (y consolidado) de la tecnología, que

1. Nuestro trabajo se centra en la fase inicial de un encargo de traducción. No obstante, la compilación de corpus especializados ha sido objeto de estudios como recurso de preparación documental en las tareas de interpretación (Gallego Hernández y Tolosa Igualada 2012).

permite una documentación rápida y eficaz siempre y cuando el traductor e intérprete cuente con unos conocimientos adecuados sobre búsquedas eficientes en internet (Sánchez Ramos y McCarroll 2019). En el caso de la TISP, uno de los ámbitos de actuación es el sanitario, en donde se intentan superar las posibles barreras que puedan establecerse entre el personal sanitario y los usuarios con características lingüísticas y culturales diversas. En este sentido, la traducción sanitaria adquiere un componente social debido a las carencias comunicativas que pueden darse entre profesionales de los servicios públicos y usuarios que no dominan con fluidez la lengua oficial del país de destino, así como de la variedad textual, que va desde artículos científicos a folletos informativos en un idioma extranjero y que están destinados a población migrante. Como indican trabajos recientes (Salgado Mariño 2013; Sánchez Ramos 2020), la documentación en la TISP presenta unas características concretas. Como hemos adelantado en la introducción, la TISP surge de la necesidad comunicativa que se establece entre los trabajadores de los servicios o instituciones públicas y los usuarios que no dominan la lengua oficial con fluidez. Así, una de las principales funciones de la TISP es proporcionar ayuda a este grupo de minorías lingüísticas y culturales que conforma nuestra sociedad. Y para ello la TISP necesita profesionales con una formación que aúne las destrezas necesarias para hacer frente a estas dificultades comunicativas. Los traductores e intérpretes en estos contextos deben realizar un proceso de documentación rápido y puede que en ocasiones la temática exija también una documentación precisa que requiera la elaboración de un material documental concreto. En palabras de Sánchez Ramos (2017), de una buena documentación dependerá que el traductor e intérprete en estos contextos adquiera los conocimientos temáticos, terminológicos, fraseológicos y textuales necesarios para llevar a cabo el proceso de trasvase. En el caso de la traducción sanitaria en la TISP, los vacíos documentales suelen centrarse en cuestiones temáticas (comprensión del texto) y terminológicas (comprensión de vocabulario especializado).

Dentro de los recursos documentales que pueden ayudar al traductor e intérprete en la TISP, los corpus y gestores de concordancias suelen ser unas herramientas poco utilizadas en este entorno (Symseridou 2018; Sánchez Ramos y Lázaro Gutiérrez 2020). Ambos trabajos inciden en la importancia de estas herramientas para así solventar las carencias lingüísticas que se dan en la TISP. En el primer trabajo, Symseridou (2018) expone el proceso de compilación de un corpus comparable (inglés-griego) especializado con la ayuda de WebBootCat. Por otro lado, el estudio llevado a cabo por Sánchez Ramos y Lázaro Gutiérrez (2020) tiene lugar dentro del aula de traducción sanitaria en la TISP y expone el uso de un corpus comparable (inglés-español) como herramienta de documentación para la traducción de textos médico-jurídicos (consentimientos informados). Además de la propuesta didáctica, resultan llamativos los resultados del cuestionario realizado a los sujetos participantes en el estudio, y donde se pone de manifiesto el desconocimiento inicial de estas herramientas y la utilidad de estas para el desarrollo y adquisición de destrezas tecnológicas y documentales.

Los trabajos anteriores son algunos ejemplos de estudios que exponen la descripción y el uso de corpus comparables en la TISP. No es nuestro objetivo extendernos en una descripción detallada sobre la tipología de corpus,[2] pero sí creemos importante detenernos en el corpus paralelo *ad hoc*, objeto de este estudio y que nos servirá para ejemplificar su explotación para la elaboración de recursos documentales, como la elaboración de glosarios especializados. El corpus paralelo está formado por originales en una lengua y sus traducciones en una o más lenguas (McEnery y Wilson 1996: 348). A esta definición podemos añadir la aportación de Frankenberg-García (2009), quien afirma que pueden ser unidireccionales o bidireccionales, o bien una combinación de ambos. Los corpus paralelos unidireccionales están formados por textos en la lengua de partida y sus traducciones en la

2. Remitimos a la obra de Zanettin (2012) para una consulta detallada de la tipología de corpus en traducción.

lengua de llegada. En palabras de FrankenbergGarcía (2009: 57): «a bidirectional corpus contains source texts in two different languages (L1 and L2) aligned with their reciprocal translations into L2 and L1». En líneas generales, el corpus paralelo será de gran utilidad para estudiar, por ejemplo, las distintas estrategias de traducción, la traducción de la fraseología empleada, o la naturaleza de la lengua traducida, entre otros. Cada día está adquiriendo una mayor popularidad la compilación de corpus de dimensiones más reducidas (monolingües, comparables o paralelos) como fuente de documentación en tareas de traducción concretas o bien como recurso pedagógico, y cuyo material no es otro que textos electrónicos extraídos de internet y que han seguido una evaluación de dichos textos y un criterio o protocolo de compilación concreto (Seghiri 2011). Se trata de los llamados corpus *ad hoc* o virtuales. Este tipo de corpus es considerablemente útil como herramienta de documentación en temas especializados, donde los recursos pueden llegar a ser escasos, como es el caso de la TISP.

3. Compilación de un corpus paralelo especializado: ejemplificación de uso y extracción terminológica

En las secciones anteriores hemos reflejado la utilidad que la compilación de un corpus puede aportar como recurso documental previo a la traducción. Sí hay que matizar que, como señalan diversos autores (Seghiri 2011, Sánchez Ramos y Vigier Moreno 2017), se precisa de un protocolo de compilación que dirija el diseño del corpus. En esta sección describiremos en detalle dicho protocolo, así como las herramientas utilizadas para su compilación de un corpus paralelo bilingüe unidireccional formado por artículos en español y sus traducciones en inglés y la extracción terminológica posterior. De este modo ofreceremos además una metodología que pueda ser provechosa a la hora de enfrentarse a la fase de preparación documental en la traducción sanitaria en los servicios públicos. Para ello, se han seguido las propuestas de Seghiri (2011), Vigier Moreno (2016) y Sánchez Ramos (2019).

El corpus aquí creado podría clasificarse según la taxonomía propuesta anteriormente por Frankenberg-García (2009), por lo que sería un corpus paralelo unidireccional (español-inglés) dado que está formado por el texto origen en español y sus correspondientes textos meta en inglés. Otro aspecto a tener en cuenta es el tamaño del corpus en cuestión. Según Bowker y Pearson (2002: 54) el tamaño de un corpus debe rondar los 35.000-45.000 *tokens* (palabras) en cada lengua. Para los fines del presente trabajo se ha considerado que el número de *tokens* propuesto por Bowker y Pearson es suficiente y por ello se ha conseguido un corpus formado por 38.355 tokens en español y 34.799 tokens en inglés.

4.1. Elección del tema

Dado que este trabajo se enmarca en la traducción en los servicios públicos se ha estimado oportuno elegir una temática ligada a esta disciplina, en concreto al ámbito médico-sanitario. Más específicamente, se decidió que el corpus objeto del presente trabajo abordaría la temática de las cardiopatías congénitas. Es importante aclarar que una cardiopatía congénita (CC) es un problema relacionado con la estructura y el funcionamiento del corazón que algunos individuos presentan ya en el momento de su nacimiento.[3] Este tipo de anomalía congénita es la más común y puede describir distintos problemas que afectan al corazón. La Dr.ª Laura Dos Subirá, coordinadora del Registro Español de Cardiopatías Congénitas, señala que los resultados de una encuesta realizada en el año 2014 muestran cómo más de 20.000 pacientes adultos con CC estaban siendo tratados en 24 Unidades de Cardiopatías Congénitas del Adulto (Sociedad Española de Cardiología 2019). Además, afirma que 3 de cada 1000 habitantes en España tiene alguna CC y que, extrapolando estos datos a la población en edad adulta en este país, se estima que habría unos 115.000 pacientes en total (*ibid*.). Además, Somos Pacientes[4] (2020), en el Día Inter-

3. MedlinePlus: https://medlineplus.gov/spanish/ency/article/001114.htm.
4. Comunidad de Asociaciones de Pacientes en España.

nacional de las Cardiopatías Congénitas de 2020, publicó un artículo que aseguraba que cada día nacen en España 10 niños con alguna CC. Esto supone, aproximadamente, un total 4000 niños cada año. Teniendo estos datos en mente y considerando que el fin de un corpus *ad hoc* es que sea útil, se ha creído pertinente centrar el trabajo en la temática de las cardiopatías congénitas puesto que es probable que, a lo largo de su trayectoria profesional, un traductor en el ámbito médico-sanitario de los servicios públicos tenga que tratar este tema y, por tanto, tenga que estar familiarizado con la terminología que rodea a esta anomalía.

Una vez decidida la temática del corpus, se han buscado en internet glosarios especializados ya existentes sobre la misma. Tras esta búsqueda se ha podido confirmar que no existen glosarios específicos sobre esta temática o que al menos no son de acceso público. Sí se han encontrado otros recursos como el *Glosario de términos y abreviaturas en cardiología: Inglés-Español* (Alegría Ezquerra y Teresa Galván 2005), que ha sido el resultado de la colaboración de las Sociedades Cardiológicas hispanoamericanas, expertos lingüistas y especialistas y con el aval de la Real Academia Española. Desgraciadamente, este y otros recursos son de pago y, por ende, no son de acceso abierto y gratuito para todo aquel que quiera hacer uso de dicho glosario.

4.2. Criterios del diseño

Antes de comenzar el proceso de compilación es esencial establecer una serie de criterios de diseño claros en lo que respecta a las fuentes documentales y a los textos que van a conformar el corpus de trabajo. Con el objetivo de crear un corpus formado por bitextos[5] se ha considerado fundamental un género en el que abunden textos originales y sus traducciones. En un primer momento se determinó que el corpus estuviera compuesto tan solo por resúmenes de artí-

5. Harris (1988: 8, en Seghiri 2017: 125) considera que un TO y su traducción no son dos textos separados, sino que forman un texto único en dos dimensiones, cada una de las cuales es un lenguaje.

culos de investigación. No obstante, para conseguir una muestra cuantitativa suficiente finalmente se decidió ampliar el contenido del corpus y no limitarse solo a los resúmenes. Esto permite que dicho corpus sea homogéneo en cuanto a su contenido y el medio es escrito. Se han excluido las notas al pie de página y descripciones de imágenes, pero sí se han incluido las abreviaturas presentes en el texto. Los resúmenes y artículos de investigación que se encuentran en varios idiomas suelen ser sometidos a un proceso de revisión (Seghiri 2017: 125), por lo que la fiabilidad de la información presente en ellos se ve incrementada. Se ha decidido determinar una antigüedad máxima de fecha de publicación de 10 años (desde 2010 hasta 2020), dado que este periodo permite obtener un número considerable de muestras que no estén desfasadas con respecto a los avances médicos. Los textos seleccionados han sido redactados por profesionales procedentes de España o que trabajan en este país con el fin de delimitar el corpus desde una perspectiva diatópica. Cabe destacar el hecho de que, si bien podían haberse elegido textos de carácter divulgativo, se ha considerado que los textos especializados podrían ser más adecuados para los fines de este trabajo. Por un lado, los artículos de investigación normalmente contienen un mayor número de términos especializados, permitiendo así obtener una muestra de términos más amplia. Por otro, aunque los traductores e intérpretes de la TISP deban actuar de puente entre los proveedores de servicios y los usuarios, el discurso por parte del proveedor del servicio, al fin y al cabo, contiene un alto grado de especificidad.

4.3. Protocolo de compilación: localización, alineación y extracción terminológica

En lo referido a la localización de los textos que conforman nuestro corpus, y con el fin de asegurar la representatividad de la muestra, el proceso de localización de los bitextos se ha restringido a los repositorios de la Revista Española de Cardiología (REC) y ScienceDirect,

puesto que se han considerado fuentes fiables y adecuadas para los fines de este trabajo. El portal de la REC ha permitido compilar un total de 18 bitextos con los criterios de diseño especificados, mientras que en ScienceDirect se han conseguido 27 bitextos. Como resultado se ha obtenido un corpus compuesto por 90 textos, 45 en español y sus 45 traducciones al inglés. Cabe mencionar el hecho de que en 10 de los 45 textos originales (con sus correspondientes 10 traducciones al inglés) se ha utilizado también el cuerpo del artículo, como se ha mencionado anteriormente, para conseguir una muestra cuantitativa suficiente. La selección de estos 10 textos se debe en su mayoría a la disponibilidad de la totalidad del artículo, puesto que en muchos de ellos solo se podía acceder de manera gratuita al resumen del artículo.

Tras la localización de los textos se ha proseguido con su descarga para poder acceder a ellos sin necesidad de conexión a internet. 17 de los documentos obtenidos de la REC se han descargado en formato PDF y 1 en formato HTML. De los 27 resultados obtenidos través del buscador de ScienceDirect, 18 de ellos se han descargado en formato PDF y 9 en formato HTML. Seguidamente, los textos se han convertido a texto plano (.txt). Tras la conversión, se dio paso a la alineación de los textos, que permite que los bitextos puedan ser procesados por los programas de gestión de corpus y además posibilita que se puedan determinar correspondencias entre los distintos niveles de segmentación (párrafo, oración y palabra). La segmentación oracional es la más frecuente (Sánchez Ramos 2019: 278) y por ello se ha decidido emplearla en este trabajo. Hoy en día existen numerosos programas informáticos, ya sean libres o comerciales, que permiten realizar esta tarea de forma automática. El programa empleado en este trabajo ha sido LF Aligner, de acceso gratuito, que facilita la creación de memorias de traducción y corpus textuales gracias a un algoritmo de código abierto de alineación de oraciones (Aligner.sf.net 2011). Además, tras esta alineación automática este programa ofrece la opción de la revisión manual mediante un editor gráfico. No obstante, existen progra-

mas como Heartsome TMX Editor que permiten editar la alineación que ha sido exportada en formato TMX. Tras haber revisado la alineación de LF Aligner también se ha optado por la revisión de archivo TMX a través de Heartsome TMX Editor, dado que ha resultado más eficaz en algunos casos y además ofrece un mayor número de opciones de edición. Gracias a la revisión y edición manual de la alineación automática se han evitado posibles fallos como, por ejemplo, la división o fusión errónea de algunas oraciones o la omisión de información.

Para realizar la extracción de los términos automáticamente se ha recurrido al programa SynchroTerm, herramienta que acelera el proceso de extracción terminológica y crea registros o entradas individuales para cada término. SynchroTerm se basa en algoritmos estadísticos y de análisis sintáctico y morfológico para la extracción automática de los términos origen y su equivalente en la lengua meta (Terminotrix s. f.). Además, posibilita la creación de las bases de datos terminológicas e incluso comparar la terminología presente en los documentos con el objetivo de mantener la coherencia terminológica.

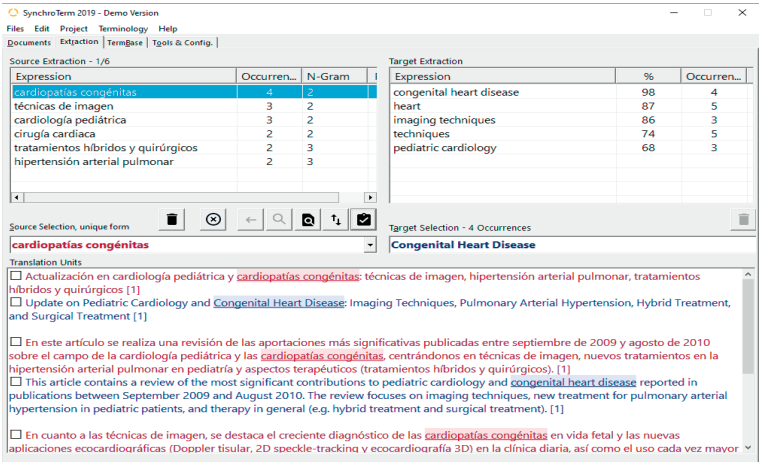


Figura 1. Extracción terminológica automática en SynchroTerm.

Como puede apreciarse en la Figura 1, SynchroTerm presenta una lista de candidatos y facilita tanto el número de ocasiones en las que aparece como su equivalente. En caso de que la extracción automática no contenga la totalidad de los términos presentes en los documentos también pueden extraerse términos que no figuran en la lista. Es recomendable revisar todos los términos individualmente para verificar su relevancia y configurar la lista en función de los parámetros deseados. En los términos extraídos se ha seguido un proceso de estandarización específico que ha consistido en unificar todos los términos en singular y, en caso de que hubiera verbos, decidir el tiempo verbal en el que van a guardarse.

SynchroTerm nos ha permitido crear una lista de términos en español y sus equivalentes en inglés sobre cardiopatías congénitas de manera automática gracias a un corpus paralelo bilingüe compilado previamente siguiendo un protocolo concreto. No obstante, se han añadido también términos de manera manual dado que SynchroTerm no extraía automáticamente todos los términos del texto. Gracias a este proceso de extracción automática proporcionada por esta herramienta terminológica, así como la posterior revisión manual, se han obtenido un total de 505 términos en español relacionados con las cardiopatías congénitas y sus equivalentes en inglés (1010 términos), cuya muestra queda recogida en el anexo a este trabajo.

4.4. Análisis de los resultados obtenidos

Gracias a la extracción realizada por SynchroTerm y a la herramienta de gestión de corpus AntConc se ha podido extraer una lista de los términos más frecuentes del corpus, así como de sus términos o palabras asociadas que hacen que formen un nuevo término.

TÉRMINO	FREC.	PALABRAS ASOCIADAS	FREC.
cardiopatía(s)	287	cardiopatía(s) congénita(s)	236
		cardiopatías adquiridas	4
		adultos con cardiopatía congénita	10
		cardiopatías congénitas complejas	4
CC	100	pacientes con CC	12
		CC complejas	4
		CC cianótica	4
cirugía(s)	123	cirugías con circulación extracorpórea	3
		cirugías con circulación extracorpórea	3
		cirugía cardiaca	23
		cirugía cardiaca en CEC	6
		cirugía correctora	11
diagnóstico(s)	128	diagnóstico(s) no esperado(s)	13
		diagnóstico precoz	4
pulmonar(es)	118	arteria(s) pulmonar(es)	21
		hipertensión pulmonar	13
		insuficiencia pulmonar	7
		estenosis pulmonar	4
		válvula pulmonar	23
insuficiencia	78	insuficiencia cardiaca	55
		insuficiencia pulmonar	7
		insuficiencia aórtica	4
ventrículo	74	ventrículo derecho	6
		ventrículo izquierdo	6
		ventrículo subpulmonar	4
		ventrículo sistémico	2
ventricular(es)	68	asistencia ventricular	19
		disfunción ventricular	13
		disfunción ventricular derecha	6
		asistencia ventricular mecánica	7
aurícula	12	aurícula derecha	10
		aurícula izquierda	2
auricular(es)	36	arritmias auriculares	10
		switch auricular	7
trasplante(s)	47	trasplante(s) cardiaco(s)	22
		trasplante pulmonar	3

válvula(s)	44	válvula(s) displástica(s)	2
		válvula tricúspide	3
		válvula mitral	4
		válvula pulmonar	23
sistémico	39	ventrículo derecho/VD sistémico	30
disfunción	36	disfunción ventricular	13
		disfunción sistólica	4
técnica(s)	67	técnicas de imagen	9
		técnica(s) diagnóstica(s)	6
estenosis	34	estenosis aórtica	5
		estenosis pulmonar	6
		estenosis valvular	3
biomarcador(es)	37	biomarcador(es) cardiaco(s)	18
ecocardiografía	35	ecocardiografía fetal	4
		ecocardiografía transtorácica	4
		ecocardiografía tridimensional	2
arritmia(s)	33	arritmias auriculares	10
		arritmias ventriculares	2
aórtica(s)	33	implante percutáneo de prótesis aórticas	2
		enfermedad aórtica	5
		estenosis aórtica	5
		insuficiencia aórtica	4
		válvula aórtica	4
cardiovascular(es)	33	enfermedad cardiovascular	10
		cirugía cardiovascular	6
circulación	26	circulación extracorpórea	19
		circulación sistémica	3
obstrucción	27	obstrucción del TSVD	8
síndrome	25	síndrome de bajo gasto cardiaco	6
		síndrome del corazón izquierdo hipoplásico	5
anomalía	39	anomalías congénitas	5
		anomalía de Ebstein	18
cateterismo	23	cateterismo diagnóstico	4
		cateterismo terapéutico	3
Fallot	20	tetralogía de Fallot	17
Fontan	15	cirugía de Fontan	4
		circulación tipo Fontan	3
corazón	19	corazón izquierdo hipoplásico	3

hipertensión	19	hipertensión arterial pulmonar	3
		hipertensión pulmonar	13
Ebstein	18	anomalía de Ebstein	18
tomografía	18	tomografía computarizada	14
tetralogía	17	tetralogía de Fallot	17
procedimiento(s)	21	procedimiento(s) quirúrgico(s)	2
		procedimiento(s) terapéutico(s)	3
malformación(es)	17	malformaciones cardiacas	3
		malformaciones congénitas	4
		malformaciones congénitas mayores	2
		malformaciones congénitas cardiacas	2
		malformaciones congénitas	6
		malformación estructural	2
oxígeno	16	saturación de oxígeno	3
		consumo de oxígeno	9
soplo	12	soplo cardiaco	2
		soplo sistólico	1
		soplo holosistólico	1
		soplo pansistólico	1

Tabla 1. Términos más frecuentes.

Como puede apreciarse en la Tabla 1 y como podía ser predecible desde un principio, el término más frecuente es *cardiopatía congénita*, seguido de *diagnóstico* y *cirugía*. También se ha percibido que algunos términos tienen más de un equivalente o que un mismo equivalente en inglés puede referirse a dos términos distintos en español. Ante esta situación se ha creído conveniente crear una entrada diferenciada para cada uno de los términos para que en caso de cambiar la direccionalidad del glosario a inglés>español aparezcan ambos términos en su orden alfabético correspondiente (el cambio de direccionalidad del glosario puede realizarse simplemente cambiando el orden de las columnas y ordenando los términos de nuevo). Estos términos aparecen en la Tabla 2. A pesar de esta diferencia en cuanto al equivalente o término en la lengua origen, para ver cuál de los términos es más adecuado o frecuente es suficiente con buscar su frecuencia o número de

apariciones dentro del corpus de trabajo a través de herramientas de gestión de corpus como AntConc.

	ES	EN
16	anomalía congénita	congenital heart defect
17	anomalía congénita	congenital anomaly
56	cardiopatía congénita	congenital heart disease
57	cardiopatía congénita	congenital cardiac malformation
133	ecocardiografía fetal	prenatal ultrasound screening
134	ecocardiografía fetal	fetal echocardiography
281	presión arterial	arterial pressure
282	presión arterial	blood pressure
283	presión arterial sistólica	systolic pressure
284	presión sistólica	systolic pressure
373	trasplante cardiaco	cardiac transplantation
374	trasplante cardiaco	heart transplantation

Tabla 2. Pluralidad de términos origen o meta.

A partir de los términos que conforman el glosario, es decir, un total de 505 términos en español y sus 505 equivalentes en inglés (1010 en total), también puede observarse algo muy interesante en lo que respecta a la composición y estructura de los términos (n-gramas). De esos 505 términos, 85 de ellos están formados por solo una palabra, 267 por dos palabras y 153 están formados por tres o más palabras (para hacer este recuento no se han contado las preposiciones y artículos; por ejemplo, *anomalía de Ebstein* se ha contado como dos palabras). Esto supone que más de la mitad de los términos están formados por dos palabras, normalmente un nombre y un adjetivo o complemento del nombre (52,87 %). La segunda construcción más frecuente es la del término formado por tres o más palabras (30,3 %), seguido en último lugar, por el término que es una única palabra (16,83 %). No obstante, cabe destacar que esos términos formados por más de una

palabra (420 términos, 83,17 % del total) son en su mayoría variaciones de aquellos términos formados por una sola palabra (85 términos, 16,83 %). Asimismo, dentro de esos 85 términos, 78 actúan como sustantivos (91,76 %) y los 7 restantes actúan como adjetivos (8,24 %). A partir de estos datos puede deducirse que el lenguaje especializado está formado en su mayoría por sustantivos y que muchas de las palabras sufren un proceso de nominalización.

También es interesante destacar algunos elementos morfológicos y sintácticos del glosario. Desde el punto de vista semántico, el hiperónimo más frecuente en este glosario es *enfermedad*. Algunos de los hipónimos de este hiperónimo, es decir, términos de significado restringido están contenidos dentro de otro de significado más amplio (el hiperónimo), incluyen *enfermedad coronaria*, *enfermedad valvular aórtica* o *miocardiopatía*. A su vez, *enfermedad coronaria*, *enfermedad valvular aórtica* y *miocardiopatía* son cohipónimos entre ellos, dado que son términos de significado restringido que comparten hiperónimo. También puede observarse la presencia de ciertos campos semánticos, que son un conjunto de palabras de la misma categoría gramatical que comparten algún sema. Por ejemplo, bajo el sema *pruebas diagnósticas* pueden encontrarse *angiografía*, *ecocardiografía*, *ecografía*, *electrocardiografía*, *ergometría*, *tomografía* y *pulsioximetría*. Bajo el sema *enfermedades* pueden incluirse *arteriopatía*, *cardiopatía*, *cromosomopatía* y *miocardiopatía*. Como puede apreciarse, las palabras de este campo semántico han sufrido un proceso de derivación mediante el sufijo -patía, que proviene del latín *pathīa* y del griego *πάθεια* (pátheia) en concreto de la raíz *παθ* (path) que significa 'sufrir', 'experimentar'. Así, este elemento compositivo denota, en este caso, 'afección' o 'dolencia'. El uso de términos derivados del latín o griego y unidades morfológicas latinas o griegas es una de las características del lenguaje médico-sanitario (Dutova 2013: 18; Ruiz Rosendo 2007: 175). Por ende, el conocimiento de las unidades morfológicas latinas y griegas puede ser una ventaja para el traductor e intérprete en este ámbito, puesto que, aunque no conozca el

equivalente exacto del término en la otra lengua de trabajo, si sabe su composición puede explicar el término con otras palabras. Desde el punto de vista morfológico, también pueden encontrarse familias léxicas, es decir, palabras que comparten el mismo lexema. Un ejemplo presente en este corpus podrían ser los términos *válvula*, *valvulopatía* y *valvuloplastia*.

5. Conclusiones

Es evidente que la documentación previa en cualquier ámbito y especialidad de la traducción e interpretación profesional en general, y de la TISP en particular, es sumamente importante, puesto que de ella dependerá la calidad del servicio prestado. Además, la TISP lleva consigo una especificidad temática y terminológica con la que los traductores no siempre están familiarizados, y es por ello por lo que deben poseer una serie de habilidades y recursos con los que hallar información sobre cualquier temática. En este sentido, los corpus paralelos son particularmente útiles para satisfacer las necesidades de documentación. Entre los diversos beneficios se encuentran, por ejemplo, su reutilización en futuros encargos, así como el hecho de que son fáciles de usar y que permiten gestionar grandes cantidades de información en cuestión de segundos. Asimismo, un corpus que sea representativo y que haya sido gestionado de manera adecuada es un instrumento muy eficaz para la identificación, extracción y traducción de unidades léxicas, permitiendo así la creación de materiales como pueden ser las bases de datos, tesauros, glosarios o memorias de traducción. Por tanto, la metodología aquí propuesta puede usarse como modelo para la creación de materiales especializados e incluso añadir nueva información, por ejemplo, en el caso de la interpretación, la inclusión en un glosario de la transcripción fonética de los términos.

También se ha reafirmado el hecho de que el diseño de un corpus requiere un protocolo de compilación que guíe la investigación y que permita obtener los resultados deseados. Además, los textos no pue-

den seleccionarse de manera aleatoria y deben tenerse en cuenta las necesidades y objetivos del corpus. En este sentido, este trabajo ha señalado la importancia de establecer unos criterios a la hora de conformar el corpus de trabajo, los cuales también deben seguirse en la fase de localización para no contaminar la muestra y asegurar que la procedencia de los documentos es fiable y representativa. De igual modo, la conversión del formato y la alineación han sido etapas esenciales en el proceso. Sin embargo, los programas que ejecutan dichas acciones automáticamente no realizaban las conversiones y alineaciones de manera satisfactoria y en algunos casos la mejor solución ha sido hacerlo manualmente.

En lo que respecta a la extracción automática de los términos, a través del presente trabajo se han podido observar algunas desventajas de las herramientas de extracción. El mayor inconveniente de estas herramientas es que en muchas ocasiones seleccionan numerosas palabras que no son términos especializados, incluso segmentos oracionales muy extensos, y excluyen términos que sí deberían haberse escogido. Por esta razón, la revisión de los términos propuestos y la posterior selección manual de aquellos términos que no se han incluido en la lista automáticamente es fundamental. A partir de esta apreciación puede concluirse que la extracción terminológica automática puede ser una buena opción para tener una primera lista de posibles términos, pero que siempre es recomendable la revisión manual para asegurar la calidad y representatividad de la muestra.

Como ya se ha mencionado, la metodología aquí propuesta puede servir como modelo para la creación de futuros materiales especializados en esta u otra combinación lingüística. Se ha buscado ofrecer unos criterios a la hora de conformar el corpus de trabajo y aportar una metodología para desarrollar dicho protocolo. Si bien es cierto que al principio el proceso puede resultar arduo, se considera que en futuras elaboraciones el tiempo invertido será menor como consecuencia de la familiarización con las etapas del proceso, reduciendo el tiempo total dedicado a la fase de documentación y permitiendo invertir más

tiempo en otras etapas. A partir de todas estas observaciones también podría concluirse que el uso de la tecnología resulta provechoso en lo que respecta a la documentación y formación del traductor e intérprete en los servicios públicos, puesto que permite obtener términos y equivalentes fiables y adecuados de manera rápida y sencilla a través de los corpus.

Referencias bibliográficas

Alegría Ezquerra, Eduardo; Teresa Galván, Eduardo de (2005): *Glosario de términos y abreviaturas en cardiología: Inglés-Español*, Madrid, Acción Médica.

Aligner.sf.net. En línea, <https://lf-aligner.soft112.com/> [Consulta: 10/04/2020].

Bowker, Lynne; Pearson, Jennifer (2002): *Working with Specialized Language: A Practical Guide to Using Corpora*, London, Routledge.

Dutova, Elana (2013): *Proyecto de glosario bilingüe español-ruso de los términos de oncología destinado al* traductor, Universidad de Alcalá, Trabajo de Fin de Máster dirigido por Anna Maevskaya.

Frankenberg-Garcia, Ana (2009): «Compiling and Using a Parallel Corpus for Research in Translation», *Babel*, vol. 21, n. 1, pp. 57-71.

Gallego Hernández, Daniel; Tolosa Igualada, Miguel (2012): «Terminología bilingüe y documentación *ad hoc* para intérpretes de conferencias. Una aproximación metodológica basada en corpus», *Estudios de Traducción*, vol. 2, pp. 33-46.

Harris, Brian (1988): «Bitext, a New Concept in Translation Theory. Working Papers on Bilingualism», *Language Monthly*, vol. 54, pp. 8-10.

Lázaro Gutiérrez, Raquel (2010): *La interpretación en los servicios sanitarios en España. Estudio de la asimetría en las consultas médicas con paciente de habla extranjera y la repercusión de la presencia de un intérprete ocasional*, Universidad de Alcalá, Tesis doctoral dirigida por M. Carmen Valero Garcés.

Macipe Costa, Rosa María; Gimeno Feliu, Luis Andrés (2011): *Competencia cultural y población inmigrante: un reto cotidiano en nuestras consultas. Pediatría Atención Primaria*, vol. 13 n. 50, pp. 299-319. En línea, <http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1139-76322011000200013> [Consulta: 10/03/2020].

McEnery, Tony; Wilson, Andrew (1996): *Corpus Linguistics. An introduction*, Edinburgh, Edinburgh University Press.

Ortega Herráez, Juan Miguel (2013): «Caracterización de la práctica profesional de la traducción en el orden jurisdiccional penal en España», *Punto y Coma*, vol. 133, pp. 10-24.

Ruiz Rosendo, Lucía (2007): «Propuesta de formación en interpretación en el ámbito de la medicina», *Panace@. Revista de medicina, lenguaje y traducción*, vol. 8, n. 26, pp. 173-182.

Salgado mariño, Jorge Eduardo (2013): *Fuentes documentales electrónicas disponibles para el traductor en los servicios públicos francés-español y español-francés: compendio y evaluación*, Alcalá de Henares, Universidad de Alcalá.

Sánchez Ramos, María del Mar (2017): «Compilación y análisis de un corpus *ad hoc* como herramienta de documentación electrónica en traducción e interpretación en los servicios públicos (TISP)», *Estudios de traducción*, vol. 7, pp. 177-190.

Sánchez Ramos, María del Mar (2019): «Corpus paralelos y traducción especializada: ejemplificación de diseño, compilación y alineación de un corpus paralelo bilingüe (inglés-español) para la traducción jurídica», *Lebende Sprachen,* vol. 64, n. 2, pp. 269-285.

Sánchez Ramos, María del Mar (2020): *Documentación digital y léxico en la traducción e interpretación en los servicios públicos (TISP): fundamentos teóricos y prácticos*, Bern, Peter Lang.

Sánchez Ramos, María del Mar; Lázaro Gutiérrez, Raquel (2020): «Traducción de consentimientos informados y lingüística de corpus: una propuesta metodológica para el aprendizaje de la traducción de textos médico-jurídicos», en Miriam Seghiri (ed.), *La lingüística de corpus aplicada desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas,* Berlin, Peter Lang, pp. 49-67.

Sánchez Ramos, María del Mar; McCarroll, Colleen (2019): «TACLAC: una plataforma colaborativa de intercambio de corpus y su aplicación en la traducción biosanitaria», *Panace@. Revista de medicina, lenguaje y traducción*, vol. xx, n. 49, pp. 73-81.

Sánchez Ramos, María del Mar; Vigier Moreno, Francisco J. (2017): «Using Parallel Corpora to Study the Translation of Legal System-Bound Terms: The Case of Names of English and Spanish Courts», *Computational and Corpus-Based Phraseology*, Springer, Cham.

Seghiri, Miriam (2011): «Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad», *Revista de lingüística teórica y aplicada*, vol. 49, n. 2, pp. 13-30.

Seghiri, Miriam (2017): «Corpus e interpretación biosanitaria: extracción terminológica basada en bitextos del campo de la Neurología para la fase documental del intérprete», *Panace@. Revista de medicina, lenguaje y traducción*, vol. 18, n. 46, pp. 123-132.

Symseridou, Elina (2018): « The Web as a Corpus and for Building corpora in the Teaching of Specialised Translation: The Example of Texts in Healthcare», *FITISPos International Journal*, vol. 5, n. 1, pp. 60-82.

Sociedad Española de Cardiología (2019): «La SEC pone en marcha el primer Registro Español de Cardiopatías Congénitas a nivel nacional». En línea, <https://secardiologia.es/comunicacion/notas-de-prensa/notas-de-prensa-sec/10602-la-sec-pone-en-marcha-el-primer-registro-espanol-de-cardiopatias-congenitas-a-nivel-nacional> [Consulta: 19/03/2020].

Somos Pacientes (2020): «Cada día nacen en España 10 niños con cardiopatías congénitas». En línea, <https://www.somospacientes.com/noticias/sanidad/cada-dia-nacen-en-espana-10-ninos-con-cardiopatias-congenitas/> [Consulta: 19/03/2020].

Terminotrix. En línea, <https://terminotix.com/docs/factsheet_SynchroTerm_en.pdf> [Consulta: 10/04/2020].

Valero Garcés, Carmen (2019): «Intercultural communication. Public service interpreting and translation», en Roberto Valdeón García, África Vidal Claramonte (eds.), *The Routledge Handbook of Spanish Translation Studies*, London, Routledge, pp. 211-226.

Vigier Moreno, Francisco Javier (2016): «Teaching the use of *ad hoc* corpora in the translation of legal texts into the second language», *Language and Law/Linguagem e Direito*, vol. 3, n. 1, pp. 100-113.

Zanettin, Federico (2012): *Translation-driven Corpora*, Manchester, St. Jerome Publishing.

Anexo 1

Muestra del glosario bilingüe (español-inglés) sobre cardiopatías congénitas.

	ES	EN	Pronunciación
1	2D speckle-tracking	two-dimensional speckle-tracking imaging	/tuː dɪˈmɛnʃənl ˈspɛkl ˈtrækɪŋ ˈɪmɪʤɪŋ/
2	ablación con radiofrecuencia	radiofrequency ablation	/ˌreɪdɪəˈfriːkwənsɪ æbˈleɪʃ(ə)n/
3	acoplamiento ventriculoarterial	ventricular-arterial coupling	/vɛnˈtrɪkjʊlər ɑːˈtɪərɪəl ˈkʌplɪŋ/
4	algoritmo terapéutico	therapeutic algorithm	/ˌθerəˈpjuːtɪk ˈælgərɪðm/
5	alteración hemodinámica	emodynamic disturbance	/ˌɪmɒdaɪˈnæmɪk dɪsˈtɜːbəns/
6	aneurisma	aneurysm	/ˈænjʊərɪz(ə)m/
7	aneurisma aórtico	aortic aneurysm	/eɪˈɔːtɪk ˈænjʊərɪz(ə)m/
8	aneurisma del seno de Valsalva	aneurysm of the sinus of Valsalva	/ˈænjʊərɪz(ə)m ɒv ðə ˈsaɪnəs ɒv vəˈlsævə/

9	aneurisma del tabique membranoso	aneurysm of the membranous septum	/ˈænjʊərɪz(ə)m ɒv ðə mɛmˈbreɪnəs ˈsɛptəm/
10	angiografía	angiography	/ˈændʒɪəʊgrəfɪ/
11	angiografía por TC	CT angiography	/siː tiː ˈændʒɪəʊgrəfɪ/
12	angiograma	angiogram	/ˈændʒɪəʊgræm/
13	angiotensina	angiotensin	/ˌændʒɪəʊˈtensɪn/
14	anillo de la válvula tricúspide	tricuspid valve annulus	/traɪˈkʌspɪd vælv ˈænjʊləs/
15	anillo valvular	valvar annulus	/ˈvælvə ˈænjʊləs/
16	anomalía congénita	congenital heart defect	/kənˈdʒɛnɪtl hɑːt dɪˈfɛkt/
17	anomalía congénita	congenital anomaly	/kənˈdʒɛnɪtl əˈnɒməli/
18	anomalía de Ebstein	Ebstein anomaly	/ˈebstɪn əˈnɒməli/
19	antagonistas de los receptores	receptor antagonists	/rɪˈsɛptər ænˈtægənɪsts/
20	ápex	apex	/ˈeɪpɛks/
21	arco aórtico	aortic arch	/eɪˈɔːtɪk ɑːʧ/
22	arco aórtico derecho	right-sided aortic arch	/raɪt ˈsaɪdɪd eɪˈɔːtɪk ɑːʧ/
23	arritmia	arrhythmia	/əˈrɪðmɪə/
24	arritmia auricular	atrial arrhythmia	/ˈeɪtrɪəl əˈrɪðmɪə/
25	arritmia supraventricular	supraventricular arrhythmia	/ˌsuːprəvenˈtrɪkjʊlə əˈrɪðmɪə/
26	arritmia ventricular	ventricular arrhythmia	/vɛnˈtrɪkjʊlər əˈrɪðmɪə/
27	arteria coronaria	coronary artery	/ˈkɒrənəri ˈɑːtəri/
28	arteria coronaria distal	distal coronary artery	/ˈdɪstəl ˈkɒrənəri ˈɑːtəri/
29	arteria coronaria proximal	proximal coronary artery	/ˈprɒksɪm(ə)l ˈkɒrənəri ˈɑːtəri/
29	arteria pulmonar	pulmonary artery	/ˈpʌlmənəri ˈɑːtəri/
31	arteria pulmonar central	central pulmonary artery	/ˈsɛntrəl ˈpʌlmənəri ˈɑːtəri/
32	arteria pulmonar principal	main pulmonary artery	/meɪn ˈpʌlmənəri ˈɑːtəri/
33	arteriopatía	arteriopathy	/ɑːtɪərɪəpəθɪ/
34	asistencia centrífuga	centrifugal pump	/sɛnˈtrɪfjʊgəl pʌmp/
35	asistencia circulatoria	circulatory assistance	/ˈsɜːkjʊlətəri əˈsɪstəns/
36	asistencia ventricular	ventricular assistance	/vɛnˈtrɪkjʊlər əˈsɪstəns/
37	asistencia ventricular mecánica	ventricular assist device	/vɛnˈtrɪkjʊlər əˈsɪst dɪˈvaɪs/
38	aspartato transaminasa	aspartate transaminase	/ˈæspɑːteɪt ˈtrænzæmɪneɪs/
39	atomoxetina	atomoxetine	/ətəˈməʊksɪtɪn/
40	atresia pulmonar	pulmonary atresia	/ˈpʌlmənəri əˈteɪːzʃə/

41	auriculoventricular	atrioventricular	/ˈeɪtrɪəvenˌtrɪkjʊlə/
42	biatrial	biatrial	/ˈbaɪətrɪəl/
43	bicava	bicaval	/ˈbaɪkəvl/
44	bicúspide	bicuspid	/baɪˈkʌspɪd/
45	biomarcador cardiaco	cardiac biomaker	/ˈkɑːdɪæk ˈbaɪəʊmeɪkə/
46	bloqueadores beta	beta blockers	/ˈbiːtə ˈblɒkəz/
47	bloqueo de rama derecha	right bundle-branch block	/raɪt ˈbʌndl brɑːnʧ blɒk/
48	bloqueo de rama izquierda	left bundle-branch block	/lɛft ˈbʌndl brɑːnʧ blɒk/
49	bolus tracking automático	automatic bolus tracking	/ˌɔːtəˈmætɪk ˈbəʊləs ˈtrækɪŋ/
50	bomba de flujo continuo	continuous flow pump	/kənˈtɪnjʊəs fləʊ pʌmp/
51	borde inferior esternal izquierdo	lower left sternal border	/ˈləʊə lɛft sternal ˈbɔːdə/
52	borde superior esternal izquierdo	upper left sternal border	/ˈʌpə lɛft stənl ˈbɔːdə/
53	bradiarritmia	bradyarrhythmia	/ˌbreɪdɪəˈrɪðmɪə/
54	canal auriculoventricular	atrioventricular septal defect	/ˈeɪtrɪavenˌtrɪkjʊlə ˈsɛptəl dɪˈfɛkt/
55	cardiología intervencionista	interventional cardiology	/ˌɪntə(ː)ˈvɛnʃən(ə)l ˌkɑːdɪˈɒlədʒi/
56	cardiopatía congénita	congenital heart disease	/kənˈʤɛnɪtl hɑːt dɪˈziːz/
57	cardiopatía congénita	congenital cardiac malformation	/kənˈʤɛnɪtl ˈkɑːdɪæk ˌmælfɔːˈmeɪʃən/
58	cardiopatía congénita compleja	complex congenital heart disease	/ˈkɒmplɛks kənˈʤɛnɪtl hɑːt dɪˈzi-ːz/
59	cardiopatía congénita mayor	major congenital heart disease	/ˈmeɪʤə kənˈʤɛnɪtl hɑːt dɪˈziːz/
60	cardiopatía congénita simple	simple congenital heart disease	/ˈsɪmpl kənˈʤɛnɪtl hɑːt dɪˈziːz/
61	cardiopatía neonatal	neonatal heart disease	/ˌniːəʊˈneɪtəl hɑːt dɪˈziːz/
62	cardiopatías adquiridas	acquired heartdisease	/əˈkwaɪəd ˈhɑːtdɪˈziːz/
63	cardiopatías congénitas del adulto	adult congenital heart disease	/ˈædʌlt kənˈʤɛnɪtl hɑːt dɪˈziːz/
64	cardiorresonancia magnética	cardiac magnetic resonance	/ˈkɑːdɪæk mægˈnɛtɪk ˈrɛznəns/
65	catéter venoso central	central venous catheter	/ˈsɛntrəl ˈviːnəs ˈkæθɪtə/
66	cateterismo terapéutico	therapeutic catheterisation	/ˌθerəˈpjuːtɪk ˌkæθetəraɪˈzeɪʃn/

67	cava superior	superior vena cava	/sju(:)ˈpɪərɪə ˈviːnə ˈkɑːvə/
68	CC cianótica	cyanotic CHD	/ˈsaɪənəʊtɪk siː eɪʧ diː/
69	cianosis	cyanosis	/ˌsaɪəˈnəʊsɪs/
70	cianosis progresiva	progressive cyanosis	/prəʊˈgrɛsɪv ˌsaɪəˈnəʊsɪs/
71	circulación extracorpórea	extracorporeal circulation	/ˌɛkstrəkɔːrˈpɔːrɪəl ˌsɜːkjʊˈleɪʃən/
72	circulación tipo Fontan	Fontan circulation	/ˈfɔːntan ˌsɜːkjʊˈleɪʃən/
73	cirugía cardiaca en CEC	cardiac surgery under CPB	/ˈkɑːdɪæk ˈsɜːʤəri ˈʌndə siː piː biː/
74	cirugía correctora en CEC	corrective surgery under CPB	/kəˈrɛktɪv ˈsɜːʤəri ˈʌndə siː piː biː/
75	cirugía de mínima invasión	minimally invasive surgery	/ˈmɪnɪmli ɪnˈveɪsɪv ˈsɜːʤəri/
76	cirugía Fontan	Fontan surgery	/ˈfɔːntan ˈsɜːʤəri/
77	cirugía transanular	transannular surgery	/trænˈzænjlə ˈsɜːʤəri/
78	cirugía valvular	valve surgery	/vælv ˈsɜːʤəri/
79	Cistatina-C	Cystatin-C	/ˈsɪstətɪn siː/
80	coartación de aorta	aortic coarctation	/eɪˈɔːtɪk kəʊɑːˈteɪʃn/
81	colapso circulatorio	circulatory collapse	/ˈsɜːkjʊlətəri kəˈlæps/
82	comisurotomía de la válvula pulmonar	pulmonary valve commissurotomy	/ˈpʌlmənəri vælv kəmɪʃʊəˈrəʊ təmɪ/
83	comunicación interauricular	atrial septal defect	/ˈeɪtrɪəl ˈsɛptəl dɪˈfɛkt/
84	comunicación interauricular tipo ostium secundum	ostium secundum atrial septal defect	/ˈɔːstɪəm sɪˈkʌndəm ˈeɪtrɪəl ˈsɛptəl dɪˈfɛkt/
85	comunicación interventricular	ventricular septal defect	/vɛnˈtrɪkjʊlə ˈsɛptəl dɪˈfɛkt/
86	concentraciones de cTn-I	cTn-I levels	/siː tiː en aɪ ˈlɛvlz/
87	copeptina	copeptin	/kəʊˈpeptən/
88	corazón izquierdo hipoplásico	hypoplastic left heart syndrome	/ˌhaɪpəˈplæstɪk lɛft hɑːt ˈsɪndrəʊm/
89	corrección videoasistida	video-assisted repair	/ˈvɪdɪəʊ əˈsɪstɪd rɪˈpeə/
90	cortocircuito intracardiaco	intracardiac communication	/ˌɪntrəˈkɑːrdiæk kəˌmjuːnɪˈkeɪʃən/
91	creatinina	creatinine	/krɪətenɪn/
92	cribado	screening	/ˈskriːnɪŋ/
93	cribado prenatal	prenatal screening	/ˌpriːˈneɪtl ˈskriːnɪŋ/

94	cromosomopatía	chromosomal abnormality	/ˌkrəʊməˈsəʊməl ˌæbnɔːˈmælɪti/
95	cuerdas tendinosas	tendinous cords	/ˈtɛndɪnəs kɔːdz/
96	cúspide valvular	valve cusp	/vælv kʌsp/
97	defecto cardiaco congénito	congenital cardiac defect	/kənˈdʒɛnɪtl ˈkɑːdɪæk dɪˈfɛkt/
98	defecto cardíaco congénito crítico	critical congenitalheart defect	/ˈkrɪtɪkəl kənˈdʒɛnɪtəl ˈhɑːrt dɪˈfɛkt/
99	defecto congénito	birth defect	/bɜːθ dɪˈfɛkt/
100	defecto intracardíaco	intracardiac defect	/ˌɪntrəˈkɑːrdiæk dɪˈfɛkt/
101	defecto septal	septal defect	/ˈsɛptəl dɪˈfɛkt/

El uso del corpus DIACOM para la enseñanza de terminología comercial: aplicaciones y posibilidades

Dunia Hourani-Martín
Universität Leipzig / Università degli Studi di Verona

1. Introducción

El proyecto DIACOM[1] está siendo desarrollado en el Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona (Italia) en el seno del área lingüística del proyecto *Patrimonio Linguistico-Culturale e Digital Humanities*.[2] El objetivo es la construcción de un corpus, denominado DIACOM, para el estudio del lenguaje especializado del comercio en francés y español que sirva como apoyo tanto a la investigación en este campo de estudio como a la docencia de los cursos universitarios que se imparten en dicho Departamento, en concreto, el Grado en Lenguas y Culturas para el Turismo y el Comercio Internacional y el Máster en Lenguas para la Comunicación Turística y Comercial.[3] Así, este corpus se construye con una doble finalidad de explotación terminológica y didáctica de la enseñanza y de la investigación.

Muchos son los autores que destacan las múltiples aplicaciones didácticas de los corpus y su potencial en la enseñanza y el aprendizaje de lenguas (Tognini-Bonelli 2001: 14-46; Bernardini 2004; Davies

1. <https://dh.dlls.univr.it/patrimonio-linguistico-culturale.html#diacom>.

2. Dicho proyecto, que engloba numerosos subproyectos, tiene como objetivo la aplicación de las humanidades digitales al patrimonio lingüístico dentro de las diferentes tradiciones lingüísticas y se enmarca, a su vez, en el macroproyecto de excelencia *Le Digital Humanities applicate alle lingue e letterature straniere (2018-2022)*, financiado por el Ministerio italiano de Educación, Universidad e Investigación (<https://dh.dlls.univr.it/>).

3. Pero su uso no solo será interno, pues el proyecto prevé tras su finalización en 2022 poner el corpus DIACOM a disposición de la comunidad científica y didáctica de forma gratuita.

2004a y 2004b; Johansson 2009: 39-40, entre otros). El empleo de los corpus favorece la participación activa de los estudiantes en el aprendizaje, la adquisición de autonomía y habilidad para identificar fenómenos lingüísticos, así como el desarrollo de una actitud crítica al observar datos y evaluarlos.

Así pues, teniendo en cuenta las posibilidades didácticas de los corpus y, al mismo tiempo, la necesidad de que los estudiantes de lenguajes de especialidad alcancen la naturalidad que requieren para transmitir el conocimiento especializado siguiendo las convenciones propias del discurso, el objetivo de la presente contribución es proponer ejemplos de explotación del corpus DIACOM que puedan ser aprovechados con estudiantes de terminología comercial en español con un nivel B2/C1.[4]

Si bien se adopta un enfoque principalmente inductivo (cfr. Johns 1991; Bernardini 2004) en el que el estudiante es protagonista activo en el proceso de aprendizaje, los ejercicios se sitúan, como se verá, entre las aproximaciones *corpus-based* y *corpus-driven* (cfr. Tognini-Bonelli 2001: 65-100). Por un lado, el corpus constituye un instrumento auxiliar para ejemplificar o contrastar datos proporcionados previamente y cuya consulta, en parte, está guiada por el profesor para llamar la atención sobre los aspectos que se quieren trabajar; por otro lado, sirve como base empírica a través de la cual el estudiante identifica fenómenos, formula hipótesis y generaliza los resultados observados en una norma lingüística.

2. El corpus DIACOM

Como se ha mencionado, DIACOM no solo se construye como fuente para la investigación del lenguaje especializado del comercio, sino también como herramienta de aplicación didáctica en el

4. En concreto, están pensadas para su empleo en la asignatura de Lengua española 3 del itinerario de comercio internacional impartida en la Universidad de Verona. El grupo meta son estudiantes de tercer año con conocimientos previos en lexicografía, morfología y comercio. No obstante, se conciben de forma abierta para que se puedan adaptar a las diferentes necesidades de los profesores, sean estas traductológicas, terminológicas o de enseñanza de lenguajes especializados.

Departamento de Lenguas y Literaturas Extranjeras de la Universidad de Verona.[5] Con tales fines, se compila el corpus DIACOM, cuyo diseño se llevó a cabo siguiendo los criterios teórico-metodológicos de la bibliografía especializada en los ámbitos de la terminología (Bowker y Pearson 2002) y la lingüística diacrónica iberorrománica (Torruella Casañas 2017), con las adaptaciones pertinentes para permitir una amplia variedad de posibilidades de estudio y aplicación didáctica. Según nuestros parámetros de diseño, DIACOM es un corpus comparable (francés-español) y especializado del dominio del comercio. Está compuesto por textos escritos íntegros de diferentes niveles de especialización que abarcan desde mediados del siglo XIX hasta el siglo XXI. En De Beni y Hourani-Martín (2020) se detallan las características de diseño y compilación del subcorpus español de DIACOM;[6] no obstante, se resumen a continuación los principales criterios:

– Criterio cronológico: el corpus DIACOM se articula en torno a tres cortes temporales correspondientes a tres periodos decisivos para la evolución del comercio a nivel internacional, a saber, 1850-1914; 1945-1970 y 1990-2018 (cfr. Valdani y Bertoli 2014).
– Criterio extralingüístico: dado que el dominio del comercio es un ámbito pluridisciplinar e interdisciplinar (cfr. Buckley y Lessard 2005; Zettinig y Vincze 2011), DIACOM se estructura en tres grandes campos temáticos —macroeconomía y economía internacional, sectores y empresa—, cada uno de los cuales se divide, a su vez, en diferentes subdominios.[7]

5. En el área de francés ya se ha aplicado DIACOM con buenos resultados en la asignatura de Langue française – Terminologie et traduction – filière commerce (Bonadonna 2020), destinada a estudiantes de tercer año de grado con un nivel de francés correspondiente al B2/C1 y conocimientos previos de lexicología y de comercio.

6. El diseño de DIACOM se llevó a cabo de forma conjunta para francés y español, la única diferencia reside en la consideración de las distintas variedades diatópicas en el subcorpus español, en concreto, se recogen textos de España y de todos los países hispanohablantes de América.

7. Esta clasificación temática del corpus realizada sobre la base de la bibliografía

Dominios	Subdominios
Macroeconomía y economía internacional (Comercio global entre países, macroáreas, comercio dentro de un país determinado, etc.)	Relaciones internacionales / Política comercial Países / Áreas de estudio Aspectos sociales
Sectores (Comercio e internacionalización de sectores específicos)	Productos Servicios
Empresa (Estrategias de internacionalización y comerciales de la empresa; actividades para el comercio internacional y la internacionalización de la empresa)	Administración (*management*) *Marketing* Logística Comercio electrónico Derecho

Tabla 1. Dominios y subdominios considerados en el corpus DIACOM (De Beni y Hourani-Martín 2020: 85).

– Criterio textual: DIACOM incorpora muestras textuales de diferente tipo para representar el dominio de especialidad en toda su amplitud y variedad textual, por ello se seleccionan textos de diverso tipo teniendo en cuenta el emisor del texto: textos institucionales, textos empresariales, textos académicos y textos periodísticos.

El objetivo final de la determinación de todos estos criterios de diseño del corpus es garantizar la representatividad del corpus y, por ende, su calidad. Al mismo tiempo, también deben servir para configurar la base de datos que almacena los textos de DIACOM y la herramienta de consulta creada en el marco del proyecto que permitirá su posterior explotación tanto de forma global como a partir de búsquedas restringidas, por ejemplo, a un periodo determinado o a un subdominio concreto.[8]

especializada (en particular, Buckley y Lessard 2005; Zettinig y Vincze 2011) fue confirmada por un experto en la materia, el profesor Fabio Cassia de la Universidad de Verona.

8. Tanto la base de datos del proyecto como la herramienta de consulta están siendo desarrolladas por el equipo de informáticos del proyecto de excelencia a partir de Alfresco <https://www.alfresco.com> y KonText <https://kontext.korpus.cz/first_form>, respectivamente.

3. Aplicaciones didácticas del DIACOM

A continuación, se presentan una serie de ejemplos de aplicaciones didácticas del corpus para el aprendizaje de terminología comercial. Dado que DIACOM aún se encuentra en su fase de construcción, para el presente trabajo se ha utilizado un subcorpus creado a partir de textos académicos de Argentina, España y México en el periodo de 1990-2018. Dicho corpus cuenta con un total de 1.064.140 *tokens* (34.124 *types*). Para la gestión del corpus y su análisis, así como para la ejemplificación de las búsquedas y posibles resultados se ha utilizado el programa Sketch Engine.[9] Cabe destacar que todos los ejemplos aquí presentados han sido extraídos del corpus de estudio.

3.1. La utilidad de las listas de palabras

El primer acercamiento a un corpus suele ser a partir de la función *Lista de Palabras*. A través de esta herramienta, el programa de gestión del corpus presenta en forma de listado las unidades que contiene el corpus tanto en su totalidad como restringidas a una categoría gramatical concreta si el corpus está anotado morfológicamente. Los resultados que genera la herramienta no solo son relevantes para la extracción de unidades terminológicas,[10] sino también para la comprensión y la estructuración del ámbito de especialidad. En este sentido, se le puede pedir al estudiante elaborar, a partir de la lista de palabras, una clasificación formal basada en el tipo de categoría gramatical de los términos encontrados (nominal, adjetival o verbal, por ejemplo, si solo tenemos en cuenta las unidades simples) o una clasificación conceptual basada en el subdominio al que se adscriban

9. <https://www.sketchengine.eu/>.

10. La extracción de términos también se puede realizar manualmente, a partir de búsquedas específicas cuando se dispone de los términos de antemano porque se conocen o se han trabajado con anterioridad, o automáticamente, por ejemplo, con la función *Palabras Clave* de Sketch Engine o mediante el programa TermoStat de la Universidad de Montreal (<http://termostat.ling.umontreal.ca>). En la extracción automática, la herramienta de análisis de corpus proporciona una lista de candidatos a término (tanto monoverbales como pluriverbales); no obstante, el carácter especializado de las unidades debe ser verificado a través de los contextos.

las unidades encontradas (política comercial, *marketing* o logística, entre otros), para esto último se puede utilizar la clasificación por sectores elaborada en el marco del proyecto (*vid*. Tabla 1).

Sin ignorar la carga terminológica que puedan tener unidades como los adjetivos, los verbos o los términos pluriverbales, la categoría morfológica por excelencia para representar y transmitir el conocimiento especializado es el sustantivo. Por ello, un buen punto de partida puede ser pedir a los estudiantes que generen una lista de unidades nominales.

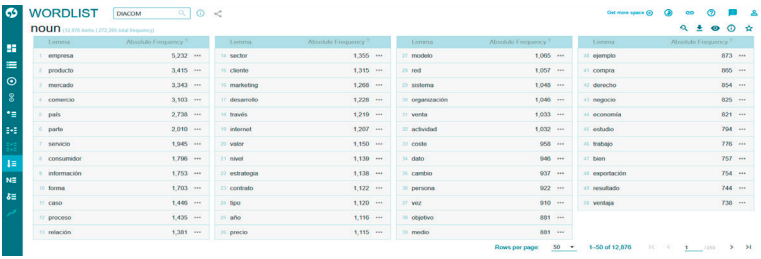


Figura 1. Primeros 50 resultados de la lista de sustantivos del subcorpus DIACOM.

En la lista de resultados que muestra la Figura 1, el estudiante podrá observar que, entre los 50 primeros sustantivos más frecuentes del subcorpus, se encuentran conceptos clave del dominio de especialidad (*empresa, producto, mercado, comercio, servicio, consumidor,* etc.). La terminología es el reflejo del desarrollo y el estado actual de todo ámbito de especialidad,[11] por lo tanto, la lista de palabras no solo es útil para la extracción terminológica en sí, sino que también

11. Puesto que en el diseño de DIACOM se han previsto tres periodos temporales (1850-1914; 1945-1970 y 1990-2018), se podrán plantear ejercicios de búsqueda global en todos los periodos o limitada a uno de ellos, por ejemplo. Este último tipo de ejercicios puede mostrar la evolución del dominio de especialidad tanto desde el punto de vista conceptual como denominativo: la transformación de la logística comercial, el surgimiento del comercio electrónico y las nuevas realidades asociadas al mismo, etc.

aporta información esencial acerca del ámbito de especialidad y las características de su terminología. Así pues, además de servir para la recopilación de unidades terminológicas, su explotación didáctica también puede destinarse a la creación de redes semánticas.

3.2. Las concordancias y la importancia del contexto

Una de las principales ventajas del trabajo con corpus reside en la recuperación de contextos comunicativos reales a través de los cuales el estudiante podrá observar el comportamiento de los términos *in vivo*. Para ello, podrá hacer uso de la herramienta *Concordancias*, con la que se pueden recuperar todos los contextos en los que aparece una unidad y que no solo proporciona ejemplos reales de uso de un término dado, sino también permite observar fenómenos de otro tipo (colocaciones y combinaciones frecuentes de palabras, términos relacionados, variantes, etc.).

Figura 2. Lista de concordancias del término *empresa*.

A partir de una lista de concordancias, como, por ejemplo, la que muestra la Figura 2 relativa al término *empresa*, se puede proponer al estudiante la realización de una variada cantidad de ejercicios relacionados con diversos fenómenos terminológicos. En lo siguiente presentamos dos posibilidades de análisis de relaciones conceptuales (hiperonimia-hiponimia) y denominativas (neología y variación).

3.2.1. Relaciones de hiperonimia e hiponimia

A partir de las concordancias, se puede pedir a los estudiantes que extraigan ejemplos de contextos en los que *empresa* aparezca dentro de términos pluriverbales,[12] a saber, aquellos formados a partir de un sintagma nominal con núcleo sustantivo y adyacente adjetival (antepuesto o pospuesto) o sintagma preposicional (*grandes empresas*, *empresa multinacional* o *empresa de comercio*) que hagan referencia a entidades.[13] Algunos ejemplos podrían ser:

1. En el presente trabajo se estudian los factores asociados con la adopción del comercio electrónico, así como el nivel de preparación digital de las *micro, pequeñas y medianas empresas* (MiPyMEs) comerciales y de servicios de Córdoba, Argentina. (AR01)
2. La importancia de emplear el comercio electrónico en las *micro y pequeñas empresas* (MYPES) del municipio de Nezahualcóyotl de giro comercial, [*sic*] radica en incrementar las ventas de productos empleando el internet como nuevo canal. (MX01)
3. El interés creciente de estudiar la adopción del comercio electrónico en las *pequeñas y medianas empresas* (PyME) de países en desarrollo radica en su potencial para la creación de valor. (AR01)
4. Las *empresas multinacionales*, [*sic*] son firmas o conglomerados de corporaciones trasnacionales, con un sistema de producción, o prestación de servicios, integrado por unidades localizadas en distintos países, que responden a estrategias centralmente planificadas en una casa matriz aunque no exclusivamente la propiedad de todo o parte del capital sea de donde está la casa matriz. (MX02)

12. Se emplea aquí esta denominación —*términos pluriverbales*— dada la falta de consenso existente en la bibliografía especializada a la hora de nombrar y concebir las unidades terminológicas de formación compleja (*vid.* Quiroz Herrera 2008: 39-56; Sanz Vicente 2011: 258-327); por ejemplo, algunos trabajos terminológicos en español se han referido a ellos como *términos polilexemáticos* (Cabré *et al*. 1996), *compuestos sintagmáticos* (Gómez de Enterría y Martí 2009), *compuestos sintagmáticos nominales* (Sanz Vicente 2011) o *compuestos nominales* (Cabezas García 2019).

13. Con términos clave del dominio de especialidad, la alta frecuencia de aparición no permitirá una criba manual y minuciosa de todos los resultados que ofrezca la lista de concordancias, pero el análisis se puede limitar a un número determinado de contextos que sirvan como punto de partida y que, después, se podrá complementar, por ejemplo, con la consulta de otros instrumentos terminográficos o con la combinación de otras funciones de búsqueda en el programa de análisis de corpus, como la herramienta *Word Sketch*, que ofrece combinaciones frecuentes de palabras.

A partir de los resultados, los estudiantes podrán crear mapas conceptuales en los que se refleje la relación de hiperonimia, hiponimia y cohiponimia entre las unidades extraídas (Figura 3).

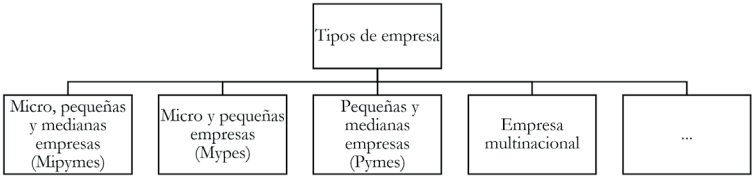


Figura 3. Mapa conceptual de las formas de empresa.

La estructuración gráfica del conocimiento favorece el aprendizaje de la terminología, así como facilita la comprensión de las relaciones conceptuales que se establecen en el dominio de especialidad. Asimismo, este esquema podrá ser también completado con definiciones o características extraídas de los propios contextos o de consultas a diccionarios especializados u otras fuentes de documentación.

3.2.2. Neología y variación

En los ejemplos (1), (2) y (3) se pueden observar variantes en forma de sigla que aparecen justo detrás de la unidad desarrollada entre paréntesis, a saber, *MiPyMEs*, *MYPES* y *PyME* respectivamente para *micro, pequeñas y medianas empresas*, *micro y pequeñas empresas* y *pequeñas y medianas empresas*. Los términos procedentes de fenómenos de reducción léxica, siglación y acronimia fundamentalmente, suelen suponer problemas de comprensión en los estudiantes, pues, al encapsular estos un sintagma pleno, pueden generar tanto opacidad si se desconoce la relación de equivalencia entre el sintagma que representan y su forma reducida como ambigüedad si se ignora su verdadero significado dentro de una situación comunicativa (Giraldo Ortiz 2008: 32). Por ello y porque constituyen un procedimiento de creación neológica habitual en este ámbito de especialidad (Gómez de Enterría 1992; Gómez de Enterría y Martí 2009: 31-

32) son interesantes para el trabajo en clase. Así, se puede pedir a los estudiantes extraer todos los contextos en los que aparezcan estas unidades para estudiar el proceso de lexicalización al que se han visto sometidas desde sus primeras apariciones como forma desarrollada hasta su incorporación a la lengua como término pleno con flexión propia, pasando por las diferentes fases de pérdida de mayúsculas, como muestran los ejemplos del corpus: *pequeñas y medianas empresas → PyME/PYME → PyMEs/PYMES/PYMEs → Pymes → pymes*.

La variación ortográfica derivada del uso combinado de mayúsculas y minúsculas en los ejemplos *PyME PyMEs*, *PYMEs* y *Pymes* es un factor esencial para determinar el grado de lexicalización del término. Al respecto resalta la Academia que, si bien es cada vez más frecuente mezclar mayúsculas y minúsculas en los componentes de siglas no lexicalizados, a saber, mayúscula en inicial de la sigla o minúscula en conectores gramaticales —o también, como vemos en los ejemplos, en marca de plural—, se trata de un método no ortodoxo pero admisible en la escritura de siglas y acrónimos (*OLE* 2010: 511). El uso reiterado en la lengua acabará dando con el producto lexicalizado que se atiene a las normas del léxico general donde esta alternancia no está permitida y cuya escritura es, por lo tanto, en minúscula. Por supuesto, no todas las unidades estudiadas acabarán siendo un producto lexicalizado, pero servirán, a su vez, para concienciar a los estudiantes acerca de la variación terminológica, en este caso procedente de la inestabilidad denominativa de las formas nuevas desde que surgen por primera vez hasta que se asientan con una forma lexicalizada en el lenguaje.[14]

3.3. Las combinaciones habituales y sus estructuras morfológicas

Los corpus son un recurso de gran utilidad a la hora de estudiar estructuras recurrentes en la lengua, como señala Aijmer (2009: 8): «Corpora draw attention to complex patterns and phraseology rather than regularities and supports [*sic*] the view of language learning as a

14. A tal efecto, también pueden servir como referencia temporal las obras lexicográficas; en el caso de *pyme*, la primera vez que se recoge esta voz en el diccionario académico fue en 2001 (*MDA* 2013: *s. v. pyme*).

complex process involving hypothesis formation and testing». Por ello, el análisis de los términos pluriverbales no es solo una forma de estudiar terminología, sino que, si se incide sobre su estructura morfológica, permitirá también extraer patrones productivos para la formación de combinaciones lingüísticas que podrán ser reproducidas.

El punto de partida de este ejercicio será uno de los términos clave del dominio de especialidad, por ejemplo, *mercado* (*vid.* Figura 1). Es necesario seleccionar para esta tarea un término que sea tanto representativo como frecuente dentro del discurso especializado.[15] Así, habrá una mayor probabilidad de que actúe dentro de otras unidades mayores, bien en forma de núcleo bien como modificador de la unidad, y, por lo tanto, de determinar patrones reiterativos y extrapolables. La tarea consiste en extraer las unidades pluriverbales más frecuentes en las que se encuentra inserto este término y clasificarlas según su esquema morfosintáctico. Para ello pueden hacer uso de la herramienta *Word Sketch* o del programa TermoStat,[16] en ambos casos se obtendrán candidatos a término cuyo valor especializado deberá ser verificado en el contexto. Algunos de los términos extraídos, ya clasificados según su esquema formal, serían los siguientes:

Sustantivo + adjetivo	Sustantivo + preposición + sustantivo
Mercado exterior, mercado internacional, mercado doméstico, mercado local, mercado potencial, mercado electrónico, mercado interno, mercado extranjero, mercado global, mercado nacional, etc.	Cuota de mercado, mercado de destino, mercado de divisas, mercado de exportación, segmento de mercado, nicho de mercado, mercado de origen, estudio de mercado, mercado de capitales, potencial de mercado, valor de mercado, etc.

Tabla 2. Términos pluriverbales que contienen *mercado* según su estructura morfológica.

15. Es cierto, no obstante, que la frecuencia no es un requisito indispensable para determinar la terminologicidad de una unidad, pero en el caso concreto de este ejercicio, en el que se están intentando extraer esquemas morfológicos productivos, tiene especial importancia.

16. Entre las funciones de TermoStat destaca la presentación de la información estadística de las estructuras morfológicas de los candidatos a término.

Los resultados ponen de manifiesto la productividad de dos esquemas: [sustantivo + adjetivo] y [sustantivo + preposición + sustantivo].[17] En cuanto al primero, se puede comprobar que, en todos los casos, el adyacente adjetival aparece pospuesto, frente a las combinaciones del tipo [adjetivo + sustantivo] que no tienen carácter especializado: *importante mercado*, *propio mercado*, *determinado mercado*, *creciente mercado*, etc. Desde el punto de vista semántico, el adjetivo califica al núcleo terminológico y convierte la unidad en hipónimo de este. Por su parte, en los términos clasificados bajo la estructura [sustantivo + preposición + sustantivo], aparece exclusivamente como nexo la preposición *de* y se manifiestan dos posibilidades: por un lado, *mercado* aparece como núcleo en primera posición [*mercado* + *de* + sustantivo], dando lugar, de nuevo, a unidades hipónimas; por otro, aparece como adyacente nominal en segunda posición [sustantivo + *de* + *mercado*] precisando a su núcleo como elemento constitutivo.

Teniendo en cuenta el enfoque eminentemente inductivo de la actividad, se puede pedir al estudiante que, a la luz de los resultados anteriores, formule una hipótesis acerca de los esquemas morfológicos más productivos para la formación de unidades pluriverbales en este ámbito de especialidad que, posteriormente, deberá verificar, refutar o precisar tras el análisis de las estructuras de las unidades pluriverbales que conforman otros términos clave.[18] Asimismo, una vez extraídas las estructuras frecuentes, estas también se podrán contrastar con las habituales en otras lenguas, como el francés, y servirán como apoyo no solo a la enseñanza de la terminología, sino también de la traducción (extracción bilingüe de terminología, elaboración de fichas terminográficas, creación de glosarios, establecimiento de equivalentes de traducción, etc.).

17. La productividad de estas estructuras también se corrobora en la bibliografía especializada (por ejemplo, Gómez de Enterría y Martí 2009: 29-30).
18. Cuando DIACOM haya sido anotado, también se podrán realizar estudios cuantitativos sobre los patrones morfológicos.

3.4. El trabajo combinado con diccionarios y corpus

Entre las 20 unidades nominales más frecuentes presentadas en la Figura 1 encontramos un préstamo lingüístico procedente del inglés: *marketing*. Esto no es de extrañar teniendo en cuenta que uno de los recursos a la hora de crear términos para denominar nuevas realidades es la incorporación de unidades procedentes de otras lenguas y que, en concreto, los ámbitos del comercio y de la economía muestran gran permeabilidad a los anglicismos (Gómez de Enterría y Martí 2009: 22-24; Mateo Martínez 2007: 197).

Los préstamos lingüísticos ya asentados en la lengua pueden servir como ejercicio de contraste y análisis entre herramientas lexicográficas y corpus. Si atendemos al ejemplo concreto de *marketing*, se puede pedir al estudiante que consulte el término en los diccionarios académicos, por ejemplo, el *Diccionario de la Lengua Española* (*DLE* 2019), el *Diccionario Panhispánico de Dudas* (*DPD* 2005), el *Nuevo Tesoro Lexicográfico de la Lengua Española* (*NTLLE* en línea) y el *Mapa de Diccionarios Académicos* (*MDA* 2013), y compendie los datos obtenidos. La búsqueda en estas obras proporcionará al estudiante una visión sobre la evolución del término en los recursos lexicográficos desde su primera entrada en el diccionario académico hasta su forma actual en la última edición de 2019, así como las recomendaciones de uso.[19]

19. No olvidamos el hecho de que se trata de diccionarios generales y, por lo tanto, puede que la aparición del término en el lenguaje de especialidad haya sido anterior. No obstante, teniendo en cuenta que el discurso del comercio presenta un carácter semitécnico e, incluso, divulgativo y que cuenta entre sus usuarios no solo con economistas teóricos y con expertos del mundo del comercio y de las finanzas, sino también con periodistas especializados en estos ámbitos y hablantes legos interesados en este mundo profesional (Mateo Martínez 2007: 193-194), es habitual que los términos acaben trascendiendo a la lengua general. En cualquier caso, consideramos que la búsqueda y recuperación de la información en obras lexicográficas complementa el trabajo con corpus y es un buen ejercicio para desarrollar tanto la destreza documental de los estudiantes como su habilidad para evaluar y analizar información.

Edición[20]	*Marketing*	*Mercadotecnia*
RAE U 1984	-	**mercadotecnia.** (De *mercado* y *-tecnia*) f. *Com*. Técnica del mercado.
RAE M 1984	[**marketing**. (Voz inglesa.) m. Conjunto de técnicas comerciales para hacer más rentable un producto.	**mercadotecnia.** f. *Com*. Técnica del mercado.
RAE M 1989	[**marketing**. (Voz inglesa.) m. Conjunto de técnicas comerciales para hacer más rentable un producto.	**mercadotecnia.** f. *Com*. Conjunto de principios y prácticas que buscan el aumento del comercio, especialmente de la demanda, y estudio de los procedimientos y recursos empleados.
RAE U 1992	**marketing.** (Voz inglesa.) m. **mercadotecnia.**	**mercadotecnia.** (De *mercado* y *-tecnia*) f. Conjunto de principios y prácticas que buscan el aumento del comercio, especialmente de la demanda, y estudio de los procedimientos y recursos tendentes a este fin.
RAE U 2001	**marketing.** (Voz inglesa.) m. **mercadotecnia.**	**mercadotecnia.** (De *mercado* y *-tecnia*) 1. f. Conjunto de principios y prácticas que buscan el aumento del comercio, especialmente de la demanda. 2. f. Estudio de los procedimientos y recursos tendentes a este fin.
DLE 2019	*marketing.* (Voz inglesa.) m. **mercadotecnia.**	**mercadotecnia.** (De mercado y *-tecnia*.) 1. f. *Econ*. Conjunto de principios y prácticas que buscan el aumento del comercio, especialmente de la demanda. 2. f. *Econ*. Estudio de la mercadotecnia.

Tabla 3. *Marketing* y *mercadotecnia* en las sucesivas ediciones de los diccionarios académicos (*s. v. marketing* / *mercadotecnia*).[20]

Además de las diferencias apreciables en la microestructura de las entradas de las sucesivas ediciones desde su primera aparición (la información etimológica, las marcas diatécnicas y, principalmente, las

20. Las obras lexicográficas del *NTLLE* se citan según su sigla correspondiente y el año de publicación: en particular, los diccionarios académicos se acompañan de M por Manual y U por Usual.

acepciones que, a su vez, revelan la evolución en la complejidad del propio ámbito de especialidad), los resultados de la búsqueda en las obras lexicográficas sirven para mostrar la convivencia de dos términos —*marketing* y *mercadotecnia*— desde 1984 hasta la imposición de este último como variante normativa de preferencia.[21]

Por su parte, la entrada de *marketing* en el *DPD* (2005: *s. v. marketing*) remite a la voz *márquetin*, adaptación gráfica del término extranjero, que se define como «conjunto de estrategias empleadas para la comercialización de un producto y para estimular su demanda» (*ibid. s. v. márquetin*). Sin embargo, esta variante propuesta no aparece en ninguna de las obras lexicográficas de la Academia consultadas.[22]

Como indicábamos *supra*, una de las mayores utilidades del trabajo con corpus es poder observar los términos dentro de su contexto comunicativo. Así, el estudiante podrá analizar si, efectivamente, existe una convivencia entre el préstamo y el término patrimonial en los textos de especialidad o si una variante se impone sobre la otra. Los resultados de la búsqueda en el subcorpus DIACOM muestran que la denominación de preferencia es *marketing* con un total de 1268 ocurrencias frente a los 26 casos de *mercadotecnia*; *márquetin* no se constata ni una sola vez. Además de la frecuencia total de aparición, otro indicador del dinamismo de un término es la cantidad de obras diferentes en las que se recoge, es decir, la cantidad de autores distintos que lo emplean, que también corrobora los datos anteriores: *marketing* aparece en 20 textos tanto de Argentina como de España y de México y *mercadotecnia* se recoge en 7 textos diferentes y, si bien también aparece en los 3 países estudiados, el número de ocurrencias es mayor en textos

21. De hecho, si el estudiante emprende la búsqueda del término *marketing* sin tener mayor conocimiento de la variación existente para expresar este concepto en español, solo se topa con la referencia —y remisión— a *mercadotecnia* en la entrada de 1992.

22. El *DPD* también indica que en muchos países americanos se emplea el término *mercadeo* como variante de *marketing*. No obstante, no entramos aquí a tratar más detenidamente esta tercera posibilidad que podría ser considerada un caso de *polisemia intradominio* (cfr. Kostina 2009: 337 y ss.) teniendo en cuenta su definición: «1. m. Acción y efecto de mercadear. 2. m. *Com*. Conjunto de operaciones por las que ha de pasar una mercancía desde el productor al consumidor» (*DLE* 2019: *s. v. mercadeo*).

de México (88 %).[23] Los resultados del corpus ponen de manifiesto no solo la vitalidad de este anglicismo en el ámbito de especialidad estudiado, sino también que los expertos no siempre siguen las recomendaciones normativas (cfr. Gómez de Enterría y Martí 2009: 24-26). Ejercicios de este tipo resaltan la importancia de los corpus como fuente de datos reales para examinar el comportamiento de los términos y, por lo tanto, su utilidad como complemento —que no sustituto— documental a las obras lexicográficas.

4. A modo de conclusión

Más allá del estudio de fenómenos lingüísticos, terminológicos y fraseológicos, DIACOM hará posible la elaboración de actividades y unidades didácticas que puedan llevarse a cabo en numerosos contextos educativos, tanto en cursos de terminología o de traducción como en el aula de español especializado del comercio o español de los negocios. La muestra de actividades aquí presentada tan solo es una gota en el océano de posibilidades pedagógicas que permitirá este recurso. Los pasos futuros en la explotación didáctica del corpus en general se pueden resumir en los siguientes puntos:

– Puesta en práctica de las actividades en los cursos universitarios del Grado en Lenguas y Culturas para el Turismo y el Comercio Internacional y del Máster en Lenguas para la Comunicación Turística y Comercial de la Universidad de Verona. También mediante el empleo de la herramienta de consulta de DIACOM creada en el marco del proyecto.
– Confección de actividades que permitan el estudio de los niveles formal, discursivo y conceptual del lenguaje de especialidad. Asimismo, elaboración de actividades centradas en fenómenos

23. En el caso del subcorpus utilizado en el presente trabajo, solo se ha podido analizar su presencia en la actualidad. Sin embargo, una vez terminado, DIACOM permitirá comprobar, asimismo, si estas voces ya aparecían con anterioridad en textos especializados y/o en qué momento comienzan a extenderse entre la comunidad de expertos, así como también cuál de las denominaciones se impone en el uso, incluso desde el punto de vista diatópico.

lingüísticos que se manifiestan en los tres ejes del subcorpus español, a saber, el tiempo, el espacio y la tipología textual: por ejemplo, actividades sobre la evolución diacrónica de los términos, su variación diatópica o la variación ocasionada por el nivel de especialización de los textos.
– Creación de actividades desde una perspectiva bilingüe y contrastiva francés-español.
– Difusión de material didáctico que facilite a los profesores la enseñanza del discurso de especialidad y mejore, a partir del acceso a material real, el aprendizaje y, por ende, la empleabilidad de los estudiantes.

Referencias bibliográficas

Aijmer, Karin (2009): «Introducion: Corpora and language teaching», en Karin Aijmer (ed.), *Corpora and Language Teaching*, Amsterdam/Philadelphia, John Benjamins (*Studies in Corpus Linguistics*, n. 33), pp. 1-10.

Bernardini, Silvia (2004): «Corpora in the classroom: An overview and some reflections on future developments», en John Sinclair (ed.), *How to Use Corpora in Language Teaching*, Amsterdam/Philadelphia, John Benjamins (*Studies in Corpus Linguistics*, n. 12), pp. 15-36.

Bonadonna, Maria Francesca (2020): «L'enseignement des notions métaterminologiques en français L2: un parcours fondé sur les corpus spécialisés», *Cahiers de Lexicologie*, n. 117, pp. 11-34.

Bowker, Lynne; Pearson, Jennifer (2002): *Working with Specialized Language. A Practical Guide to Using Corpora*, London, Routledge.

Buckley, Peter J.; Lessard, Donald R. (2005): «Regaining the Edge for International Business Research», *Journal of International Business Studies,* vol. 36, n. 6, pp. 595-599.

Cabezas García, Melania (2019): *Los compuestos nominales en terminología: formación, traducción y representación*, Universidad de Granada, Tesis doctoral dirigida por Pamela Faber Benítez y Pilar León Arauz.

Cabré, M.ª Teresa; Estopà, Rosa; Lorente, Mercè (1996): «Terminología y fraseología», en *Actas del V Simposio de Terminología Iberoamericana*. En línea, <https://www.academia.edu/5304839/Terminolog%C3%ADa_y_Fraseolog%C3%ADa._MT_Cabre_M_LORENTE_R_ESTOP%C3%80__Actas_del_V_Simposio_de_Terminolog%C3%ADa_Iberoamericana_1996> [Consulta: 21/09/2020].

Davies, Mark (2004a): «Student Use of Large, Annotated Corpora to Analyze Syntactic Variation», en Guy Aston, Silvia Bernardini, Dominic Stewart (eds.), *Corpora and Language Learners*, Amsterdam/Philadelphia, John Benjamins, pp. 259-269.

Davies, Mark (2004b): «Student Use of Large Corpora to Investigate Language Change», en Ulla Connor, Thomas A. Upton (eds.), *Applied Corpus Linguistics: A Multidimensional Perspective*, Amsterdam, Rodopi, pp. 207-221.

De Beni, Matteo; Hourani-Martín, Dunia (2020): «El discurso especializado del comercio en español: diseño de un corpus para su estudio diacrónico», *Lexis*, vol. 44, n. 1, pp. 75-112.

DPD = Real Academia Española; Asociación de Academias de la Lengua Española (2005)*: Diccionario Panhispánico de Dudas*. En línea, <https://www.rae.es/dpd/> [Consulta: 21/09/2020].

DLE = Real Academia Española; Asociación de Academias de la Lengua Española (2019): *Diccionario de la Lengua Española*, versión electrónica 23.3. En línea, <http://www.rae.es> [Consulta: 21/09/2020].

Giraldo Ortiz, John Jairo (2008): *Análisis y descripción de las siglas en el discurso especializado de genoma humano y medio ambiente*, Universitat Pompeu Fabra, Tesis doctoral dirigida por María Teresa Cabré Castellví.

Gómez de Enterría, Josefa (1992): «Las siglas en el lenguaje de la economía», *Revista de filología románica*, n. 9, pp. 267-274.

Gómez de Enterría, Josefa; Martí, Manuel (2009): «El vocabulario de la economía actual (1998-2008)», en Natividad Gallardo San Salvador, Josefa Gómez de Enterría (eds.), *Estudios de léxico especializado*, Granada, Atrio (*Traducción en el atrio*, n. 17), pp. 15-133.

Johansson, Stig (2009): «Some Thoughts on Corpora and Second-Language Acquisition» en Karin Aijmer (ed.), *Corpora and Language Teaching*, Amsterdam/Philadelphia, John Benjamins (*Studies in Corpus Linguistics*, n. 33), pp. 33-44.

Johns, Tim (1991): «Should You Be Persuaded: Two Samples of Data-Driven Learning Materials», *English Language Research Journal*, vol. 4, (*Classroom Concordancing*, ed. Tim Johns, Philip King), pp. 1-16.

Mateo Martínez, José (2007): «El lenguaje de las ciencias económicas», en Enrique Alcaraz Varó, José Mateo Martínez, Francisco Yus Ramos (eds.), *Las lenguas profesionales y académicas*, Barcelona, Ariel, pp. 191-203.

MDA = Real Academia Española (2013): *Mapa de diccionarios*. En línea, <http://web.frl.es/ntllet> [Consulta: 21/09/2020].

NTLLE = Real Academia Española: *Nuevo tesoro lexicográfico de la lengua española.* En línea, <http://buscon.rae.es/ntlle/SrvltGUILoginNtlle> [Consulta: 21/09/2020].

OLE = Real Academia Española (2010): *Ortografía de la lengua española*. En línea, <http://aplica.rae.es/orweb/cgi-bin/buscar.cgi> [Consulta: 21/09/2020].

Quiroz Herrera, Gabriel Ángel (2008): *Los sintagmas nominales extensos especializados en inglés y en español: descripción y clasificación en un corpus de genoma*, Universitat Pompeu Fabra, Tesis doctoral dirigida por Mercè Lorente Casafont.

Sanz Vicente, María Lara (2011): *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español*, Universidad de Salamanca, Tesis doctoral dirigida por Joaquín García Palacios.

Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*, Amsterdam/Philadelphia, John Benjamins (*Studies in Corpus Linguistics*, n. 6).

Torruella Casañas, Joan (2017): *Lingüística de corpus: génesis y bases metodológicas de los corpus (históricos) para la investigación en lingüística*, Frankfurt am Main, Peter Lang.

Valdani, Enrico; Bertoli, Giuseppe (2014): *Marketing internazionale*, Milano, EGEA.

Zettinig, Peter; Vincze, Zsuzsanna (2011): «The Domain of International Business: Futures and Future Relevance of International Business», *Thunderbird International Business Review*, vol. 53, n. 3, pp. 337-349.

Corpus paralelos y enseñanza del español LE/L2 a sinohablantes. Estado de la cuestión y propuestas didácticas*

Weiqian Wei
Universitat de Barcelona

Mar Cruz Piñol
Universitat de Barcelona

Kris Buyse
KU Leuven / Universidad Nebrija

1. Presentación

En este capítulo pretendemos llamar la atención sobre las aplicaciones de un tipo especial de corpus multilingües, los corpus paralelos, en el contexto específico de la enseñanza del Español como Lengua Extranjera o Segunda (en adelante, ELE) a estudiantes sinohablantes. Para ello, en el apartado 2 revisamos el estado de la cuestión en relación con los corpus paralelos y sus usos prácticos en el marco del ELE. En el apartado 3 ofrecemos una selección argumentada de corpus paralelos chino-español disponibles en línea y los analizamos a partir de una serie de criterios. Finalmente, en el apartado 4 proponemos dos actividades y otros posibles usos de los corpus paralelos para enseñar ELE a los estudiantes sinohablantes.[1]

2. Los corpus paralelos y sus aplicaciones didácticas

2.1. ¿Qué es un corpus paralelo?

Los corpus que recogen muestras en más de una lengua pueden ser bilingües (si incluyen textos en dos lenguas) o multilingües (si contienen

* Este trabajo tiene su origen y se basa en Wei (2020).

1. En el momento de redacción de estas páginas, todavía no se han llevado a la práctica las propuestas didácticas que aquí se presentan. En el estudio experimental que está desarrollando una de las autoras (Weiqian Wei) en el marco de su tesis doctoral, se pondrán en práctica estas y otras propuestas.

más lenguas). En ambos casos, los textos que conforman el corpus se tienen que seleccionar según idénticos criterios para todas las lenguas representadas (Hallebeek 1999). Si además el corpus contiene la misma colección de textos en dos o más idiomas (uno de ellos es el idioma original y los otros, la traducción), estaremos hablando de corpus paralelos (McEnery y Hardie 2012). Los corpus paralelos pueden clasificarse a partir de dos criterios: según los idiomas incluidos y según la dirección de traducción (McEnery y Xiao 2007). Los corpus paralelos juegan un papel destacable especialmente para las investigaciones en traducción, los estudios contrastivos, la lingüística computacional y la enseñanza de lenguas extranjeras. En comparación con los corpus monolingües, los corpus paralelos aportan ventajas como mostrar la correspondencia y el contraste entre dos lenguas (Barlow 2000; Kenning 2010). Según los idiomas que incluya, un corpus paralelo puede ser bilingüe (los que constan de textos en dos idiomas) o multilingüe (los que constan de textos en más idiomas). Según la dirección de traducción, se pueden clasificar en unidireccionales, si la dirección de la traducción es unívoca (por ejemplo, textos españoles traducidos al chino); bidireccionales, cuando la traducción se efectúa en los dos sentidos (por ejemplo, textos españoles traducidos al chino y viceversa), o multidireccionales, cuando un mismo original es traducido a varios idiomas.

En este contexto conviene llamar la atención hacia un tipo de corpus muy especial que se ha llegado a denominar «corpuscionario». Puesto que las fronteras entre el corpus y el diccionario vienen desapareciendo en los últimos tiempos, el neologismo «corpuscionario» (acuñado por Alonso Ramos 2009) se refiere tanto a los diccionarios enriquecidos con corpus (generalmente de acceso libre y con información textual) como a los corpus complementados con un diccionario. Según Alonso Ramos (2009: 1195), un ejemplo de estos tipos híbridos podría ser la «reutilización de un corpus paralelo aunque complementado con otros repertorios lexicográficos». Así, las plataformas *Linguee*,[2] *Glosbe* y *Re-*

2. Un diccionario en línea que recopila las traducciones realizadas por humanos, alineadas con su texto de origen, y tomadas de sitios web multilingües (Alonso Jiménez 2013). Cabe mencionar que, a pesar de que en *Linguee*, un «corpuscionario»

verso muestran dicha interacción entre corpus y diccionario, es decir, son diccionarios bi-/multilingües de libre acceso en línea que están basados en (o complementados con) corpus paralelos alineados.

Con el objetivo de conocer mejor los corpus paralelos y compararlos entre sí, a continuación, vamos a definir tres conceptos clave en la Lingüística de corpus (en adelante, LC): las concordancias paralelas, los listados de frecuencias y la lematización. En primer lugar, a diferencia de las concordancias *KWIC*,[3] las concordancias paralelas[4] (Figura 1) se refieren a las listas de ocurrencias de una palabra o una combinación de palabras junto con sus contextos y alineadas con sus traducciones en una o más lenguas de forma paralela (Frankenberg-Garcia 2005). Las concordancias que ofrecen los corpus son una herramienta de suma importancia y muestran múltiples utilidades que pueden ser de gran ayuda para todas las personas interesadas en la didáctica de lenguas (Alonso Pérez-Ávila 2006).

Figura 1. Concordancias paralelas en CPEIC (Corpus Paralelo de Español, Inglés y Chino).[5]

que tiene mucha popularidad, se incluye el chino, solo se puede combinar con el inglés para formar par de idiomas. Es decir, en nuestro caso, no nos proporciona las concordancias paralelas chino-español.

3. Sobre las concordancias *KWIC*, es decir, *Key Words in Context* ('palabras clave en contexto'), véase Pérez Hernández (2002).

4. También denominadas concordancias bilingües.

5. *CPEIC* es un corpus paralelo trilingüe elaborado en la Universidad Nacional de Cheng Kung (NCKU) desde el año 2007 con el propósito de facilitar el análisis contrastivo entre las tres lenguas.

En segundo lugar, disponer de la información sobre las palabras más frecuentes de una lengua puede constituir una gran ventaja a la hora de elaborar materiales para la enseñanza del ELE (Cruz Piñol 2012). Asimismo, las listas de frecuencias pueden ayudar a los alumnos a elegir las estructuras lingüísticas más habituales en el uso real (Contreras Izquierdo y Martos Eliche 2020). En el caso específico que nos ocupa, si los diccionarios tienen en cuenta la frecuencia (Gargallo Gil 1992; Jiménez Ríos 2001; *DLE* 2020), el estudiante no nativo podrá identificar las traducciones más frecuentes para el contexto en cuestión.

En tercer lugar, en un corpus es fundamental la lematización, que consiste en vincular cada forma con su lema. La forma se refiere a la palabra tal como aparece en su contexto de uso, mientras que el lema es la entrada que, por convención, engloba todas las formas flexionadas de esa palabra. Si tomamos como ejemplo el paradigma verbal, la forma se refiere a cada una de las posibilidades de flexión (*hablo, hablaste, hablen...*) y el lema será el verbo en infinitivo (*hablar*). La ausencia de lematización muestra varias limitaciones al usar los corpus, especialmente cuando se trata de lenguas con mucha flexión (como, en nuestro caso, el español), porque habría que ejecutar la búsqueda de todas las formas flexionadas posibles para poder cubrir todas las ocurrencias de un lema.

2.2. Diseño y elaboración de los corpus paralelos

Conocer cómo se construye un corpus paralelo nos sirve de gran ayuda para entender mejor sus características, su naturaleza y su utilidad. Según Doval (2017), en el proceso de elaboración de un corpus paralelo, destacan tres fases principales: la compilación de textos y su preprocesamiento, la alineación de los textos y la anotación documental y lingüística. Asimismo, durante el proceso de la compilación de textos, de acuerdo con Doval (2017), hay que tener en cuenta la importancia de la cantidad y la calidad de los textos. Además, cabe subrayar la importancia del equilibrio de los datos en cuanto a la dirección de traducción.

Por lo que respecta al proceso de alineación, es una tarea crucial e imprescindible en la construcción y elaboración de un corpus paralelo (Castillo Rodríguez 2011). En palabras de Castillo Rodríguez (2009: 7), la alineación es «la reestructuración de los textos de tal forma que se pueda establecer una correspondencia entre los párrafos, oraciones y/o palabras de los textos involucrados». La alineación consta de dos tareas, una previa de segmentación de los textos y la posterior, la vinculación de esos segmentos a partir de la correspondencia entre las dos lenguas para formar bisegmentos, es decir, pares alineados de segmentos (Doval 2017).

En cuanto a los niveles de la segmentación, las unidades de alineación pueden ser párrafos, frases o palabras. Entre ellos, lo más habitual actualmente es la alineación a nivel oracional. Con respecto a la anotación, siguiendo a McEnery y Hardie (2012), un corpus contiene básicamente los siguientes tres tipos de información añadida para la investigación de los datos en el corpus: los metadatos, el marcado textual y la anotación lingüística (Kübler y Zinsmeister 2015). Más concretamente, los metadatos se refieren a la información externa relacionada con la obra original, la traducción y otros datos bibliográficos. Los metadatos nos permiten, posteriormente, recuperar información relevante del corpus. El marcado textual, en cambio, proporciona la división interna de los textos tales como capítulos o partes a fin de localizar una cadena de texto dentro de una obra. Por último, la anotación lingüística está relacionada con el análisis propiamente lingüístico de las palabras que conforman el corpus (lema, categoría gramatical o función sintáctica, principalmente).

2.3. La web como corpus paralelo

En la mayoría de las ocasiones, los corpus paralelos tienen un tamaño relativamente limitado y es habitual que las concordancias paralelas mostradas sean insuficientes para extraer conclusiones fiables. Ante este problema, una solución es utilizar también la web como corpus.

En este sentido, cabe adelantar que, de los cuatro[6] corpus que hemos seleccionado para nuestra propuesta, tres se compilan a partir de los datos de la web: *Glosbe*, *Eshelper* y *ReversoContext*. Los corpus de este tipo son fáciles de manejar y contienen gran cantidad de datos bi-/multilingües, pero también presentan algunos problemas a la hora de utilizarlos en el aula. Uno de estos consiste en «la baja calidad de los contextos encontrados» (Gelbukh *et al.* 2002: s/p), como veremos en el apartado 3.4. Por ejemplo, los textos pueden ser producidos por autores poco cultos, la alineación automática sin revisión puede causar incidencias, los textos pueden estar mal segmentados, etc.

2.4. Aplicaciones de los corpus paralelos en la enseñanza de lenguas

Los corpus paralelos se están empleando cada vez más en el ámbito de la enseñanza de lenguas, un terreno en el que este tipo de corpus presenta varias ventajas. Como sostiene Barlow (2000), utilizar los textos paralelos permite a los aprendices de lenguas investigar directamente la correspondencia entre las palabras o estructuras en dos lenguas. Por su parte, Doval (2018) ilustra muy bien las posibilidades que presenta el corpus paralelo creado por la misma autora[7] en el aula de ELE: dicho corpus pretende ser un recurso útil para los estudiantes avanzados de español o de alemán, para que dispongan de diversos contextos de uso y ejemplos de traducción.

De acuerdo con Kenning (2010), los corpus paralelos permiten a los estudiantes acceder a la evidencia contrastiva que destaca las similitudes y divergencias entre las diferentes lenguas. Además, el acceso a la traducción, tanto la completa como la parcial, posibilita que los estudiantes estén expuestos a las ocurrencias que contienen vocabulario o frases que no conocen. De ahí que se favorezca el *Aprendizaje Basado*

6. Datos correspondientes a abril de 2021. Hay que decir que previamente contemplamos también el corpus *TAUS Data Cloud*, pero aquí lo hemos descartado porque en 2021 dejó de estar accesible de forma gratuita.

7. Se trata del corpus *PaGeS*, un corpus paralelo español-alemán, disponible en https://www.corpuspages.eu/corpus/about/about?lang=en.

en Datos (Data-Driven Learning)[8] (Kenning 2010). A continuación, se verá que los estudios distinguen cuatro tipos de usos de los corpus paralelos en la enseñanza de lenguas:

a) Uso directo de las concordancias paralelas en el aula (p. ej. Barlow 2000; Olímpio de Oliveira Silva y Penadés 2016; Lu y Cheng 2016; Doval 2018):[9] el uso directo de los corpus paralelos en el aula ha sido investigado por muchos autores, especialmente en lo concerniente a los corpus que incluyen el inglés y desde la década de 1990 (Zanettin 1994; Barlow 2000; St. John 2001, etc.). Utilizando los datos bilingües inglés-francés, Barlow (2000) sugiere que se puede sensibilizar a los estudiantes sobre la manera de expresar significados equivalentes, gracias a los abundantes contextos que proporcionan. Señala que, con la ayuda de la consulta en paralelo de los pronombres reflexivos, los alumnos pueden observar que las formas reflexivas en inglés no se traducen necesariamente por una forma reflexiva en francés, e incluso para algunas formas no existen equivalentes similares en otra lengua. Por su parte, St. John (2001) plantea la utilidad de un corpus paralelo inglés-alemán con alumnos de nivel inicial. Las tareas propuestas en este caso consisten en la comparación entre la lengua meta y la lengua de origen para extraer conclusiones de esta observación: a partir de la consulta de la palabra *also* en alemán, que puede ser una partícula o un adverbio dependiendo de la sintaxis y del contexto, y que es un *falso amigo* entre estas dos lenguas, los alumnos pueden observar la función, la posición y el uso como modificador de esta palabra e intentar derivar independientemente las reglas que rigen su uso. Otro ejemplo lo encontramos en Olímpio de Oliveira Silva y Penadés (2016), quienes presentan cómo utilizar los «corpuscionarios» de tipo *Linguee*, en la enseñanza-aprendizaje de las unidades fraseológicas en el aula de ELE. Afirman que el uso de las concordancias paralelas

8. Sobre *Data-Diven Learning* (DDL), puede verse Johns (2000).

9. Véase, también, Zanettin (1994); St. John (2001); Wang (2001); Frankenberg-Garcia (2005); Cao *et. al* (2018).

portugués-inglés en el aula facilitará la observación de diferentes equivalencias en distintos contextos, la comprensión de los conceptos culturales, la reflexión sobre los errores interlingüísticos, etc. Concretamente sobre las aplicaciones de los corpus paralelos chino-español en la enseñanza del ELE a estudiantes sinohablantes, Lu y Cheng (2016: 133) apuntan que la utilización del corpus CPEIC (Figura 1) podrá «facilitar la manipulación de la lengua meta (español), a través de las experiencias y conocimientos de su primera lengua extranjera (inglés) y su lengua materna (chino)». En resumen, todos los estudios mencionados en este apartado nos han inspirado para valorar la utilidad de los corpus paralelos chino-español en el aula.

b) Uso indirecto de los corpus paralelos para la elaboración de materiales didácticos (Chang 2004; Römer 2008; Doval 2018): aunque en la propuesta de nuestro trabajo no nos vamos a centrar en el uso indirecto de los corpus paralelos, conviene describirlo para tener una visión más global sobre las aplicaciones didácticas de los corpus. En su uso indirecto, los corpus constituyen una base para el diseño y la preparación de materiales didácticos, manuales de referencia o pruebas de nivel (Doval 2018). Cabe mencionar que este tipo de uso se considera muy ventajoso en el ámbito de la enseñanza de lenguas con fines específicos. Como indica Römer (2008), un corpus paralelo, sobre todo si contiene textos de especialidad, es adecuado para servir de base para la elaboración de materiales didácticos y manuales de referencia como diccionarios bilingües, puesto que en este tipo de corpus se muestra la equivalencia y correspondencia.

c) Uso como herramienta de consulta autónoma para mejorar la expresión escrita (p. ej. Cruz Piñol 2012; Bluemel 2014; Buyse 2020):[10] Frankenberg-Garcia (2004) señala que la consulta tipo *self-access* de los corpus paralelos puede ser de gran ayuda para los

10. Véase, también, Frankenberg-Garcia (2004); Buyse y Verlinde (2013); Buyse (2014); Cruz Piñol (2015); Buyse (2020).

alumnos durante las tareas de escritura en la lengua meta, puesto que, por un lado, los corpus paralelos tipo *Linguee* ofrecen tanto el significado de las palabras desconocidas para los alumnos como los contextos donde se usan las palabras adecuadamente, y, por otro, les permite observar en ocasiones las unidades léxicas que no tienen equivalente directo en otra lengua. Debido a que la consulta en este caso la inician los alumnos mismos según su propia necesidad, este tipo de utilización favorecerá significativamente su aprendizaje de la lengua meta (siempre y cuando los alumnos sepan extraer las conclusiones relevantes de la consulta, para lo cual es recomendable que el profesor forme a los alumnos en el uso experto de estas herramientas).

d) Uso como base de la investigación en lingüística contrastiva (p. ej. Lu y Cheng 2016; Lu y Lu 2016; Lu *et al.* 2017):[11] por lo que respecta al tema de este trabajo, en los últimos años, el estudio contrastivo entre el chino y el español a partir de corpus ha suscitado el interés de algunos profesores y estudiantes de español en China. Por ejemplo, en las tres publicaciones mencionadas se destaca la utilidad del corpus *CPEIC* para facilitar varios estudios contrastivos sobre los pronombres de complemento directo e indirecto, los verbos copulativos, los tiempos y las voces gramaticales, etc.

En el contexto específico que nos ocupa, de estos cuatro usos creemos que los que más pueden favorecer el aprendizaje del ELE son el (a) y el (c), es decir, el uso directo (llevar las concordancias paralelas al aula, con la intervención del docente) y el uso como herramienta de consulta (que los alumnos consulten los corpus de forma autónoma, sobre todo en las tareas de escritura). Por eso, en el próximo apartado vamos a mostrar los corpus paralelos chino-español existentes en línea y, tras analizarlos de acuerdo con unos criterios, seleccionaremos los cuatro que nos parecen más adecuados para el uso directo y para el uso como herramienta de consulta.

11. Véase, también, Luo (2017), McEnery y Xiao (2007).

3. Corpus paralelos de nativos chino-español útiles para la enseñanza del ELE[12]

3.1. Cuatro de diez: justificación de la selección

Tras consultar algunas selecciones previas[13] y utilizar motores de búsqueda,[14] hemos localizado diez corpus paralelos de nativos chino-español, que se pueden clasificar en los siguientes tres grupos:

- Grupo A: Los repositorios de traducciones humanas que presentan el texto original y la traducción de forma alineada, y los «corpuscionarios» en línea de tipo *Linguee*. Generalmente los textos proceden de sitios web bilingües o multilingües y han pasado por un procedimiento altamente automatizado (*MyMemory*, *Glosbe*, *Eshelper*, *ReversoContext*, *OPUS*).[15]
- Grupo B: Los corpus paralelos que solo permiten a los usuarios descargar los textos alineados, pero no cuentan con un sistema de consulta. Los corpus de este grupo no muestran la hibridación entre diccionario y corpus, sino que sirven como base de datos para los corpus de grupo A (*The Holy Bible* [Resnik *et al.* 1999], *Corpus Paralelo de las Naciones Unidas* [Ziemski *et al.* 2016]).
- Grupo C: Los corpus paralelos establecidos por un grupo de investigación y diseñados para una finalidad puntual de investigación lingüística (*PaChEs* [Doval *et al.* 2019], *CPEIC* [Lu 2016], *RST Spanish-Chinese Treebank* [Cao *et al*. 2018]).

Sin embargo, para la propuesta didáctica que mostraremos enseguida, no vamos a utilizar todos los corpus mencionados arriba (once en total) porque, por una parte, el uso de algunos corpus encierra mayor

12. Cabe aclarar que en nuestro trabajo nos centramos en los corpus (paralelos) de nativos y no nos ocupamos de los corpus (paralelos) de aprendices (McEnery y Xiao 2011).

13. Por ejemplo, la página personal de Chelo Vargas Sierra (https://personal.ua.es/es/chelo-vargas/enlaces/corpus-paralelos-en-linea.html), y la página de *Tagpacker* de Mar Cruz Piñol (https://tagpacker.com/user/mar.cruz.pinol?t=corpus,paralelos%255Fy%255Fmultilingues).

14. Por ejemplo, *Google*, *Baidu*, *Bing*, etc.

15. También *Taus Data Cloud* era de este tipo, pero lo descartamos porque ya no se puede consultar en abierto (véase la nota 6).

grado de dificultad y, por otra parte, los corpus como *PaChEs* todavía están en construcción o no están en línea hasta la fecha. Por eso, en nuestro trabajo hemos decidido utilizar los cuatro corpus paralelos chino-español fácilmente manejables en línea, es decir, *Glosbe*, *Eshelper*, *ReversoContext* y *CPEIC*. Partiremos de las propuestas de Buyse (2011, 2021) como referencia para comparar las posibilidades que estos corpus presentan en el aula de ELE. Además, compararemos los corpus teniendo en cuenta las características y conceptos clave en el proceso de la construcción de los corpus, como se ha visto en el apartado 2. Es decir, analizaremos los corpus a partir de las siguientes preguntas:

1) Idiomas: ¿cuántos idiomas contiene este corpus? En cuanto al par de idiomas chino-español, ¿qué variedades incluye?
2) Propósito: ¿con qué objetivo se ha construido este corpus?
3) Tamaño: ¿cuál es el tamaño de este corpus?
4) Selección de textos: ¿qué textos se han seleccionado?
5) Lematización: ¿es un corpus lematizado?
6) Palabra clave resaltada en ambas versiones de lengua: ¿la palabra que buscamos está resaltada tanto en el texto original como en la traducción?
7) POS-tagging (*Part of Speech*, es decir, categorías gramaticales): ¿hay etiquetación gramatical?
8) Definición: ¿se muestra la definición de las palabras que buscamos?
9) Combinación de palabras: ¿se pueden hacer búsquedas de combinaciones de palabras?
10) Búsqueda filtrada: ¿se pueden filtrar los resultados según el tema, el tipo de textos, etc.?
11) Alineación: ¿cuál es el nivel de alineación de este corpus?
12) Información anotada: ¿hay información de los metadatos y el marcado textual?
13) Dirección de traducción: ¿es un corpus unidireccional, bidireccional o multidireccional?
14) Particularidad: ¿tiene algunas características destacables?

3.2. Presentación breve de los cuatro corpus seleccionados

Glosbe es un «corpuscionario» multilingüe en línea creado a partir de memorias de traducción, provenientes de traducciones hechas por humanos. La interfaz permite ver no solo la traducción de la palabra, sino también cómo se comporta en la frase. Actualmente cuenta con 375.000 frases traducidas alineadas en el diccionario chino-español. Comparado con *TAUS Data Cloud*, *Glosbe* nos proporciona más información sintáctico-semántica, por ejemplo, la etiquetación de la categoría gramatical (POS), la definición, etc.

Eshelper es un «corpuscionario» diseñado especialmente para los aprendices sinohablantes de español. Recopila 450.000 bisegmentos alineados a partir de los datos de la web. Además, permite acceder al diccionario bilingüe de términos jurídicos, económicos y de negocio disponible en *Eshelper* (*XiBanYaYuFaLuWaiMaoJingMaoCiHui*). Puesto que es un recurso diseñado especialmente para los aprendices sinohablantes, las frases recopiladas parecen más cotidianas y fáciles de comprender.

ReversoContext es un «corpuscionario» que utiliza millones de textos bilingües, procesados por complejos algoritmos elaborados con inteligencia artificial. Incluye más de 100 pares de idiomas y miles de millones de traducciones extraídas de diálogos de películas, documentos oficiales, páginas web, periódicos. En comparación con los dos «corpuscionarios» mencionados arriba, una particularidad de *ReversoContext* consiste en que se permite ver el contexto más amplio de una traducción y verificar la fuente de un ejemplo.

El *Corpus Paralelo de Español, Inglés y Chino* (*CPEIC*) fue creado por un grupo de investigación de la National Cheng Kung University con el objetivo de proporcionar datos fiables para facilitar el análisis contrastivo de tres lenguas: español, inglés y chino, que son respectivamente la lengua meta (L3), la primera lengua extranjera (L2) y la lengua materna (L1) de los estudiantes sinohablantes. Se ha recogido una gran cantidad de materiales idiomáticos (p. ej. cuentos, textos de la Biblia, documentos de las Naciones Unidas, subtítulos de películas etc.), y se continúa mejorando la construcción del corpus. A diferen-

cia de los corpus mencionados más arriba, los textos en *CPEIC* son seleccionados por expertos.

3.3. Tabla comparativa de los cuatro corpus

Para tener una visión general, a continuación (Tabla 1), mostramos la comparación de estos cuatro corpus siguiendo los criterios mencionados en el apartado 3.1.[16] Se puede observar que cada corpus tiene sus puntos a favor y en contra.[17]

Corpus / Criterios	*Glosbe*	*ReversoContext*	*Eshelper*	*CPEIC*
Multilingüe	+	+	–	+
Finalidad[17]	+	+	+	+
Selección y revisión de los textos	–	–	–	+
Lematizado	+	–	+	–
Palabra clave destacada en dos versiones de lengua	–	+	–	+
Etiquetado gramatical (POS-tagging)	+	+	+	+
Etiquetado semántico (definiciones)	+	+	+	–
Combinaciones de palabras	+	+	+	–
Búsqueda filtrada por tema, tipo de texto, o categoría gramatical	–/–/–	–/–/+	–/–/+	+/+/+
Alineado por frase	+	+	+	+
Metadatos y marcado textual	Fuente	Fuente	Título	Título
Traducción multidireccional	+	+	+	+
Otras funciones	+	+	+	–

Tabla 1. Comparación de los corpus *Glosbe*, *ReversoContext*, *Eshelper* y *CPEIC*.

16. En la tabla vamos a excluir el criterio del tamaño porque, por un lado, la información sobre el tamaño del corpus chino-español no está registrada en la página de *ReversoContext* y, por otro lado, en los otros tres corpus los editores utilizan diferentes parámetros para describir el tamaño y somos conscientes de que estos parámetros no son comparables.
17. Esta tabla se ha diseñado desde la perspectiva de la utilidad de los corpus de aprendices para la enseñanza-aprendizaje de lenguas.

3.4. Ventajas y desventajas del uso de los corpus en el aula de ELE

Tras presentar las características de los cuatro corpus seleccionados, es preciso llamar la atención sobre algunas de sus limitaciones, dado que pueden ser relevantes a la hora de usarlos en el aula de ELE. En primer lugar, conviene tener en cuenta que, excepto uno (*CPEIC*), los corpus que analizamos (*Glosbe*, *ReversoContext* y *Eshelper*) toman la web como fuente; es decir, los textos incluidos en estos corpus no han sido seleccionados ni revisados manualmente por expertos. En este sentido, no hay que olvidar que, si bien los corpus basados en los datos extraídos de la web tienen mayor tamaño, la calidad de los textos puede ser peor (Gelbukh *et al.* 2002).

En segundo lugar, se observa que entre los cuatro corpus seleccionados solo *Glosbe* y *Eshelper* están lematizados, y el nivel de lematización de *Glosbe* es muy limitado. Sin duda alguna, la ausencia de lematización limitará su utilidad en el aprendizaje-enseñanza del ELE (Cruz Piñol 2012). Por ejemplo, en el caso del aprendizaje de una colocación que incluya un verbo, un corpus que permita la búsqueda por lemas posibilitará que los alumnos dispongan de más ejemplos para observar la combinación más habitual y extraer sus propias conclusiones.

En tercer lugar, la información de anotación (la anotación lingüística, los metadatos, el marcado textual) se encuentra muy limitada en estos cuatro corpus. Por una parte, la anotación lingüística limitada restringe el nivel de análisis al que se pueda someter un corpus. Por otra parte, la ausencia de los metadatos sobre la obra, el autor, la fecha, etc. impide la recuperación de información relevante por parte de los usuarios cuando tienen dudas.

En cuarto lugar, la posibilidad de comparar entre géneros, temas, países, etc. es baja o inexistente en la mayoría de estos corpus. Esto se debe, por una parte, a que los textos que se recogen en estos corpus muestran poca diversidad en cuanto al género, medio, tema, área geográfica, etc. Por otra parte, entre estos cuatro corpus, solo hay uno (*CPEIC*) que nos permite seleccionar el tema que deseemos buscar; es decir, la mayoría de los corpus no disponen de un sistema para realizar una búsqueda filtrada.[18]

18. Esta constatación nos lleva a poner de relieve que la interacción con los

Teniendo en cuenta todo lo expuesto, consideramos que el que más se acerca al «corpus paralelo ideal» para utilizar en el aula de español sería el de *Eshelper*, por ser una herramienta especialmente diseñada para los alumnos sinohablantes de español y tener una calidad de textos relativamente buena. Sin embargo, cabe destacar que cada corpus tiene sus pros y contras. Así que es recomendable utilizar estos corpus de forma combinada, con el objetivo de tener más ejemplos variados. Además, consideramos que la utilización de los corpus basados en los datos de la web (*Glosbe*, *ReversoContext* y *Eshelper*) por parte de los alumnos de nivel inicial requiere la instrucción o supervisión del docente, ya que las frases en estos corpus no están siempre bien redactadas y traducidas. Por supuesto, los cuatro corpus pueden ser de ayuda para los alumnos de nivel avanzado para hacer consultas autónomas (de tipo *self-access*) durante las tareas de escritura, puesto que los alumnos de este nivel ya son capaces de distinguir los pares de textos que no tienen correspondencia y escoger los ejemplos que les parecen útiles. Asimismo, el corpus que cuenta con un sistema para filtrar los resultados de búsqueda (*CPEIC*) puede ser una buena herramienta a la hora de comparar un fenómeno lingüístico entre diferentes tipos de textos o áreas temáticas.

4. Propuesta didáctica

4.1. Actividades para aprender locuciones

Una de las propuestas que lanzamos en este trabajo va orientada al aprendizaje de las locuciones con los corpus paralelos. Este tipo de unidades fraseológicas constituyen un claro obstáculo para los aprendices sinohablantes de español debido a su idiomaticidad, fijación,

corpus debe avanzar hacia interfaces intuitivas adaptadas a los hábitos actuales de navegación en la red, lo cual incluiría opciones de búsqueda más visuales, como mapas interactivos (Barros 2021, Cruz Piñol 2021, Trigo 2021), en la línea de lo que se ha venido a denominar «humanidades espaciales» (Gregory y Geddes 2014; Sartor 2019). En este sentido, en los cuatro corpus seleccionados se echan en falta representaciones gráficas que aporten una dimensión espacial a la hora de visualizar los datos lingüísticos.

sentido cultural, etc. (Wu 2016). En este contexto, a nuestro juicio, los corpus paralelos chino-español pueden ser de gran utilidad para elaborar un amplio repertorio de ejercicios destinados al aprendizaje de locuciones por parte de los estudiantes sinohablantes. Se pueden utilizar para, por un lado, ayudar a los aprendices a inducir los significados idiomáticos y, por otro lado, para comprobar si el alumno ha captado sus significados y usos. Con la ayuda de los corpus paralelos, los estudiantes pueden aprender las locuciones a nivel textual, comunicativo y pragmático. A continuación, mostramos una propuesta de actividad[19] (Figura 2).

Observa estos ejemplos y las traducciones, y busca en las oraciones en chino los equivalentes de las locuciones destacadas en español. Asocia, después, las locuciones con sus equivalentes en chino.

Se me abría la boca por haber dormido poco. 因为睡眠时间少，我老打哈欠。
Cuando ve los pasteles, **se le hace la boca agua**. 他看到蛋糕的时候，就馋的流口水。
La noticia **corrió de boca en boca** y ya lo saben todo. 那消息已口口相传，现在大家都已知道此事
Te fuiste de la boca y ahora lo sabe todo. 是你说漏了嘴，现在他全知道了。

(Fuente: Eshelper y elaboración propia)

Irse de boca/ irse la boca a uno	说漏嘴
Abrirse la boca	打呵欠
Hacerse la boca agua a uno	口口相传
Andar/correr/ir de boca en boca	流口水

Figura 2. Actividad para aprender locuciones (a partir del corpus *Eshelper*).

19. Se pueden ver más propuestas en Wei (2020).

En esta actividad (duración: 10 minutos aproximadamente), los alumnos están expuestos a las concordancias paralelas extraídas del corpus *Eshelper*, que les ayudarán a establecer la correspondencia entre las dos lenguas. Esta correspondencia facilitará la comprensión y memorización de las locuciones para los alumnos sinohablantes, puesto que la mayoría de estas locuciones comparten el elemento *boca* (口/嘴) y las otras tampoco son difíciles de comprender con la ayuda del contexto y la traducción. Después, la actividad para asociar las locuciones con sus equivalentes obliga a los alumnos a repasar estas locuciones, lo cual favorecerá la retención de estas unidades léxicas.

4.2. Actividades para aprender marcadores del discurso

Otro ejemplo del uso de los corpus paralelos chino-español que proponemos en este trabajo sería para el desarrollo de la competencia discursiva. Más concretamente, vamos a centrarnos en el conector concesivo *aunque* (Figura 3). Desde nuestro punto de vista, los corpus paralelos podrían ser de gran utilidad en la enseñanza de los marcadores discursivos, puesto que en un diccionario convencional no se puede encontrar una «definición» de un marcador y una manera más eficaz para saber qué significa un marcador es verlo en su uso y contexto. Además, gracias a la observación de las concordancias paralelas, los aprendices pueden comparar las similitudes y diferencias del uso de los marcadores en dos lenguas, extraer su propia conclusión de modo inductivo y reflexionar sobre sus errores interlingüísticos causados por las interferencias de su lengua materna.

ACTIVIDAD 1

Aunque vive conmigo, no tenemos ninguna vinculación.

虽然跟我住在一起，***但是***我们没有任何交集。

(Fuente: Eshelper y elaboración propia)

1. En este ejemplo arriba, has visto que para expresar 虽然 en español, se usa el conector *aunque*. Busca en Glosbe, ReversoContext, Eshelper y CPEIC el conector *aunque* y observa los resultados. ¿Has encontrado otra traducción de este conector además de 虽然? y, ¿has observado alguna similitud o diferencia del uso de estos conectores en chino y en español?

Figura 3. Actividad para aprender el conector *aunque* (a partir del corpus *Eshelper*).

En esta actividad, empezamos con un ejemplo extraído de un corpus paralelo (en este caso, el corpus *Eshelper*). En el ejemplo, el conector ha sido traducido por 虽然. Sin embargo, el conector *aunque* puede corresponder a muchos marcadores en chino en diferentes contextos (*p. ej.* 虽然, 即使, 尽管, 纵使, etc.). Después de este ejemplo, se pide que los alumnos hagan búsquedas en los cuatro corpus. A modo de ejemplo, si un alumno busca la palabra *aunque* en el corpus *ReversoContext*, va a obtener los resultados que se muestran en la Figura 4:

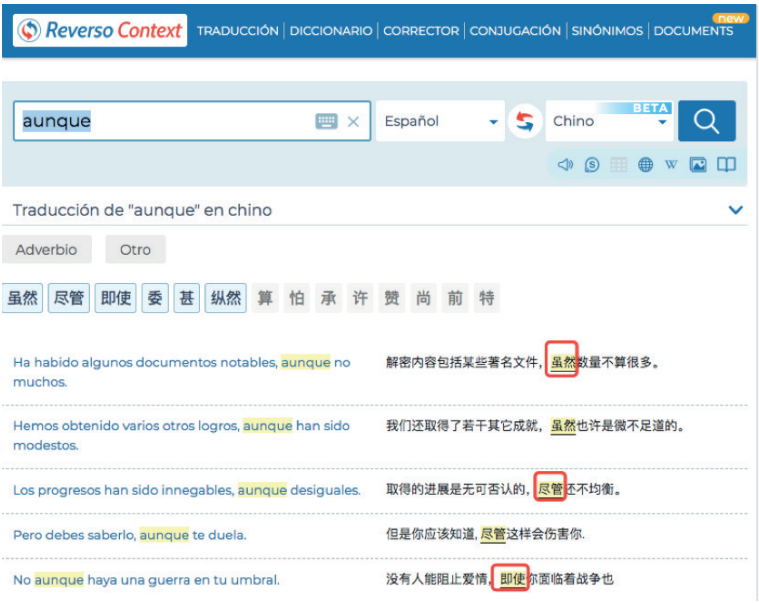


Figura 4. Búsqueda de la palabra *aunque* en *ReversoContext*.

Igual que en el estudio de Barlow (2000), tras buscar una determinada palabra y observar las concordancias paralelas repetidamente, los alumnos pueden llegar a ser conscientes de que una palabra puede corresponder a varias unidades léxicas en distintos contextos. Por ejemplo, el conector *aunque* puede corresponder a muchos marcadores en chino en diferentes contextos (p. ej. 虽然, 即使, 尽管, 纵使, etc.). Además, esta actividad, cuya duración es de unos 20 minutos, también tiene el objetivo de que los alumnos descubran las similitudes y diferencias entre el chino y el español en cuanto al uso de estos marcadores. Por ejemplo, a través del ejemplo que se ha mostrado más arriba (Figura 4: ***Aunque*** vive conmigo, no tenemos ninguna vinculación. 虽然跟我住在一起，但是我们没有任何交集。), se ha visto que, por una parte, el fragmento español, igual que el chino, ha sido dividido en dos unidades y que las primeras unidades en

dos lenguas empiezan con un marcador del discurso que muestra una relación concesiva. Por otra parte, el uso de los marcadores discursivos ha mostrado cierto contraste en dos lenguas. De acuerdo con Cao *et al.* (2016), para expresar el mismo significado, es necesario incluir dos marcadores del discurso en este enunciado en chino: 虽然 (*aunque*) al principio de la primera unidad y 但是 (*pero*) al principio de la segunda unidad. No obstante, en español solo se usa un marcador (*aunque*) al principio de la primera unidad. Tras comparar repetidamente las concordancias paralelas en dos lenguas, esperamos que los alumnos investiguen este tipo de contraste ellos mismos y consigan evitar este tipo de error interlingüístico causado por la interferencia de su lengua materna en su producción. Así, en esta actividad, los alumnos se han convertido en investigadores, lo cual aumentará su autonomía de aprendizaje.

4.3. Otras posibilidades

En las dos propuestas presentadas más arriba, hemos explorado la utilización de los corpus paralelos en la enseñanza del ELE para el desarrollo de la competencia léxica y discursiva. Se ha comprobado que, en comparación con los corpus monolingües, los corpus paralelos tienen la ventaja de mostrar la correspondencia y el contraste entre dos lenguas, como se ha visto en el apartado 2.4. Por lo tanto, con la finalidad de utilizar esta herramienta especialmente para los alumnos sinohablantes de español, sería recomendable prestar especial atención a los análisis contrastivos entre el chino y el español, tomándolos como punto de partida a la hora de diseñar las actividades. Aparte de los dos ejemplos que hemos propuesto, los corpus paralelos también pueden aportar muchas posibilidades de la enseñanza del ELE a nivel sintáctico. Así, para aprender las estructuras de la voz pasiva en español, los alumnos podrían buscar la palabra «被» (uno de los marcadores de la voz pasiva en chino) con el objetivo de obtener ejemplos en voz pasiva en español y familiarizarse con las estructuras.

Sin duda, son variados los usos de los corpus paralelos en la enseñanza de lenguas. Y es que, gracias a la manera como presentan los datos auténticos, los alumnos tienen la oportunidad de explorar, aprender e investigar ellos mismos con mayor autonomía.

5. Recapitulación

En este capítulo se ha abordado la aplicación de los corpus paralelos chino-español en el contexto específico de la enseñanza del ELE para los estudiantes sinohablantes. Se ha visto que los estudios distinguen cuatro tipos de usos de los corpus paralelos en la lingüística aplicada: el uso directo de las concordancias paralelas en el aula, el uso indirecto de los corpus paralelos para la elaboración de materiales didácticos, el uso como herramienta de consulta para mejorar la expresión escrita y el uso como base para la investigación en lingüística contrastiva. A continuación, hemos presentado unos criterios para escoger los corpus más adecuados para utilizar en el aula, que nos han permitido seleccionar cuatro corpus, que recomendamos utilizar de forma combinada, teniendo en cuenta el nivel de los alumnos y el tipo de instrucción del docente. Por último, hemos planteado dos propuestas de actividades utilizando los corpus seleccionados para aprender locuciones y conectores, y hemos dejado abiertas otras posibilidades para el desarrollo de diferentes competencias. Esperamos que esta aportación pueda servir de ayuda tanto para los investigadores como para los profesores interesados en utilizar los corpus paralelos para la enseñanza del ELE a aprendices sinohablantes.

Referencias bibliográficas

Alonso Jiménez, Elisa (2013): «*Linguee* y las nuevas formas de traducir», *Skopos: revista internacional de traducción e interpretación*, vol. 2, pp. 5-28. En línea, <https://rio.upo.es/xmlui/bitstream/handle/10433/851/2013_Alonso_Linguee_Skopos.pdf?sequence=1> [Consulta: 11/11/2020].

Alonso Pérez-Ávila, Elena (2006): «El corpus lingüístico en la didáctica del léxico en el aula de E/LE», memoria de máster publicada en *RedELE*, vol. 11, pp.

1-83. En línea, <https://www.educacionyfp.gob.es/dam/jcr:c9948929-a65a-4afc-8df2-9c104e1e6175/2010-bv-11-01alonso-pdf.pdf> [Consulta: 11/11/2020].

Alonso Ramos, Margarita (2009): «Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionario?», *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus*, Murcia, AELINCO, pp. 1191-1207. En línea, <http://www.dicesp.com/app/webroot/files/file/CILC%2009.pdf> [Consulta: 11/11/2020].

Barlow, Michael (2000): «Parallel Texts in Language Teaching», en Simon Botley *et al.* (eds.), *Multilingual Corpora in Teaching and Research*, Amsterdam, Rodopi, pp. 106-115. En línea, <https://books.google.es/books?id=eBwAj1nyRu4C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false> [Consulta: 11/11/2020].

Barros, Benamí (2021): «Representar visualmente los resultados de la investigación sobre el español LE/L2», en Mar Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 228-247.

Bluemel, Brody (2014): «Learning in Parallel: Using Parallel Corpora to Enhance Written Language Acquisition at the Beginning Level», *Dimension*, vol. 31, pp. 31-48. En línea, <https://files.eric.ed.gov/fulltext/EJ1080263.pdf> [Consulta: 11/11/2020].

Buyse, Kris (2011): «¿Qué corpus en línea utilizar para qué fines en la clase de ele?», en Javier de Santiago Guervós, Hanne Bongaerts, Jorge J. Sánchez Iglesias, Marta Seseña Gómez (eds.), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*, Salamanca, ASELE, pp. 277-288. En línea, <https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0277.pdf> [Consulta: 11/11/2020].

Buyse, Kris (2014): «Una hoja de ruta para integrar las TIC en el desarrollo de la expresión escrita recursos y resultados», *Journal of Spanish Language Teaching*, vol. 1, n. 1, pp. 101-115.

Buyse, Kris (2020): «Qué tipo de corpus para qué tipo de texto: de la teoría a la práctica», en Javier de Santiago Guervós, Lourdes Díaz (eds.), *Lingüística textual y enseñanza de español LE/L2*, London/New York, Routledge, pp. 282-306.

Buyse, Kris (2021): «Corpus textuales de nativos para investigar sobre la enseñanza/aprendizaje del español LE/L2», en Mar Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 109-136.

Buyse, Kris; Verlinde, Serge (2013): «Possible Effects of Free on Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of *Linguee* and the *Interactive Language Toolbox*», *Procedia – Social and Behavioral Sciences*, vol. 95, pp. 507-512. En línea, <https://doi.org/10.1016/j.sbspro.2013.10.675> [Consulta: 11/11/2020].

Cao, Shuyuan; da Cunha, Iria; Iruskieta, Mikel (2016): «A Corpus-based Approach for Spanish-Chinese Language Learning», *Proceedings of the 3rd Wor-*

kshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016), Osaka, The COLING 2016 Organizing Committee, pp. 97-106. En línea, <https://www.aclweb.org/anthology/W16-4913/> [Consulta: 11/11/2020].

Cao, Shuyuan; da Cunha, Iria; Iruskieta, Mikel (2018): «The RST Spanish-Chinese Treebank», *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, Santa Fe, Association for Computational Linguistics, pp. 156-166. En línea, <https://www.aclweb.org/anthology/W18-4917.pdf> [Consulta: 11/11/2020].

Castillo Rodríguez, Cristina (2009): «La elaboración de un corpus *ad hoc* paralelo bilingüe», *Tradumàtica*, n. 7. En línea, <http://www.lexytrad.es/assets/castillo2009a_0.pdf> [Consulta: 11/11/2021].

Castillo Rodríguez, Cristina (2011): «La alineación de un corpus paralelo multilingüe: propuesta de fases para la didáctica de traducción especializada inversa», *Cadernos de Tradução*, vol. 1, n. 27, pp. 117-142. En línea, <https://periodicos.ufsc.br/index.php/traducao/article/view/2175-7968.2011v1n27p117> [Consulta: 11/11/2021].

Chang, Baobao (2004): «Chinese-English Parallel Corpus Construction and its Application», *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation*, Tokyo, Waseda University, Logico-Linguistic Society of Japan, pp. 283-290. En línea, <http://aclweb.org/anthology/Y04-1030> [Consulta: 11/11/2020].

Contreras Izquierdo, Narciso; Martos Eliche, Fermín (2020): «Las secuencias formulaicas en la enseñanza del español como LE/L2. La frecuencia léxica como criterio de nivelación», *Porta Linguarum*, n. 33, pp. 111-127. En línea, <http://www.ugr.es/~portalin/articulos/PL_numero33/7_Narciso%20Contreras.pdf> [Consulta: 11/11/2021].

Cruz Piñol, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*, Madrid, Arco/Libros [2.ª edición: 2017].

Cruz Piñol, Mar (2015): «Léxico y ELE: enseñanza/aprendizaje con tecnologías», *Journal of Spanish Language Teaching*, vol. 2, n. 2, pp. 165-179.

Cruz Piñol, Mar (2021): «e-Research: fundamentos metodológicos y aplicaciones a la investigación sobre el español LE/L2», en M. Cruz Piñol (ed.), *e-Research y espanol LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 1-39.

DLE = Real Academia Española; Asociación de Academias de la Lengua Española (2020): «Advertencias», *Diccionario de la Lengua Española*, versión electrónica 23.4. En línea, <https://dle.rae.es/contenido/advertencias> [Consulta: 11/11/2020].

Doval, Irene (2017): «La construcción de un corpus paralelo bilingüe multifuncional», *Moenia*, vol. 27, pp. 125-141. En línea, <https://revistas.usc.es/index.php/moenia/article/view/4071> [Consulta: 11/11/2021].

Doval, Irene (2018): «Corpus paralelos en la enseñanza de lenguas extranjeras: un ejemplo de aplicación basado en el corpus PaGeS», *CLINA: Revista Interdisciplinaria de Traducción, Interpretación y Comunicación Intercultural*, vol. 4, n. 2, pp. 65-82. En línea, <https://doi.org/10.14201/clina2018426582> [Consulta: 11/11/2021].

Doval, Irene *et al.* (2019): *Protocolo para constituir el corpus chino-español del Proyecto PaChEs*, Santiago de Compostela / Barcelona [trabajo inédito].

Frankenberg-Garcia, Ana (2004): «Lost in Parallel Concordances», en Aston Guy *et al.* (eds.), *Corpora and language learners*, Amsterdam, John Benjamins, pp. 213-229. En línea, <https://www.researchgate.net/publication/253553681_Lost_in_parallel_concordances> [Consulta: 11/11/2020].

Frankenberg-Garcia, Ana (2005): «Pedagogical Uses of Monolingual and Parallel Concordances», *ELT Journal*, vol. 59, n. 3, pp. 189-198.

Gargallo Gil, José Enrique (1992): «La ordenación de acepciones en algunos artículos lexicográficos del DRAE: cuestiones de norma y uso», en Manuel Ariza Viguera (ed.), *Actas del II Congreso Internacional de la Historia de la Lengua Española*, Madrid, Pabellón de España, t. I, pp. 1043-1054. En línea, <http://data.cervantesvirtual.com/manifestation/230923> [Consulta: 11/11/2020].

Gelbukh, Alexander *et al.* (2002): «Corpus virtual, virtual: Un diccionario grande de contextos de palabras españolas compilado a través de Internet», en Julio Gonzalo, Anselmo Peñas, Antonio Ferrández (eds.), *Multilingual Information Access and Natural Language Processing. International Workshop at IBERAMIA-2002: VII Iberoamerican Conference on Artificial Intelligence*, Sevilla, Elsnet y Ritos, s/p. En línea, <https://nlp.cic.ipn.mx/Publications/2002/MLIA-2002-Corpus.pdf> [Consulta: 11/11/2020].

Gregory, Ian N.; Geddes, Alistair (2014): *Toward Spatial Humanities. Historical GIS and Spatial History*, Bloomington, Indiana University Press.

Hallebeek, Jos (1999): «El corpus paralelo», *Procesamiento del lenguaje natural*, n. 24, pp. 58-69. En línea, <https://core.ac.uk/download/pdf/16355934.pdf> [Consulta: 11/11/2020].

Jiménez Ríos, Enrique (2001): «Notas a propósito de la ordenación de las acepciones en el Diccionario de la Academia», *Anuario de Estudios Filológicos*, vol. 24, pp. 255-272. En línea, <https://dialnet.unirioja.es/servlet/articulo?codigo=59018> [Consulta: 11/11/2020].

Johns, Tim (2000): *Tim Johns' data-driven learning Page*. En línea, <http://archive.ecml.at/projects/voll/our_resources/graz_2002?ddrivenlrning/whatisddl/resourses/tim_ddl_learning_page.htm> [Consulta: 11/11/2020].

Kenning, Marie-Madeleine (2010): «What Are Parallel and Comparable Corpora and How Can We Use Them», en Anne O'Keeffe, Michael McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*, London/New York, Routledge, pp. 487-500.

Kübler, Sandra; Zinsmeister, Heike (2015): *Corpus Linguistics and Linguistically Annotated Corpora*, London, Bloomsbury.

Lu, Huichuan; Cheng, Anchung (2016): «Rasgos semánticos a partir de corpus y su aplicación en el aprendizaje de la estructura 'ser/estar+adjetivo' en español», *Círculo de Lingüística Aplicada a la Comunicación*, vol. 68, pp. 117-137. En línea, <https://doi.org/10.5209/clac.54524> [Consulta: 11/11/2020].

Lu, Huichuan *et al.* (2017): «Glocalized rhetoric practices on medical writing», *Círculo de lingüística aplicada a la comunicación*, vol. 69, pp. 217-236. En línea, <https://revistas.ucm.es/index.php/CLAC/article/view/55320> [Consulta: 11/11/2020].

Lu, Huichuan; Lu, Wenxiang (2016): *The Acquisition of the Spanish Object Pronouns as a Third Language*, National Cheng Kung University, informe de resultados de investigación. En línea, <http://ir.lib.ncku.edu.tw/bitstream/987654321/175889/1/1010203001-000001.pdf> [Consulta: 11/11/2020].

Luo, Xiaoqing (2017): 基于语料库的中文政治文本英西译本的语言特征差异研究, Beijing Foreign Studies University, Trabajo de Fin de Máster dirigido por Yin Jun.

McEnery, Tony; Hardie, Andrew (2012): *Corpus Linguistics: Method, theory and practice*, Cambridge, Cambridge University Press. En línea, <https://epdf.pub/corpus-linguistics-method-theory-and-practice2f8645efd2df79651f7a90fe3faf67255873.html> [Consulta: 11/11/2020].

McEnery, Tony; Xiao, Richard (2007): «Parallel and Comparable Corpora: What Are They Up to?», en Gunilla Anderman, Margaret Rogers (eds.), *Incorporating Corpora. The Linguist and the Translator*, Clevedon/Buffalo/Toronto, Multilingual Matters, pp. 18-31. En línea, <https://doi.org/10.21832/9781853599873-005> [Consulta: 11/11/2020].

McEnery, Tony; Xiao, Richard (2011): «What Corpora Can Offer in Language Teaching and Learning», en Eli Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*, New York, Routledge, vol. 2, pp. 364-380. En línea, <https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/McEnery_Xiao_teaching.PDF> [Consulta: 11/11/2020].

Olímpio De Oliveira Silva, M.ª Eugênia; Penadés, Inmaculada (2016): «*Linguee* como herramienta de enseñanza-aprendizaje de las unidades fraseológicas», en Gloria Corpus Pastor *et al.* (eds.), *Computerised and Corpus-Based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, Genève, Editions Tradulex, pp. 367-374. En línea, <http://www.tradulex.com/varia/Europhras2015.pdf> [Consulta: 11/11/2020].

Pérez Hernández, María Chantal (2002): *Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*. *ELiEs 18*, Barcelona, RedIris/UAB. En línea, <http://elies.rediris.es/elies18/> [Consulta: 11/11/2020].

Resnik, Philip *et al.* (1999): «The Bible as a Parallel Corpus: Annotating the "Book of 2000 Tongues"», *Computers and the Humanities*, vol. 33, pp. 129-153.

Römer, Ute (2008): «Corpora and Language Teaching», en Anke Lüdeling, Merja Kytö (eds.), *Corpus linguistics: an international handbook*, Berlin/New York, Walter de Gruyter, vol. 1, pp. 112-130. En línea, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.9749&rep=rep1&type=pdf> [Consulta: 11/11/2020].

Sartor, Elisa (2019): «Corpus textuales y mapas digitales: un estudio de la circulación de *Arte de los metales* (1640) de Álvaro Alonso Barba en Europa y América durante la Edad Moderna a través de las humanidades espaciales», en Matteo De Beni (ed.), *Imagen y discurso técnico-científico en español. Miradas interdisciplinarias*, Mantova, Universitas Studiorum (Colección *Pliegos Hispánicos*), pp. 197-219.

St. John, Elke (2001): «A Case for Using a Parallel Corpus and Concordancer for Beginners of a Foreign Language», *Language Learning and Technology*, vol. 5, n. 3, pp. 185-203. En línea, <https://scholarspace.manoa.hawaii.edu/bitstream/10125/44573/1/05_03_stjohn.pdf> [Consulta: 11/11/2020].

Trigo, Beatriz (2021): «Investigar en el contexto de las Humanidades Digitales y el español LE/L2», en Mar Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 24-39.

Wang, Lixun (2001): «Exploring Parallel Concordancing in English and Chinese», *Language Learning and Technology*, vol. 5, n. 3, pp. 174-184. En línea, <https://scholarspace.manoa.hawaii.edu/bitstream/10125/44572/1/05_03_wang.pdf> [Consulta: 11/11/2020].

Wei, Weiqian (2020): *Propuesta de aplicación didáctica de los corpus paralelos en el aula de ELE para los estudiantes sinohablantes,* Universitat de Barcelona, Trabajo de Fin de Máster dirigido por Mar Cruz Piñol. En línea, <http://hdl.handle.net/2445/172281> [Consulta: 11/11/2020].

Wu, Fan (2016): «Fraseología comparada del español y del chino: su aplicación a la enseñanza en la clase de español como lengua extranjera», *marcoELE, Revista de Didáctica Español Lengua Extranjera*, n. 22. En línea, <https://www.redalyc.org/pdf/921/Resumenes/Resumen_92153185011_1.pdf> [Consulta: 11/11/2020].

Zanettin, Federico (1994): «Parallel Words: Designing a Bilingual Database for Translation Activities», *Corpora in Language Education and Reserch: A Selection of Papers from Talc*, vol. 94, n. 4, pp. 99-111. En línea, <https://www.academia.edu/3490875/Parallel_Words_Designing_a_Bilingual_Database_for_Translation_Activities> [Consulta: 11/11/2020].

Ziemski, Michal *et al.* (2016): *The United Nations Parallel Corpus*, Language Resources and Evaluation (LREC'16), Portorož. En línea, <https://conferences.unite.un.org/UNCORPUS/es> [Consulta: 11/11/2020].

Corpus

Corpus Paralelo de las Naciones Unidas. En línea, <https://conferences.unite.un.org/uncorpus/> [Consulta: 11/11/2020].

Corpus Paralelo de Español Inglés y Chino (CPEIC). En línea, <http://140.116.245.146/cpeic/TriLin.html?role=teacher> [Consulta: 11/11/2020].

Eshelper. En línea, <https://www.esdict.cn/> [Consulta: 11/11/2020].

Glosbe. En línea, <https://es.glosbe.com/> [Consulta: 11/11/2020].

Linguee. En línea, <https://www.linguee.com/> [Consulta: 11/11/2020].

MyMemory. En línea, <https://mymemory.translated.net/> [Consulta: 11/11/2020].

OPUS. En línea, <http://opus.nlpl.eu/> [Consulta: 11/11/2020].

PaGeS. En línea, <https://www.corpuspages.eu/corpus/about/about?lang=en>. [Consulta: 11/11/2020].

Reverso. En línea, <https://diccionario.reverso.net/> [Consulta: 11/11/2020].

ReversoContext. En línea, <https://context.reverso.net/traduccion/> [Consulta: 11/11/2020].

RST Spanish-Chinese Treebank. En línea, <http://ixa2.si.ehu.es/rst/zh/index.php> [Consulta: 11/11/2020].

The Holy Bible. En línea, <http://users.umiacs.umd.edu/~resnik/parallel/bible.html> [Consulta: 11/11/2020].

La obtención de datos culturales mediante el uso de la web como corpus para la elaboración de productos lexicográficos de ELE*

Antoni Nomdedeu-Rull
Universitat Rovira i Virgili

1. Introducción

Los principales y más recientes avances técnicos y metodológicos sobre la lexicografía tienen que ver con las diferentes herramientas de análisis y de compilación de datos usadas para elaborar productos lexicográficos. Entre las últimas preocupaciones, hallamos cuestiones como la selección de los lemas, la selección de los elementos del significado de los lemas o el tratamiento de los datos.

Desde hace varias décadas, los diccionarios generales de calidad suelen compilarse a partir de datos procedentes de corpus.[1] Sin embargo, en el ámbito del español, la elaboración de productos lexicográficos basados en corpus no ha supuesto que las palabras o expresiones incorporaran de manera explícita, sistemática ni sistematizada datos culturales. Sin duda, los aprendices no nativos de lenguas necesitan adquirir información cultural cuando leen o redactan textos para poder entender lo que leen o escuchan o para poder expresarse cuando escriben o hablan sin equívocos ni malentendidos, por lo que la ausencia de datos culturales en los diccionarios de español no permite que dichos aprendices puedan satisfacer mediante su consulta sus necesidades relacionadas normalmente con funciones comunicativas o cognitivas. De ahí que uno de los retos para la lexicografía pedagó-

* Agradezco al Dr. Pedro A. Fuertes-Olivera los valiosos datos proporcionados de los *Diccionarios Valladolid-UVa* para poder realizar este artículo.

1. Por *corpus* se entiende, en términos de Parodi (2008: 106), «una colección o conjunto de textos que está formado por al menos dos o más textos (dicho de otro modo, corpus aquí sería algo así como corpus textual)». Sobre las diferentes aportaciones a la lingüística de corpus, cfr. Villayandre Llamazares (2010).

gica sea la selección y, sobre todo, la representación de datos culturales en diferentes tipos de productos lexicográficos. En Nomdedeu-Rull (2020) se ha presentado una propuesta sobre cómo representarlos en el *Write Assistant* de Ordbogen, publicado en 2019 y todavía en fase de desarrollo. La materialización de dicha propuesta dependerá de cómo los futuros productos lexicográficos afronten las tres acciones siguientes: 1) personalización del producto lexicográfico, 2) integración del producto lexicográfico en otras herramientas (p. ej. asistentes de escritura) y 3) desarrollo del producto como herramienta de aprendizaje y no solo de consulta.

En este artículo, no se aborda cómo representar los datos culturales en diferentes tipos de productos lexicográficos, sino que se persigue ejemplificar la viabilidad de seleccionar datos culturales directamente de Internet con el fin de elaborar productos lexicográficos destinados a aprendices de español como lengua extranjera (ELE) sin la necesidad de tener que crear corpus *ad hoc*. Para ello, en el segundo apartado se repasan las principales aportaciones de la elaboración de diccionarios por medio de corpus, los diferentes tipos y extensión de los corpus con fines lexicográficos, las principales tecnologías aplicadas y la discusión sobre si es preferible el uso de corpus creados *ad hoc* o el uso de Internet directamente como corpus. En el tercer apartado, se presenta una breve crítica acerca de cómo los principales diccionarios para aprendices no nativos de español no representan suficientes datos culturales para satisfacer las necesidades comunicativas y cognitivas del aprendiz al que se dirigen. En el cuarto apartado, se presenta una propuesta de selección de elementos culturales por lema basada en el método de trabajo lexicográfico de la *Teoría funcional de la lexicografía* (TFL) que actualmente se está aplicando en los *Diccionarios en línea del español Valladolid-Uva*. Con este método de trabajo, se demuestra que la elaboración de productos lexicográficos de diferente tipo se puede realizar en relativamente poco tiempo e invirtiendo muchos menos recursos que mediante la elaboración de un corpus textual *ad hoc*.

2. Sobre la elaboración de diccionarios por medio de corpus

Un ejemplo que ilustra la metodología de elaboración de diccionarios con datos procedentes de corpus lo constituye el *Cambridge Learner's Dictionary*, con frases de ejemplo extraídas del *Cambridge English Corpus*. Los corpus permiten, como ha señalado Rojo (2009: 6), construir el diccionario sobre los datos que proceden de ellos, como, por ejemplo: palabras y acepciones realmente usadas; detección de los significados asociados a ellas; identificación de las características sintácticas del uso de cada palabra en sus acepciones; ejemplos reales del uso de cada palabra en sus acepciones. En otras palabras, los corpus permiten al lexicógrafo acceder a un número importante de datos con relativa rapidez gracias a las herramientas actuales que posibilitan obtener resultados sobre, por ejemplo, una palabra, frase, patrón gramatical o colocación en particular. Sin duda, dichos datos son de gran valor para la investigación lingüística y para la realización de diferentes tareas relacionadas con la compilación de productos lexicográficos.

En la década de 1980, la aparición de los proyectos lexicográficos basados en corpus supuso un cambio radical con respecto a la lexicografía anterior en la concepción del trabajo lexicográfico (cfr. Rojo 2009). La aparición en 1987 del *Collins Cobuild English Language Dictionary* basado en el *Bank of English* supervisado por Sinclair (1987) es el ejemplo más claro. Esta nueva base documental fue acompañada de otros cambios en el modo de concebir el diccionario: 1) se simplificó el diseño y la tipografía de los diccionarios existentes hasta ese momento al no usar abreviaturas y símbolos; 2) se acompañó al lema de explicaciones informales, con patrones sintácticos, fraseología, pragmática, ejemplos y colocaciones según lo revelado por los datos del corpus. Lamentablemente, la lexicografía pedagógica del español no ha incorporado gran parte de estas aportaciones. Esta manera de trabajar redujo parcialmente la subjetividad del proceso de selección de datos, como, por ejemplo, el que se siguió en la primera edición del *Oxford English Dictionary*, publicada por entregas entre 1884 y 1928, en donde cientos de colaboradores leían textos (escritos)

de diferentes épocas y seleccionaban ejemplos que enviaban al equipo redactor a partir de las indicaciones del equipo central y también según su propio criterio personal.

La consideración de la frecuencia (alta o baja) de las palabras como elemento central del proceso de selección tanto de lemas como de significados ha sido central en el trabajo con corpus. Sin embargo, representa algunos problemas. Según nuestra experiencia con trabajo en corpus con fines lexicográficos, la presencia de palabras con mucha frecuencia hace inmanejable la cantidad enorme de ejemplos existentes por parte de los redactores del diccionario y, por tanto, ralentiza excesivamente su trabajo, mientras que la existencia de palabras con frecuencia muy baja obliga a la ampliación del corpus, lo que también provoca que el proyecto se demore en exceso. Entre una acción y otra, el trabajo lexicográfico se convierte en interminable, como así lo demuestran los numerosos proyectos de corpus con fines lexicográficos de ELE en español inacabados. Sin embargo, desde el punto de vista lexicográfico, el problema real es otro:

> hay que distinguir entre *frecuencia lingüística* (de corpus) y *frecuencia lexicográfica* (de diccionario). Es ante todo el último tipo de frecuencia que es relevante para un nuevo proyecto de diccionario que está destinado a satisfacer necesidades reales y que, además, tiene previsto publicarse en varias fases. (Tarp 2018b: 64)

No existe correspondencia total entre la frecuencia de una palabra en Internet (y/o en un corpus) y la frecuencia con la que los usuarios buscan esta palabra en un diccionario.

Hoy en día, los listados con las palabras más buscadas pueden fácilmente elaborarse a partir de los *log files* de diccionarios que ya han sido publicados en línea. Como ha explicado Tarp (2018b: 63), Bergenholtz y Norddahl (2012) descubrieron, en un estudio basado en *log files*, que

> solo el 33 por ciento de los más de 110.000 lemas contenidos en los *Diccionarios en Línea de Danés* habían sido consultados por lo menos una vez varios años después de su publicación a pesar de que el número total de consultas en ese

momento llegara a casi 20 millones. (Es una tendencia que se muestra estable hasta hoy cuando el número de consultas es más del doble).

Es una solución posible el uso de *log files*, aunque no la única, pues puede haber términos que no aún no hayan sido buscados. Los *Diccionarios en Línea de Español Valladolid-UVa* consiguieron en 2016 un lemario compuesto por las 30.000 palabras más buscadas en otros diccionarios en línea de español. El proceso se llevó a cabo a partir del análisis de 60 millones de consultas en diccionarios monolingües y bilingües realizadas durante dos meses en los diccionarios de español que la empresa danesa Ordbogen (https://www.ordbogen.com/en/) tiene en su portal lexicográfico. Se consideró que un lema era «frecuente» cuando el mismo (o algunas de sus formas) había sido buscado en, al menos, el 80 % de los diccionarios del portal que están relacionados con el español (Fuertes-Olivera 2019).

Los diferentes tipos de productos lexicográficos basados en corpus han partido de corpus de diferente tipo y extensión. Los distintos tipos de corpus se pueden clasificar de una manera u otra de acuerdo con diferentes parámetros: la modalidad de la lengua (corpus textuales u orales), el número de lenguas a que pertenecen los textos (monolingües, bilingües o multilingües, etc.), el tamaño o cantidad de textos que conforman el corpus, la variedad lingüística o el grado de especialización de los textos (corpus generales o especializados), el periodo temporal que abarcan los textos (sincrónicos o diacrónicos) o el tratamiento aplicado al corpus. Por ejemplo, por lo que respecta a los corpus con fines lexicográficos, la editorial Macmillan cuenta con uno general y tres especializados. El corpus general incluye una amplia variedad de textos informativos e imaginativos, que van desde libros y revistas académicas hasta novelas populares y literarias y periódicos nacionales y locales. En estos momentos[2] contiene casi 1600 millones de palabras de inglés hablado y escrito. Los corpus especializados son: 1) el *Macmillan Curriculum Corpus*, usado para producir el *Macmillan School Dictionary* y el *Macmillan Study Dictionary* es una base de

2. Consulta realizada el 16/09/2020.

datos de 20 millones de palabras compuesta por cientos de libros de texto escolares y programas de exámenes, que cubren materias escolares desde agricultura hasta zoología; 2) un corpus de ciencias ambientales de 60 millones de palabras: el primero de una serie planificada de nuevos corpus para dominios específicos; 3) el *Learner Corpora* creado por el Centre for English Corpus Linguistics (CECL) en la Université catholique de Louvain-la-Neuve en Bélgica.[3]

Existen corpus especializados en ámbitos temáticos, como el fútbol o la contabilidad, entre otros, que han dado lugar a diccionarios de diferente tipo: el *Diccionario de fútbol* de Nomdedeu-Rull (2009), un diccionario dirigido al gran público con el objetivo de ayudar a la comprensión de textos futbolísticos; o el *Diccionario Inglés-Español de Contabilidad: Traducción* y el *Diccionario Inglés-Español de Contabilidad: Traducción de Frases y Expresiones* de Fuertes-Olivera, ambos son diccionarios construidos para solucionar las necesidades específicas de los traductores de textos de contabilidad. En el *Diccionario de fútbol* se empleó un método híbrido: se constituyó un corpus textual y se usó Internet directamente como corpus. En los diccionarios de Fuertes-Olivera, se usó Internet directamente como corpus.

Hay diccionarios basados en corpus con características especiales, lo que representa una excepción en el panorama lexicográfico internacional. Un ejemplo de ello lo constituye el *Diccionario Básico Escolar* dirigido por Eloína Miyares (2005). Este diccionario, del que se han publicado cuatro versiones,[4] está basado en un corpus de textos elaborados por los propios usuarios finales de dicho diccionario. Se confeccionó con el propósito de solucionar las dificultades de los estudiantes de «segundo ciclo» de la educación primaria, secundaria y preuniversitaria de Cuba, desde los 10 hasta los 18 años. Varios investigadores del Centro de Lingüística Aplicada de Santiago de Cuba realizaron entre 1990 y 1995 una investigación sobre el vocabulario activo-funcional del escolar cubano para conocer el léxico de los es-

3. Cfr. http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/.
4. La 1.ª de julio de 2005, la 2.ª en mayo de 2008, la 3.ª de noviembre de 2009 y la 4ª. de octubre de 2013. Me consta que la 5.ª está en marcha.

tudiantes cubanos. El corpus estuvo compuesto por un total de 7001 composiciones (de las cuales 5873 fueron escritas y 1128 orales) que dieron lugar a 700.000 palabras, que facilitaron la compilación del *Diccionario Básico Escolar* y de *Diccionario Escolar Ilustrado*.[5]

En las últimas décadas, dichos corpus han sido sometidos a diferentes tecnologías, que han facilitado que sean cada vez de mayor tamaño. Constituyen ejemplos claros de ello, para el inglés, el *Oxford English Corpus* https://www.sketchengine.eu/oxford-english-corpus/ o el *World English Corpus* de Macmillan http://www.macmillandictionaries.com/Terms/world-english-corpus/ y, para el español, los diferentes corpus de la Real Academia Española (*Corpus de Referencia del Español Actual* http://corpus.rae.es/creanet.html y *Corpus Diacrónico del Español* http://corpus.rae.es/cordenet.html) o el *Corpus del español* de Mark Davies https://www.corpusdelespanol.org/. A raíz de la existencia de corpus como los mencionados, se ha hablado de *big data*,[6] esto es, datos masivos.

La extensión de los corpus ha centrado buena parte del debate sobre la compilación de corpus con fines lexicográficos. Se observa una evolución de los conjuntos compuestos por un millón de formas a corpus constituidos por miles de millones de formas y, posteriormente, a la línea de trabajo *Web as Corpus*, esto es, el uso de todo lo que está contenido en la red. No obstante, ningún corpus, por grande que sea, puede compararse con el corpus que representa Internet. De ahí que se hayan desarrollado métodos enfocados a aprovechar el uso de Internet directamente como corpus y crear *smart data*, esto es, datos adaptados a las necesidades de los usuarios en lugar de datos sin procesar. Como afirma Tarp (2017: 59), ello

5. Los resultados del análisis de este corpus permitieron determinar el creciente —y cambiante— vocabulario activo de los estudiantes rurales y urbanos en cinco grados (de 8 a 12 años de edad) y sirvieron para conocer las dificultades de los estudiantes a la hora de producir textos, escritos y orales. Para todos los detalles, véase Miyares Bermúdez (2006), obra en la que se publicaron los resultados de tal estudio.

6. En un sentido estricto, el término *big data* «se refiere a los grandes conjuntos de información que por sus características no pueden ser obtenidos, gestionados ni procesados por herramientas tradicionales en un período de tiempo razonable» (González-Fernández 2016: 92).

constituye sin duda ninguna un desafío cada vez más relevante para la lexicografía, no solo por el aumento significativo de datos empíricos sino también por el tiempo que se ahorra cuando ya no hay que componer corpus específicos para los respectivos proyectos lexicográficos.

Así, en los últimos años la discusión se ha centrado en si es mejor crear corpus *ad hoc* o lo es usar la web como corpus. Varios son los autores que consideran que Internet es un tipo de corpus lexicográfico (Fuertes-Olivera 2012: 51; Kilgarrif y Grefenstette 2003: 334). Hay dos maneras de usarlo: 1) componiendo un corpus a partir de los textos encontrados en Internet; 2) usando Internet directamente como corpus. Para las ventajas y desventajas del uso de Internet directamente como corpus, Tarp y Fuertes-Olivera (2016: 277-278) y Tarp (2018b) se han referido a ello (*vid.* Tabla 1):

Ventajas	Desventajas
Los lexicógrafos pueden acceder muchos más textos que los que están incluidos en cualquier corpus de textos seleccionados.	No se puede controlar la cualidad y origen de los textos.
Los textos siempre están actualizados.	Puede que los autores de algunos textos no sean seres humanos reales
Se ahorran tiempo y dinero a no tener que componer un corpus separado.	Los autores pueden tener una baja competencia en la lengua en cuestión.
El proceso de búsqueda puede limitarse a regiones geográficas específicas, algo que es especialmente relevante para una lengua multinacional como el español.	Puede que los textos no hayan sido revisados y corregidos.
El uso del Internet puede conducir a la identificación y selección de más elementos de significado que los que pueden encontrarse en un corpus separado.	Es difícil calcular la frecuencia de los fenómenos lingüísticos que aparecen en el texto.

Tabla 1. Ventajas y desventajas del uso de Internet directamente como corpus.

Los diferentes *Diccionarios Valladolid-UVA* dirigidos por Fuertes-Olivera (Tarp y Fuertes-Olivera 2016) constituyen un claro ejem-

plo de elaboración de diccionarios por medio del uso de Internet directamente como corpus. Se hallan bajo el portal lexicográfico con objetivos comerciales denominado *Diccionarios Valladolid-Uva* —financiado y diseñado por varios investigadores de las universidades de Valladolid y Aarhus y de la empresa danesa Odbogen A/S— que incluirá diccionarios generales monolingües de español, diccionarios especializados inglés-español/español-inglés y diccionarios generales bilingües español-inglés/inglés-español.[7] Dicho portal

> está pensado para ofrecer a usuarios humanos y a usuarios máquinas herramientas de información (diccionarios de internet) gracias a las cuales puedan satisfacer sus necesidades de información de la forma más rápida y mejor posible y con el menor coste (en tiempo y en dinero) posible. (Fuertes-Olivera y Bergenholtz 2018: 171)

Actualmente, los escasos diccionarios para aprendices de ELE que se están elaborando crean un corpus *ad hoc*. El hecho de que existan varios proyectos de diccionarios de este tipo inacabados demuestra la necesidad de evolucionar en el método de compilación de diccionarios basado en corpus. Un ejemplo de ello lo demuestra el *Diccionario de español para aprendices sinohablantes* llevado a cabo por el grupo de investigación en Lexicografía y Enseñanza de Lenguas Extranjeras (LEXELE) de la Universitat Rovira i Virgili que ha necesitado más de tres años para elaborar un lemario aún no terminado. En la tradición lexicográfica hispánica, hay varios proyectos inacabados. En este contexto, el reto actual al respecto es desarrollar métodos que permitan aprovechar el corpus de Internet para crear *smart data* adaptados a las necesidades de los usuarios, confeccionarlos en un tiempo breve e invirtiendo pocos recursos económicos y humanos.

El debate sobre la selección de métodos de uso de corpus es relevante desde el punto de vista científico, aunque también lo es por las enormes consecuencias prácticas y económicas que conlleva. Como se ha analizado en varias investigaciones, la lexicografía del español, en particular la pedagógica (para nativos y para no nativos de español), vive desde hace

7. Se espera que parte de estos diccionarios estén operativos en 2021.

dos décadas en plena crisis, tanto productiva como económica (cfr. Nomdedeu-Rull 2018, 2020, 2022; Nomdedeu-Rull y Tarp 2018; Nomdedeu-Rull y Barcroft 2022 en prensa, entre otros). Ello puede resolverse integrando métodos eficaces de compilación de datos, tecnologías fáciles de usar y lexicógrafos motivados y entrenados (Tarp 2017: 71). Como se está demostrando con los *Diccionarios Valladolid-UVa*, el uso de Internet directamente como corpus aumenta la productividad en el proceso de compilación de diccionarios sin comprometer su calidad, al mismo tiempo que representa un buen ejemplo para hacer frente a la crisis mencionada.

En la lexicografía hispánica puede ser útil trabajar con este método de selección de datos, pero para ello es necesario un cambio de paradigma. Dicho cambio tiene que producirse en el terreno metodológico de compilación, producción y presentación de productos lexicográficos. Pero dicho cambio servirá de poco si la concepción de los productos lexicográficos en español no evoluciona hacia una lexicografía concebida como disciplina con un núcleo propio y con vocación interdisciplinaria por las consecuencias que conlleva en la selección y presentación de los datos, más allá de todas las implicaciones propias del proceso de concepción y compilación del producto lexicográfico. Para profundizar sobre este concepto de lexicografía, véase Tarp (2018a). El caso más claro de ello lo representa la TFL, una construcción teórica iniciada en la década de 1990 en el *Centre for Lexicography* de la Universidad de Aarhus, que presenta la lexicografía como disciplina independiente y centrada en el desarrollo de principios que puedan trasladarse a orientar el diseño y construcción de diccionarios. Como explica Fuertes-Olivera (2012: 11):

> Deben ser diccionarios que puedan satisfacer las necesidades de un tipo específico de usuario con problemas específicos que se presentan en una situación de uso específica. En este marco la teoría funcional cambia el foco de atención de los usuarios reales del diccionario a los diccionarios potenciales del mismo y a la situación social en la que se encuentran.

Ello adquiere mucha relevancia cuando abordamos el tratamiento de los datos culturales en los productos lexicográficos. La lexicografía

debe tratarse como una disciplina de información (cfr. Tarp 2018a). De acuerdo con este principio, los productos lexicográficos deben presentar datos culturales en los casos necesarios para que los usuarios puedan aprender a usar las palabras en los contextos en los que se emplean. De poco servirá aprovechar las ventajas del uso de Internet directamente como corpus para crear *smart data* adaptados a las necesidades de los usuarios si se omite buena parte de estas necesidades, concretamente las relacionadas con todos tipos de datos culturales asociados a un lema.

3. La representación de los datos culturales en productos lexicográficos de ELE. Crítica de los diccionarios existentes

Cuando nos referimos a la introducción de datos culturales en diferentes tipos de productos lexicográficos, aludimos a aquellas palabras o expresiones que conllevan diferencias culturales entre la lengua de partida y la lengua meta y cuyo conocimiento es esencial para entender su sentido completo. Nida (1945) fue el primero en abordarlas. Más adelante, otros autores comenzaron a interesarse por ellas: Newmark (1988) se refirió a ellas como *cultural words* o Nord (1997) las denominó *términos culturales* o *culturemas*, algunos años después de que Vermeer (1983) creara el término, *i.e.*, palabras que identifican las nociones culturales específicas de un país o de un ámbito cultural, que no cuentan con un paralelo exacto en otras lenguas y solo tienen sentido dentro del estilo de vida de la sociedad que las usa. Casos prototípicos de culturemas en español serían *olé*, *siesta*, *llegar y besar el santo*, *las cuentas del Gran Capitán*, *Casanova*, *huevo de Colón* o *donjuán*. Hay muchas palabras o expresiones cuyo desconocimiento de los rasgos culturales que las caracterizan pueden conducir a importantes malentendidos: se trata de expresiones con valor idiomático, ya que el hablante no se refiere directamente a los elementos intralingüísticos que figuran en la oración, como ¡Mucha mierda! ('mucha suerte', en inglés *break a leg!*) (cfr. Nomdedeu-Rull 2020).

Zgusta (1989: 3), quien se interesó por la representación de los datos culturales en los productos lexicográficos, expuso que

> since language is embedded in culture, cultural data are important to the learner not only for steering his linguistic behavior but frequently for choosing the correct lexical equivalent. Such cultural information can be understood in a broad way, so that it can pertain to political and administrative realities of the country or countries whose language is being learned, and so on. Undoubtedly, a good part of this information is of encyclopedic character; be this as it may, it belongs to what the learner has to learn.

Pero, además este tema ha sido abordado por numerosos lexicógrafos y estudiosos de la lexicografía. Entre otros, Stark (1999) lo ha considerado desde los diccionarios enciclopédicos para aprendices; Welker (2008: 382) desde la lexicografía pedagógica; Cablitz (2011) desde el conocimiento cultural en diccionarios de lenguas en peligro de extinción; Bergenholtz y Nielsen (2013) desde la lexicografía en general; Klosa (2015) desde las ilustraciones en los diccionarios.

Sin embargo, a pesar de opiniones autorizadas como las anteriores, como ha analizado Nomdedeu-Rull (2018, 2020), la presencia de la dimensión cultural de las palabras en los diccionarios es anecdótica en algunas tradiciones lexicográficas, como la del español, y no se le ha prestado atención de un modo explícito, sistemático ni sistematizado. En los diccionarios del inglés (*learner's dictionaries*) —concretamente en los denominados *big five*— la situación es mejor. Ambos tipos de diccionarios presentan algunas características similares (cfr. Nomdedeu-Rull y Barcroft 2022 en prensa).

El hecho de que los datos culturales se hayan quedado fuera de los diccionarios para aprendices de ELE puede explicarse por dos motivos. En primer lugar, porque el tratamiento de los datos culturales en los diccionarios se ha solido plantear como un problema de traducción. Por esta razón, el tipo de diccionario en donde mayormente se ha tratado ha sido el bilingüe, ofreciendo soluciones en forma de traducciones equivalentes, aunque ello no quiere decir que los productos bilingües muestren o puedan mostrar más datos culturales que los monolingües (cfr. Szerszunowicz 2011). El hecho de que su inclusión en los diccionarios monolingües no haya merecido la atención por parte de

los lexicógrafos no permite que se puedan satisfacer las exigencias y necesidades de los actuales aprendices no nativos de lenguas. Los datos culturales en ellos complementarían la equivalencia de traducción o la referencia a otra lengua y cultura que ofrecen algunos productos bilingües, sobre todo porque la equivalencia léxica es parcial e incompleta o imperfecta, como así lo expresa la idea de *anisomorfismo* apuntada por Zgusta (1971: 294-297) aplicada a los diccionarios.

En segundo lugar, se puede explicar por la perspectiva adoptada en la tradición lexicográfica hispánica según la cual el diccionario monolingüe de lengua debe contener únicamente datos lingüísticos, ya que los datos culturales se consideran enciclopédicos (cfr. Sánchez-Pérez 2010: 107). La separación entre diccionarios de lengua y enciclopedias es «ill-conceived, irrelevant and potentially confusing» (Fuertes-Olivera 2013: 1069). Luego, la distinción históricamente realizada entre *definición lexicográfica* y *definición enciclopédica* (cfr., p.ej., Bosque 1982, Werner 1982, Rey 1988 o Lara 1990) no es relevante desde la perspectiva de la lexicografía orientada al usuario, pues las clasificaciones lexicográficas deberían basarse en las funciones del diccionario, en donde los usuarios y sus necesidades constituyen el foco principal de atención. Ello es propio de una teoría lexicográfica orientada al usuario como lo es la TFL, uno de cuyos enunciados axiomáticos fundamentales es que:

> Las necesidades de información de los usuarios se resuelven dando acceso a datos lexicográficos explícitamente preparados, de los cuales los usuarios pueden extraer la información requerida que luego pueden emplear para un sinfín de propósitos relacionados con las situaciones extra-lexicográficas en las que las necesidades originalmente se produjeron. (Tarp 2015: 35)

A este respecto, comparto con Tarp (2000: 204) que «the important concern is whether or not the definition or explanation meets the user's needs and not whether it can be analysed in different parts». Lo que se ha venido llamando incorrectamente *diccionario de lengua* es un *diccionario cultural*, motivo por el cual no considero que distinguir entre ambos tipos de diccionarios sea necesario ni productivo.

Es posible que el usuario pueda entender sin mayores problemas, con la información que recibe de estos diccionarios, el sentido de culturemas como los antes mencionados, pero es muy probable que sienta la necesidad de ampliar su conocimiento con datos alusivos al porqué de estas expresiones. Los pocos datos culturales que contienen los diccionarios de ELE existentes se suelen introducir en la definición y a veces, implícitamente, en los ejemplos. En ningún caso se incluyen en notas explicativas, que son atípicas en la tradición lexicográfica hispánica. Además, si planteamos la introducción de datos culturales en diccionarios del español desde una perspectiva panhispánica, las carencias son mucho mayores, al tratarse de diccionarios elaborados desde la visión castellanocentrista del español. Esta visión supremacista sobre la lengua no es exclusiva del español. En los *learner's dictionaries* también predomina la visión de la cultura británica sobre las demás comunidades que tienen al inglés como lengua materna (cfr. Algeo 1995), como EE. UU. o Australia.

Con todo, lo que está claro es que, como se demuestra en Fuertes-Olivera y Bergenholtz (2018), Nomdedeu-Rull (2018 y 2020) o en Nomdedeu-Rull y Tarp (2018), los datos lingüísticos que ofrecen los diccionarios de ELE no son suficientes para satisfacer las necesidades comunicativas del aprendiz no nativo de español, en la misma línea que Tseng (2003) y Miller (2018) en referencia a los diccionarios para aprendices de inglés. En Fuertes-Olivera y Bergenholtz (2018), después de analizar la palabra *esvástica* en diferentes diccionarios, se afirma que «los diccionarios de español dan datos veraces de la palabra *esvástica*, pero los datos que dan no son correctos por razones diversas y por lo tanto no sirven al fin primario del diccionario en bastantes situaciones de uso» (Fuertes-Olivera y Bergenholtz 2018: 174). Ninguno de los diccionarios analizados permite que el usuario pueda saber qué religión utiliza este símbolo, qué significa en tal o cual religión, no puede saber el valor cultural de la esvástica, etc. En Nomdedeu-Rull (2018) se analizan diez expresiones propias de la cultura religiosa española (*acabar como el rosario de la aurora*,

colgar a alguien el sambenito, *estar como unas pascuas*, *estar hecho un cristo*, *hacer algo en un santiamén*, *llorar como una Magdalena*, *que Dios nos coja confesados*, *ser más falso que Judas*, *tener más años que Matusalén*, *venderse por un plato de lentejas*) y diez términos simples relativos a ámbitos típicos de la cultura española, conocidos internacionalmente (*sanfermines*, *procesión*, *fallas*, *tapa*, *gazpacho*, *sangría, paella*, *aguinaldo*, *donjuán*, *olé*). Se concluye que el tratamiento de los datos culturales ofrecido por los diccionarios monolingües para aprendices de español como LE/L2 existentes son insuficientes y no satisfacen las necesidades comunicativas y cognitivas que dichos aprendices puedan presentar. En Nomdedeu-Rull (2020) se analizan las expresiones *hat-trick* y ¡Mucha mierda!. En ambos casos, se concluye que ni los diccionarios de ELE ni los *learner's dictionaries* del inglés se suelen caracterizar por introducir de manera explícita, sistemática ni sistematizada datos culturales de las palabras o expresiones que incorporan, por lo que dichos aprendices no puedan satisfacer mediante su consulta sus necesidades relacionadas normalmente con funciones comunicativas o cognitivas.

Las carencias señaladas por estos autores se agravan por 1) la falta de delimitación del usuario del diccionario y por 2) el casi nulo aprovechamiento de la alta tecnología que llevan a cabo las editoriales encargadas de la producción de estos diccionarios enfocado en una mayor personalización de sus productos.

4. Cómo seleccionar datos culturales para productos lexicográficos

Veamos cómo se puede solucionar la selección y la representación de todos los elementos del significado, en nuestro caso en referencia a datos culturales, operando con los datos provenientes directamente de Internet y con un método de trabajo lexicográfico como lo es el establecido según el marco teórico de la TFL. Los enunciados fundamentales de esta teoría fueron formulados en la tesis doctoral de Tarp (1992) y popularizados por Bergenholtz y Tarp (1995), entre varias otras publicaciones. Debido a las duras críticas de Wiegand (2001),

fue renovada con una terminología mejor definida, principalmente por medio de las publicaciones de Bergenholtz y Tarp (2002, 2003, 2004) y culminando con Tarp (2008). El objetivo de la TFL es «disponer de una teoría transformadora que sirva de herramienta teórica capaz de guiar la práctica, es decir, la concepción y compilación de futuras obras lexicográficas» (Tarp 2015, 32). Elaborada con este objetivo, la TFL se basa en los dos siguientes enunciados axiomáticos fundamentales:

> Cualquier tipo de diccionario es una herramienta de uso concebida para ser consultada de forma rápida y fácil con el fin de satisfacer necesidades puntuales de información relacionadas con cualquier ciencia, disciplina y área de actividad humana, y que tienen tipos específicos de usuarios que se encuentran en tipos específicos de situaciones extra-lexicográficas.
> Las necesidades de información de los usuarios se resuelven dando acceso a datos lexicográficos explícitamente preparados, de los cuales los usuarios pueden extraer la información requerida que luego pueden emplear para un sinfín de propósitos relacionados con las situaciones extra-lexicográficas en las que las necesidades originalmente se produjeron. (Tarp 2015: 32)

En otras palabras, la TFL aporta las directrices teóricas necesarias para guiar la labor de los equipos lexicográficos en la elaboración de las nuevas obras de referencia aprovechando los logros aportados por la lexicografía en papel pero también adaptándose a la era digital. A este respecto, tiene en cuenta una serie de conceptos que constituyen la esencia misma de la lexicografía, como *usuario*, *acceso rápido y eficaz a los datos* y definición de *situación* y *necesidades lexicográficamente relevantes*. Más concretamente, la TFL adopta como punto de partida la definición del perfil del usuario y de las necesidades surgidas en las distintas situaciones.

El tratamiento de los términos ligados a la cultura es amplio y abarca desde el proceso de selección de las unidades, hasta su tratamiento y presentación (cfr. Manukyan Vardanyan 2015). En el proceso general de compilación del diccionario descrito por Fuertes-Olivera y Tarp (2014: 85), hay tres situaciones en los que los lexicógrafos pueden necesitar acceso a datos empíricos para hacer un buen trabajo:

1. Cuando buscan información sobre las necesidades lexicográficas de los usuarios previstos para preparar un concepto de diccionario que pueda ayudar a estos usuarios a satisfacer sus necesidades.
2. Cuando seleccionan y preparan los datos lexicográficos para incluirlos en el diccionario.
3. Cuando evalúan la utilidad del diccionario en términos de satisfacción del usuario.

Posteriormente, Tarp y Fuertes-Olivera (2016) indicaron que hay una cuarta situación en la que se pueden requerir datos empíricos externos, aunque está más relacionada con el aspecto comercial del proyecto que con los aspectos lexicográficos: el análisis del mercado para determinar las posibilidades de venta del producto.

En cada uno de estos casos hay un conjunto de métodos que pueden parecer más o menos fiables y rápidos en términos de productividad y calidad del producto final. Con respecto a la segunda de las situaciones en las que los lexicógrafos pueden necesitar acceso a datos empíricos para hacer un buen trabajo (cuando seleccionan y preparan los datos lexicográficos para incluirlos en el diccionario), es necesario consultar varios tipos de fuentes. Para seleccionar los datos culturales con fines lexicográficos podemos partir de tipos diferentes de fuentes de datos. Apuesto por usar Internet directamente como corpus mediante buscadores. Si un lexicógrafo puede encontrar inspiración en la gran colección de textos que es Internet, entonces Internet también puede considerarse un tipo de «corpus» lexicográfico de acuerdo con la definición anterior. Este es también el punto de vista de Kilgarriff y Grefenstette (2003: 334) quienes escriben que «The answer to the question "Is the web a corpus?" is yes».

Para poder satisfacer las necesidades del usuario previsto, en la compilación de diccionarios hay que identificar y seleccionar el número adecuado de lemas y de elementos de significado. A continuación, me centro en el segundo de los procesos.

Para la selección de los elementos culturales de cada lema, se puede establecer la misma metodología para seleccionar los elementos

de significado (esto es, acepciones) de cada lema que se sigue en los *Diccionarios en línea del español Valladolid-Uva* (cfr. Fuertes-Olivera 2019), un método inspirado por otro similar utilizado en los diccionarios en línea del danés (cfr. Bergenholtz y Agerbo 2014), pero que tiene algunos rasgos propios. Dicha metodología se lleva a cabo por medio de una interfaz de usuario del lexicógrafo, una herramienta desarrollada por Ordbogen.com y concebida por Henning Bergenholtz del *Centre for Lexicography* de la Universidad de Aarhus (Tarp y Fuertes-Olivera (2016: 279). En términos generales, el método de selección de significado al que he aludido abarca los 15 pasos explicados en Tarp y Fuertes-Olivera (2016: 280-283) y en Tarp (2018b: 64-67):

1. Se elige un lema contenido en la base de datos en la interfaz de usuario del lexicógrafo.
2. El botón «Google» a la izquierda en la interfaz del lexicógrafo está activado.
3. Aparece un resultado de búsqueda de Google «tradicional».
4. Las primeras (1-3) páginas se omiten porque solo contienen datos lexicográficamente irrelevantes.
5. Los minitextos que aparecen en cada página se leen para tener una idea general de lo que se trata.
6. Usando el método «copiar y pegar», las partes relevantes de los minitextos se copian en un documento de Word.
7. Simultáneamente, se seleccionan colocaciones, ejemplos, sinónimos, antónimos y formaciones de palabras para introducirlos en los campos respectivos en la ficha que representa el sentido en cuestión en la interfaz del lexicógrafo. Los modismos y las expresiones fijas se envían al editor en jefe para una evaluación adicional.
8. Se revisan varias páginas de Google hasta que no aparecen más datos nuevos y todo se repite. El número de páginas depende de las características de cada lema, así como de la intuición del lexicógrafo basada en la experiencia.
9. Una vez que se ha seleccionado una cantidad satisfactoria de datos empíricos, estos datos se agrupan según su significado.
10. En base a los grupos de datos, las primeras definiciones se escriben de acuerdo con las instrucciones lexicográficas preparadas por el editor en jefe.
11. Ahora el lexicógrafo decide si está satisfecho o si es necesario repetir el proceso, o parte del proceso, para obtener una cantidad satisfactoria de evidencia empírica.

12. Cuando el lexicógrafo ha finalizado la selección de significado y ha escrito las definiciones de los sentidos dirigidos a un lema, se envía un mensaje al editor en jefe.
13. El editor en jefe revisa las definiciones y las compara con las que aparecen en cuatro diccionarios de español. Si falta algo, esto puede conducir a un nuevo proceso de búsqueda, ya que es un principio básico en el proyecto que no se copia ninguna definición de otros diccionarios.
14. Si las definiciones están relacionadas con términos especializados que aparecen en el lenguaje general, se puede consultar a expertos externos para controlar su corrección.
15. Cuando el editor en jefe está satisfecho, y se incluyen otros datos relevantes como gramática, sinónimos, antónimos, formaciones de palabras, colocaciones y oraciones de ejemplo, el lema en cuestión se indica para su publicación en línea.

Aunque no hay referencia explícita a los datos culturales en este patrón, podemos considerar que se trata de unos datos que requieren la consulta de fuentes especializadas (paso 14). En el caso de que los datos culturales no pudieran obtenerse consultando a expertos directamente, habría que consultar fuentes especializadas. Por un lado, sería necesario consultar obras que abordan la dimensión lingüístico-cultural de la lengua, como las que se contienen en el proyecto *Diccionario lingüístico-cultural* dirigido por Luque Nadal (2009: 19). Por otro lado, podrían documentarse datos en obras especializadas en el ámbito específico en el que se usan las palabras o expresiones con datos culturales. Así, para la palabra *hat-trick* usada en el ámbito del fútbol, entre otros, pueden consultarse obras como el *The Dictionary of cricket* de Rundell (1985), el *Diccionario de fútbol* de Nomdedeu-Rull (2009) o el artículo «Hat-Trick» de Nomdedeu-Rull (2017) en el que se explica la historia de esta palabra y su evolución morfológica y semántica con una importante base de documentación histórica.

El establecimiento de este patrón permite la obtención de muchísimos más datos culturales de los que proporcionan, por ejemplo, los diccionarios del español en relativamente poco tiempo y gastando muchos menos recursos que mediante la elaboración de un corpus textual *ad hoc*. La entrada de la palabra *esvástica* en los *Diccionarios Valladolid-Uva* (*vid*. Cuadro 1) constituye un ejemplo ilustrativo:

Nombre <una esvástica, la esvástica, unas esvásticas, las esvásticas>
Definiciones
1 Signo solar muy antiguo que presenta variadas formas, normalmente de forma circular y en forma de cruz con cuatro brazos iguales cuyos extremos se doblan en ángulo recto en la misma dirección.
Sinónimo: cruz gamada
Imagen: https://es.wikipedia.org/wiki/Esv%C3%A1stica#/media/Archivo:Vaso_celt%C3%ADbero_con_anillas_colgantes_(M.A.N._Inv.1920-37-94)_01.jpg

Ejemplo: La esvástica tiene una larga historia. Se usaba al menos 5.000 años antes de que Adolf Hitler diseñara la bandera nazi.

2 Símbolo hindú sagrado; tiene forma de cruz con cuatro brazos iguales, cada uno de los cuales se inclina en el medio formando un ángulo de 90 grados; en la religión hindú aparecen un punto en cada brazo; si se orienta hacia la derecha representa al dios creador; si se orienta hacia la izquierda representa al dios destructor; es un símbolo de la buena suerte, de la paz y de la prosperidad.
Sinónimo: suástica
Imagen: https://es.wikipedia.org/wiki/Esv%C3%A1stica#/media/Archivo:HinduSwastika.svg

Ejemplo: La esvástica se encuentra por todas partes en los templos de la religión hindú, así como en símbolos, altares, escenas e iconografía en India y Nepal, tanto en el pasado como en nuestros días.

3 Símbolo budista sagrado; tiene forma de cruz con cuatro brazos iguales; está en ángulo recto; se pueden orientar hacia la derecha o hacia la izquierda; es más frecuente la orientación hacia la izquierda; representa prosperidad, vida, fortuna y fortaleza;
Sinónimo: sauvástica
Imagen: https://i1.wp.com/xixerone.com/wp-content/uploads/2012/08/esvasticas_japon1.jpg?resize=560%2C372&ssl=1

Ejemplo: La esvástica puede «girar» a la derecha (dextrogira) cuando el brazo superior apunta hacia la derecha o «girar» a la izquierda (levogira) cuando apunta a la izquierda.

4 Símbolo antiguo adoptado por los nazis en 1920; tiene forma de cruz con cuatro brazos iguales orientados principalmente hacia la derecha; la cruz aparece girada 45 grados; representa a la raza aria y es un símbolo del antisemitismo y del racismo
Sinónimo: cruz gamada
Imagen: https://es.wikipedia.org/wiki/Esv%C3%A1stica#/media/Archivo:Nazi_Swastika.svg

Ejemplo: Los teóricos nazis asociaron el uso de la esvástica con sus tesis que afirmaban la ascendencia cultural del pueblo alemán de la llamada raza aria.

Cuadro 1. Esvástica en los *Diccionarios en Línea de Español Valladolid-Uva*.

Con este método, en los *Diccionarios Valladolid-Uva* se pretende que el diccionario solucione una de las lagunas observadas en la lexicografía española: no hay diccionarios de definiciones sino diccionarios generales definitorios. Para ello, se defienden ideas como las siguientes:

> (a) la discusión sobre la naturaleza del significado está fuera del ámbito de la lexicografía; (b) los datos lexicográficos deben ser correctos, pero «correcto» no es lo mismo que «verdadero» en lexicografía; (c) la selección de los datos lexicográficos a incluir en el diccionario debe estar basada en las necesidades del usuario evitando en todo momento la sobrecarga informativa; (d) un diccionario de internet como el *diccionario español de definiciones* es un diccionario dinámico con artículos dinámicos necesarios para favorecer el proceso de individualización que se configura como «el Santo Grial» de la lexicografía de internet; (e) para poder hacer este tipo de diccionarios hay que diseñar *ex profeso* los tres componentes básicos de un diccionario de internet, a saber, la base de datos lexicográfica o editor, el motor de búsqueda y la interfaz del diccionario. (Fuertes-Olivera y Bergenholtz 2018: 169)

5. Conclusiones

En este estudio se ha repasado de qué manera la selección de datos directamente de Internet permite proporcionar productos lexicográficos destinados a diferentes tipos de usuarios. La ausencia de datos culturales en los diccionarios de ELE mostrada por los diferentes estudios que han analizado los datos culturales en diferentes tipos de diccionarios exige que se tenga que comenzar a trabajar con esta metodología de trabajo y desde la perspectiva de la TFL para compilar y producir productos lexicográficos que satisfagan las necesidades de información (fundamentalmente, de cuestiones comunicativas y cognitivas) de los diferentes tipos de aprendices (tanto nativos como no nativos) de español.

En este artículo, y ante la ausencia de un proyecto lexicográfico para tal fin, se han usado datos de los *Diccionarios en línea del español Valladolid-Uva*, así como los proporcionados por diferentes investigaciones que se han centrado en el análisis de los datos culturales en los diccionarios. Con el método de trabajo para la elaboración de estos diccio-

narios, la confección de productos lexicográficos de diferente tipo se puede realizar en relativamente poco tiempo e invirtiendo muchos menos recursos que mediante la elaboración de un corpus textual *ad hoc*.

Referencias bibliográficas

Algeo, John (1995): «British and American biases in English dictionaries», en Braj B. Kachru, Henry Kahane (eds.), *Cultures, Ideologies and the Dictionary*, Tübingen, Max Niemeyer Verlag, pp. 205-212.

Bergenholtz, Henning; Agerbo, Heidi (2014): «Extraction, selection and distribution of meaning elements for monolingual information tools», *Lexicographica*, n. 30, pp. 488-512. En línea <https://doi.org/10.1515/lexi-2014-0015> [Consulta: 28/09/2020].

Bergenholtz, Henning; Nielsen, Sandro (2013): «The treatment of culture-bound items in dictionaries», en Rufus H.Gouws, Ulrich Heid, Wolfgang Schweickard, Herbert Ernst Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, Berlin/Boston, Walter De Gruyter, pp. 469-481.

Bergenholtz, Henning; Norddahl, Bjarni (2012): «Ordbogsartikler som ingen læser», *LexicoNordica*, n. 19, pp. 207-223.

Bergenholtz, Henning; Tarp, Sven (eds.) (1995): *Manual of Specialised Lexicography*, Amsterdam, John Benjamins.

Bergenholtz, Henning; Tarp, Sven (2002): «Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen», *Lexicographica*, n. 18, pp. 253-263.

Bergenholtz, Henning; Tarp, Sven (2003): «Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions», *Hermes, Journal of Linguistics*, n. 31, pp. 171-196.

Bergenholtz, Henning; Tarp, Sven (2004): «The concept of "dictionary usage"», en Cay Dollerup (ed.), *Worlds of Words. A tribute to Arne Zettersten. Special Issue*, *Nordic Journal of English Studies*, vol. 3, n. 1, pp. 23-36

Bosque, Ignacio (1982): «Sobre la teoría de la definición lexicográfica», *Verba*, n. 9, pp. 105-123. En línea, https://minerva.usc.es/xmlui/handle/10347/4963?show=full [Consulta: 28/09/2020].

Cablitz, Gabriele H. (2011): «Documenting Cultural Knowledge in Dictionaries of Endangered Languages», *International Journal of Lexicography*, vol. 24, n. 4, pp. 446-462. En línea, <https://doi.org/10.1093/ijl/ecr017> [Consulta: 28/09/2020].

Fuertes-Olivera, Pedro A. (2012): «Lexicography and the Internet as a (Re-) source», *Lexicographica*, n. 28, pp. 49-70.

Fuertes-Olivera, Pedro A. (2013): «Electronic encyclopedias», en Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard, Herbert Ernst Wiegand (eds.), *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, Berlin/Boston, Walter De Gruyter, pp. 1069-1081.

Fuertes-Olivera, Pedro A. (2019): «Designing and Making Commercially Driven Integrated Dictionary Portals: the Diccionarios Valladolid-Uva», *Lexicography*, vol. 6, n. 1, pp. 21-41. En línea, <https://link.springer.com/article/10.1007%2Fs40607-019-00056-8> [Consulta: 28/09/2020].

Fuertes-Olivera, Pedro A.; Bergenholtz, Henning (2018): «Diccionario Español de Definiciones», *Círculo de Lingüística Aplicada a la Comunicación*, n. 74, pp. 169-194. En línea, <http://webs.ucm.es/info/circulo/no74/fuertes.pdf> [Consulta: 12/03/2020].

Fuertes-Olivera, Pedro A.; Tarp, Sven (2014): *Theory and Practice of Specialised Online Dictionaries: Lexicography versus Terminography*, Berlin/New York, De Gruyter.

González-Fernández, Adela (2016): *Más allá del corpus: Big data en la investigación lingüística. Evolución, análisis y predicción del uso de la lengua a través de Twitter*, Córdoba, Universidad de Córdoba, Tesis doctoral dirigida por María Luisa Calero Vaquera y Gloria Guerrero Ramos. En línea, <https://helvia.uco.es/xmlui/handle/10396/13668> [Consulta: 13/07/2021].

Kilgarriff, Adam; Grefenstette, Gregory (2003): «Introduction to the Special Issue on the Web as Corpus», *Computational Linguistics*, n. 29, pp. 333-347. En línea, <https://www.aclweb.org/anthology/J03-3001.pdf> [Consulta: 16/09/2020].

Klosa, Annette (2015): «Illustrations in Dictionaries; Encyclopaedic and Cultural Information in Dictionaries», en Philip Durkin (ed.), *The Oxford Handbook of Lexicography*, Oxford, Oxford University Press, pp. 515-531.

Lara, Luis Fernando (1990): «Diccionario de lengua, enciclopedia y diccionario enciclopédico: el sentido de sus distinciones», en Luis Fernando Lara, *Dimensiones de la lexicografía*, México, El Colegio de México, pp. 213-231.

Luque Nadal, Lucía (2009): «Los diccionarios lingüístico-culturales y el estudio de los fraseologismos», *Boletín Hispánico Helvético*, n. 11, pp. 5-23. En línea, <https://vdocuments.mx/los-diccionarios-lingueistico-culturales-y-el-estudio-de-los-fraseologismos-lucia-luque-nadal.html> [Consulta: 28/09/2020].

Manukyan Vardanyan, Lilit (2015): *Culture-bound terms en los diccionarios bilingües (español-ruso y español-armenio)*, Barcelona, Universitat Autònoma de Barcelona, Tesis doctoral dirigida por Xavier Blanco Escoda. En línea, <https://ddd.uab.cat/record/148809> [Consulta: 16/09/2020].

Miller, Julia (2018): «Learners' dictionaries of English», en Pedro A. Fuertes-Olivera (ed.), *The Routledge Handbook of Lexicography*, New York, Taylor & Francis Group, pp. 353-366.

Miyares Bermúdez, Eloína (dir.) (2005): *Diccionario Básico Escolar*, Santiago de Cuba, Centro de Lingüística Aplicada.

Miyares Bermúdez, Eloína (dir.) (2006): *Léxico activo-funcional del escolar cubano*, Santiago de Cuba, Centro de Lingüística Aplicada.

Newmark, Peter (1988): *A Textbook of Translation*, Hertfordshire, Prentice Hall.

Nomdedeu-Rull, Antoni (2009): *Diccionario de fútbol*, *Anexos* de *Revista de Lexicografía*, n. 11, A Coruña, Servizo de Publicacións-Universidade da Coruña.

Nomdedeu-Rull, Antoni (2017): «Hat-trick», *Martes neológico*. En línea, <http://blogscvc.cervantes.es/martes-neologico/hat-trick/> [Consulta: 28/09/2020].

Nomdedeu-Rull, Antoni (2018): «El desafío del tratamiento de los datos culturales en un diccionario en línea para aprendices de español como LE/L2», *Revista de Filología de la Universidad de La Laguna*, 36, pp. 279-308. En línea, <https://riull.ull.es/xmlui/handle/915/8621> [Consulta: 28/09/2020].

Nomdedeu-Rull, Antoni (2020): «How to Select and Present Cultural Data: A Challenge to Lexicography», *Lexicographica. International Annual for Lexicography*, vol. 36, pp. 39-57. En línea, <https://www.degruyter.com/document/doi/10.1515/lex-2020-0003/html> [Consulta: 30/11/2020].

Nomdedeu-Rull, Antoni (2022): «Lexicographic Perspectives and L2 Spanish Vocabulary», en Joe Barcroft, Javier Muñoz-Basols (eds.), *Spanish Vocabulary Learning in Meaning-Oriented Instruction*, New York, Taylor & Francis Group, pp. 89-111.

Nomdedeu-Rull, Antoni; Barcroft, Joe (2022, en prensa): «Los diccionarios de ELE», en Sergi Torner, Paz Battaner, Irene Renau (eds.), *Lexicografía hispánica / The Routledge Handbook of Spanish Lexicography*, New York, Taylor & Francis Group.

Nomdedeu-Rull, Antoni; Tarp, Sven (2018): «Hacia un modelo de diccionario en línea para aprendices de español como LE/L2», *Journal of Spanish Language Teaching*, vol. 5, n. 1, pp. 50-56. En línea, https://doi.org/10.1080/23247797.2018.1459279 [Consulta: 28/09/2020].

Parodi, Giovanni (2008): «Lingüística de corpus: una introducción al ámbito», *Revista de Lingüística Teórica y Aplicada*, vol. 46, n. 1, pp. 93-119. En línea, <https://www.scielo.cl/pdf/rla/v46n1/art06.pdf> [Consulta: 13/07/2021].

Rey, Alain (1988): *Enciclopedias y diccionarios*, México, Fondo de Cultura Económica.

Rojo, Guillermo (2009): «Sobre la construcción de diccionarios basados en corpus», *Tradumàtica*, n. 7, pp. 1-7. En línea, <https://ddd.uab.cat/record/52620> [Consulta: 16/09/2020].

Sánchez-Pérez, Aquilino (2010): «The Treatment of Cultural and/or Encyclopaedic Items in Specialised Dictionaries for Learners», en Pedro A. Fuer-

tes-Olivera (ed.), *Specialised Dictionaries for Learners*, Berlin/New York, Walter de Gruyter, pp. 107-129.

Sinclair, John (1987): «Introduction», *Collins Cobuild English Language Dictionary*, London, HarperCollins Publishers, pp. XV-XXI.

Stark, Martin (1999): *Encyclopedic Learners' Dictionaries. A Study of their Design Features from the User Perspective*, Tübingen, Max Niemeyer Verlag.

Szerszunowicz, Joanna (2011): «The Cultural Component in Bilingual Dictionaries of Phraseological Units», en Kaoru Akasu, Satoru Uchida (eds.), *LEXICOGRAPHY: Theoretical and Practical Perspectives*, Tokyo, ASIALEX, pp. 628-637. En línea, <https://www.researchgate.net/publication/308522330_The_cultural_component_in_bilingual_dictionaries_of_phraseological_units_in_LEXICOGRAPHY_Theoretical_and_Practical_Perspectives_Papers_submitted_to_the_Seventh_ASIALEX_Biennial_International_Conferenc> [Consulta: 16/09/2020].

Tarp, Sven (1992): *Prolegomena til teknisk ordbog*, Aarhus, Århus, Handelshøjskolen i Århus, Spansk Institut, Tesis doctoral.

Tarp, Sven (2000): «Theoretical Challenges to Practical Specialised Lexicography», *Lexikos*, n. 10, pp. 189-208. En línea, <10.5788/10-0-894> [Consulta: 16/09/2020].

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

Tarp, Sven (2015): «La teoría funcional en pocas palabras», *Estudios de Lexicografía*, n. 4, pp. 31-42. En línea, <https://cc.au.dk/fileadmin/04._La_teori__a_funcional_en_pocas_palabras.pdf> [Consulta: 13/07/2021].

Tarp, Sven (2018a): «Lexicography as an Independent Science», en Pedro A. Fuertes-Olivera (ed.), *Routledge Handbook of Lexicography*, London/New York, Routledge, pp. 19-33.

Tarp, Sven (2018b): «¿Adiós a los corpus para fines lexicográficos?», Leonel Ruiz Miyares (ed.), *Estudios de Lexicología y Lexicografía: Homenaje a Eloína Miyares Bermúdez*, Santiago de Cuba, Ediciones Centro de Lingüística Aplicada, pp. 57-77.

Tarp, Sven; 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*, n. 26 (AFRILEX-reeks/series 26), pp. 273-295. En línea, <10.5788/26-1-1349> [Consulta: 16/09/2020].

Tseng, Taiyuan (2003): «The Cultural Component in an Ideal Monolingual English Learner's Dictionary», en Li Ming, Hua Hue-Feng, Zhang Ji Zhuang (eds.), *Aspects of English Language Teaching and Research*, Suzhou, P.R. China, Suzhou University Press, pp. 215-226.

Vermeer, Hans J. (1983): «Translation Theory and Linguistics», Pauli Roinila, Ritua Orfanos, Sonja Tirkkonen-Condit (eds.), *Näkökohtia kääntämisen tutkimuksesta*, Joensuu, Joensuu University, pp. 1-10.

Welker, Herbert A. (2008): *Panorama Geral da Lexicografia Pedagógica*, Brasilia, Thesaurus Editora.

Werner, Reinhold (1982): «La definición lexicográfica», en Günter Haensch, Lothar Wolf, Stefan Ettinger, Reinhold Werner (eds.), *La lexicografía*, Madrid, Gredos, pp. 259-328.

Villayandre Llamazares, Milka (2010): *Aproximación a la lingüística computacional*, León, Universidad de León Departamento de Filología Hispánica y Clásica, Tesis Doctoral dirigida por Salvador Gutiérrez Ordóñez y Manuel Iglesias Bango.

Zgusta, Ladislav (1971): *Manual of lexicography*, Den Haag-Paris, Mouton.

Zgusta, Ladislav (1989): «Idle Thoughts of an Idle Fellow; or, Vaticinations on the Learners' Dictionary», en Makhan L. Tickoo (ed.), *Learner's Dictionaries: The State of the Art*, Singapore, SEAMO Regional Language Centre, pp. 1-9.

En este volumen se pretende ofrecer una muestra de las que, en la actualidad, son las líneas de investigación más destacadas y avanzadas acerca de la compilación o la explotación de corpus especializados en distintos ámbitos, como el comercio, el sector agroalimentario o la medicina, con una orientación dirigida hacia la traducción y la didáctica del español LE/L2. Por ello, se ha dado voz a investigadores que trabajan en este marco aplicando una pluralidad de enfoques metodológicos y herramientas de gestión y análisis de corpus. Los capítulos que componen este libro se vertebran bajo un eje sincrónico y se reparten en dos secciones, la primera dedicada a la compilación y explotación de los corpus en el contexto de la traducción, mientras que la segunda hace hincapié en su aprovechamiento en la enseñanza del español como lengua extranjera.

ISBN 978-88-3369-128-2

Euro 25,00

UNIVERSITAS STUDIORUM