

Reflexión sobre la extracción y traducción de unidades fraseológicas especializadas en el ámbito científico

Eva Lucía Jiménez-Navarro
Universidad de Córdoba
lucia.jimenez@uco.es

Abstract

This chapter presents a set of observations prompted in the extraction and translation of specialized phraseological units in scientific language. For so doing, the starting point is a specialized corpus consisting of research articles written in English, which is the base for the extraction of phraseological units. Thus, the objectives defined are twofold: first, to explain the differences between applying two distinct methods for this task, and second, to describe three ways to translate these groupings of words from English into Spanish, by exemplifying each of them. The methodology employed for the fulfilment of the second purpose was based on a bilingual dictionary and a Spanish reference corpus. Furthermore, the attainment of both objectives was possible with the use of the software *Sketch Engine*. Preliminary results reveal the existence of phraseological units ranging from more to less specific and demonstrate the value of a corpus in the target language to produce a more accurate translation.

Key words: corpus, extraction, scientific language, *Sketch Engine*, translation, specialized phraseological unit.

Resumen

Este capítulo presenta una serie de consideraciones implicadas en la extracción y traducción de unidades fraseológicas especializadas en el lenguaje científico. Para ello, partimos de un corpus especializado compuesto de artículos científicos en inglés, base para la extracción de unidades fraseológicas. Así, los objetivos propuestos son los siguientes: primero, explicar las diferencias entre la aplicación de dos métodos distintos para esta tarea, y segundo, describir tres vías para la traducción de inglés a español de estos conjuntos de palabras, ejemplificando cada una de ellas. Para este segundo propósito, la metodología empleada se ha basado en un diccionario bilingüe y en un corpus de referencia en español. Además, el software *Sketch Engine* ha posibilitado el logro de ambos objetivos. Los resultados preliminares indican la existencia de unidades fraseológicas graduables en función de su especificidad y demuestran el valor de un corpus en la lengua meta para una traducción más exacta.

Palabras clave: corpus, extracción, lenguaje científico, *Sketch Engine*, traducción, unidad fraseológica especializada.

1. Introducción

El Diccionario de la Real Academia Española define *ciencia* como el “Conjunto de conocimientos obtenidos mediante la observación y el razonamiento, sistemáticamente estructurados y de los que se deducen principios y leyes generales con capacidad predictiva y comprobables experimentalmente”¹. Uno de los medios más demandados para la transmisión de esos conocimientos es el artículo científico, cuyo propósito fundamental es transferir información de manera objetiva. Sin embargo, esta tarea podría complicarse debido a la gran carga de *términos* que porta, es decir, vocabulario propio del mismo y que lo caracteriza. En particular, las combinaciones que generan estas unidades léxicas han cobrado especial importancia en los últimos años y han sido el objeto de estudio de numerosos autores (p. ej., Jiménez-Navarro, 2019; Luzón Marco, 2000; Pérez-Llantada, 2014).

Su investigación se enmarca en el contexto de la fraseología especializada, área que abarca combinaciones de palabras usadas en lenguajes especializados y que surgió como campo de estudio dentro de la disciplina de la Terminología (Bevilacqua, 2004; Pavel, 1993; Sanz Vicente, 2011). Las características que definen a las unidades fraseológicas (de ahora en adelante, UFs) especializadas no difieren significativamente de las que identifican a las UFs de la lengua general, como son su estructura fija y estable a nivel semántico y morfosintáctico, graduable en función de su tipología (Corpas Pastor, 1996). No obstante, Lorente Casafont (2002) destaca algunas diferencias básicas, ya que las UFs especializadas se componen de términos y representan a una comunidad especialista. Además, acarrear conocimiento especializado y pueden contener un fuerte componente cultural, lo cual dificulta la labor traductológica. De hecho, Caballero Artigas (2018) sostiene que una traducción especializada “no se basa simplemente en reproducir un mensaje de una lengua a otra” (p. 48), sino que el traductor debe actuar de enlace entre dos culturas.

En este contexto, y teniendo en cuenta que entendemos la traducción especializada como un proceso de adaptación de una situación comunicativa a otra (Parra Membrives, 2013), los objetivos perseguidos son dos: en primer lugar, reflexionar sobre dos vías de extracción de UFs especializadas en el ámbito de la ciencia, y, en segundo lugar, profundizar en diferentes técnicas para su traducción, en este caso, de inglés a español. Para alcanzar estas metas, el capítulo se estructura de la siguiente manera. La Sección 2 describe la metodología empleada en este estudio, indagando más detenidamente en el primer objetivo planteado. A continuación, la Sección 3 ahonda mayormente en los resultados obtenidos para el segundo objetivo. Finalmente, la Sección 4 presenta las conclusiones alcanzadas.

2. Metodología

2.1. Corpus representativo: Condición *sine qua non*

La realización de un estudio lingüístico sobre cualquier tipo de lenguaje requiere, en primera instancia, la existencia de un corpus representativo (Biber y Reppen, 2015; Lindquist, 2009; Meyer, 2002; Tognini-Bonelli, 2001). Puesto que conseguir esto no es siempre tarea fácil, la construcción de un nuevo corpus ocurre, normalmente, cuando el/la investigador/a no tiene a su disposición uno que atienda a sus necesidades, algo muy común en los estudios sobre lenguajes especializados.

Ilustremos lo dicho presentando UCOSCIENCOR, el corpus monolingüe especializado en el que se ha basado este trabajo. Fue compilado para el análisis del

¹ <https://dle.rae.es/ciencia?m=form> (Último acceso: 02/03/2022).

lenguaje utilizado por renombrados/as académicos/as de la Universidad de Córdoba (España) y recoge seiscientos artículos científicos extraídos manualmente de páginas de acceso libre. Aunque estos textos fueron redactados por hablantes no nativos/as, se debe subrayar que fueron publicados en revistas científicas internacionales de reconocido prestigio, como *Nature Reviews Immunology*, *The Lancet*, *Nature Reviews Molecular Cell Biology* o *Vaccine*, lo que indica un alto nivel lingüístico-discursivo. La compilación de estos textos y su anotación se llevó a cabo con la ayuda del software *Sketch Engine*² y se considera que el resultado fue satisfactorio, ya que el corpus, de más de tres millones de palabras, es representativo del género en cuestión.

2.2. Métodos para la extracción de unidades fraseológicas especializadas

Cuando existe un corpus representativo del lenguaje objeto de estudio, la extracción de UFs especializadas también se puede realizar con la herramienta *Sketch Engine*. Teniendo en cuenta que es posible la detección de UFs más o menos específicas, mostraremos dos funciones que ofrece este software, las cuales nos ayudaron a lograr nuestro primer objetivo.

Por un lado, las UFs más específicas son las que definen la temática del corpus. Esto es, la función *Keywords* permite extraer conjuntos de palabras típicas de un corpus especializado comparando los resultados con aquellos obtenidos en un corpus de referencia. Cuando la lengua de trabajo es el inglés, el corpus English Web 2020 ('enTenTen20') es elegido por defecto, ya que es el corpus más denso disponible en este idioma (cuenta con más de 38 100 millones de palabras). Para su identificación, la función adopta un enfoque híbrido y aplica criterios estadísticos y lingüísticos. Existe, empero, el inconveniente de que solo extrae UFs nominales. Para ilustrar mejor, la Figura 1 muestra las 50 primeras UFs extraídas del UCOSCIENCOR³:

The screenshot shows the 'KEYWORDS' tool interface. At the top, the search term 'UCOSCIENCOR' is entered. Below, there are tabs for 'SINGLE-WORDS' and 'MULTI-WORDS TERMS'. The results are displayed in a table with 50 rows, each containing a word and its frequency in the target corpus (UCOSCIENCOR) and the reference corpus (English Web 2013). The words are sorted by frequency in the target corpus.

Word	Word	Word	Word
1 p p p	14 cardiovascular risk	27 necrosis factor	40 fusarium oxysporum
2 p p p p	15 metastatic colorectal cancer	28 total cholesterol	41 cardiovascular disease
3 p p p p p	16 insulin sensitivity	29 renal function	42 serum calcium
4 p p	17 wild boar	30 postprandial lipemia	43 statistical analysis
5 ankylosing spondylitis	18 vascular access	31 serum phosphate	44 fatty acid
6 oxidative stress	19 insulin resistance	32 postprandial response	45 peripheral blood
7 disease activity	20 view project	33 postprandial state	46 dietary intervention
8 metabolic syndrome	21 dna demethylation	34 gondii infection	47 peritoneal dialysis
9 present study	22 serum albumin	35 virgin olive oil	48 dialysis dose
10 mediterranean diet	23 bluetongue virus	36 situ hybridization	49 transcription factor
11 colorectal cancer	24 endothelial function	37 logistic regression	50 nutritional status
12 downloaded file	25 gene expression	38 endothelial dysfunction	
13 red deer	26 renal failure	39 lipid metabolism	

Fig. 1: Extracción de unidades fraseológicas especializadas con *Keywords*.

Por otro lado, las UFs menos específicas están más cercanas al lenguaje general. En este caso, la función *Word Sketch* incorpora, también, una perspectiva híbrida que combina un criterio estadístico, basado en la medida *logDice* y la frecuencia de la combinación, con otro lingüístico, consistente en la relación sintáctica entre ambos

² <https://www.sketchengine.eu/> (Último acceso: 02/03/2022).

³ Queda patente en los primeros resultados de la tabla (p. ej., “p p p”, “p p p p”) que la anotación automática del corpus no fue completamente correcta, hecho que puede suceder como apuntan algunos/as autores/as (Pežik, 2018, p. 66; Vargas Sierra, 2010, p. 29).

componentes de la UF. La Figura 2 muestra los 10 primeros resultados obtenidos y ordenados según su frecuencia en el UCOSCIENCOR (la frecuencia mínima establecida, dada por defecto, fue 5):

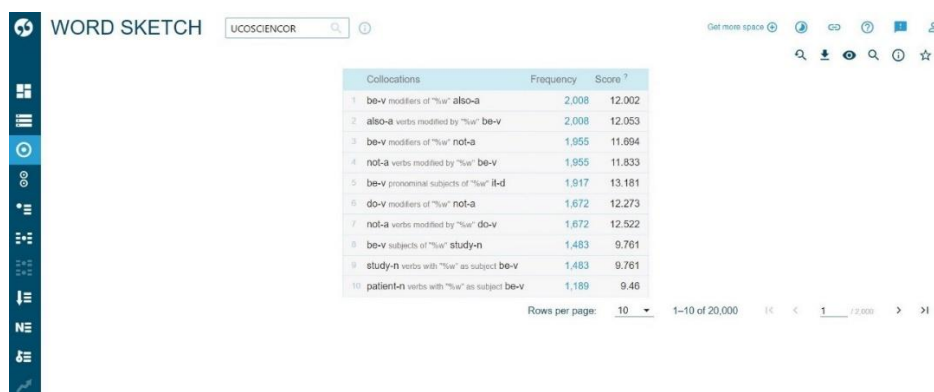


Fig. 2: Extracción de unidades fraseológicas especializadas con *Word Sketch*.

2.3. Técnicas para la traducción de unidades fraseológicas especializadas

Como hemos comentado anteriormente, la traducción especializada no es solo un proceso lingüístico, sino también cultural. El análisis de tres técnicas susceptibles de ser utilizadas para esta labor nos ayudará a conseguir nuestro segundo objetivo, y son las siguientes: primero, el/la traductor/a puede confiar en su experiencia previa y no recurrir a otras fuentes; segundo, el/la traductor/a puede utilizar un diccionario bilingüe, por ejemplo, el *Cambridge Dictionary*⁴, para conocer los significados de cada una de las palabras que componen la UF y/o para comprobar si la UF en cuestión aparece recogida en el mismo; tercero, el/la traductor/a puede hacer uso de un corpus en la lengua meta para valorar la traducción de la UF en contexto. Por otra parte, el/la traductor/a puede emplear un método híbrido y combinar dos o más posibilidades. La siguiente sección presenta algunos ejemplos.

3. Traducción de Unidades Fraseológicas Especializadas en el Lenguaje Científico: Inglés-Español

La primera de las técnicas mencionadas en la sección anterior, es decir, la confianza plena en la experiencia previa del/de la traductor/a para dar por buena una traducción especializada, puede resultar exitosa con UFs que admitan una traducción literal y cuyos elementos ya formen parte del léxico de la persona que lleva a cabo la tarea. La Tabla 1 muestra algunos ejemplos extraídos del UCOSCIENCOR y la función con la que fueron obtenidos:

Tab. 1: Traducción de unidades fraseológicas especializadas basada en la experiencia previa del/de la traductor/a.

Función	UF en inglés	Traducción al español
<i>Keywords</i>	<i>cardiovascular risk</i>	riesgo cardiovascular
<i>Keywords</i>	<i>metabolic syndrome</i>	síndrome metabólico
<i>Keywords</i>	<i>oxidative stress</i>	estrés oxidativo
<i>Word Sketch</i>	<i>subject + consume</i>	el sujeto consume

⁴ <https://dictionary.cambridge.org/> (Último acceso: 02/03/2022).

<i>Word Sketch</i>	<i>indicate + difference</i>	indicar una diferencia
<i>Word Sketch</i>	<i>show + decrease</i>	mostrar una disminución

La segunda de las técnicas consiste en el uso de un diccionario bilingüe para revisar la traducción de cada uno de los componentes de la UF por separado y también como combinación de palabras. Pongamos por caso la UF *renal failure*, extraída mediante la función *Keywords*. La traducción del primero de los términos es sencilla, dado que procede del latín y tiene una grafía similar en las dos lenguas de estudio, pero el segundo término podría resultar problemático. Su significado más común, con el que posiblemente el/la traductor/a esté más familiarizado/a, es *fracaso*, como se puede ver en la Figura 3, que representa un extracto de la entrada de *failure* en el *Cambridge Dictionary*⁵:



Fig. 3: Extracto de la entrada de *failure* en el *Cambridge Dictionary*; traducción más común.

Sin embargo, esto es insuficiente, ya que la UF **fracaso renal* no resulta del todo natural en español. Para solucionarlo, basta con consultar la entrada completa del término, la cual recoge UFs idénticas y nos permite saber que la traducción *insuficiencia renal* sería más acertada, como ilustra la Figura 4:



Fig. 4: Extracto de la entrada de *failure* en el *Cambridge Dictionary*; traducciones de unidades fraseológicas especializadas similares a *renal failure*.

La última de las técnicas de traducción puestas a prueba conlleva el uso de un corpus en la lengua meta. La opción ideal sería disponer de un corpus especializado, pero, si no existiera, se podría consultar uno general. Tomemos, de nuevo, el ejemplo de *renal failure*. Puesto que *Sketch Engine* no dispone de un corpus científico en español, el Spanish Web 2018 ('esTenTen18') puede aportar información sobre el uso de la traducción *insuficiencia renal*. Concretamente, la función *Concordance* permite el acceso a ejemplos directamente extraídos de los textos del corpus, como muestra la Figura 5⁶:

⁵ <https://dictionary.cambridge.org/dictionary/english-spanish/failure> (Último acceso: 02/03/2022).

⁶ También se comprobó la frecuencia de la traducción *fracaso renal* y el corpus de referencia solo produjo 636 resultados (frente a los 40 078 encontrados para *insuficiencia renal*, como refleja la Figura 5).

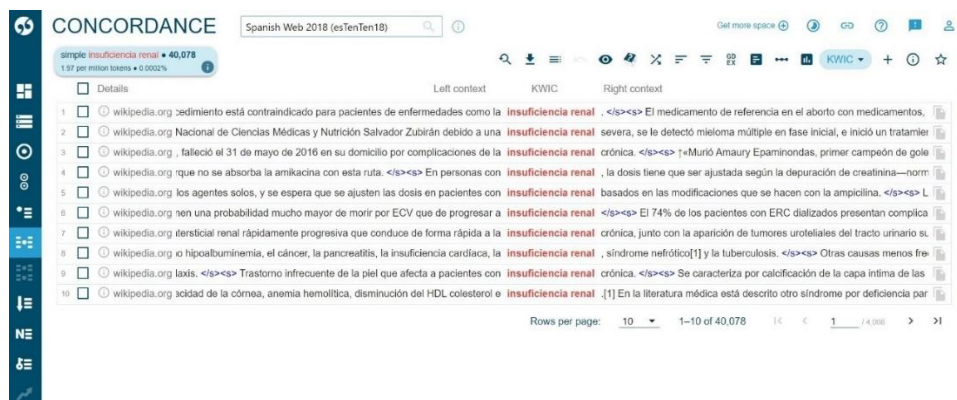


Fig. 5: Líneas de concordancia de *insuficiencia renal* en el ‘esTenTen18’.

Si en algunos casos esas pequeñas muestras obtenidas de los textos no fueran suficientes para verificar la especialización de la UF, se podrían extender los contextos haciendo clic sobre ellas. La Figura 6 demuestra que la línea de concordancia número 1 pertenece a un texto del ámbito científico:

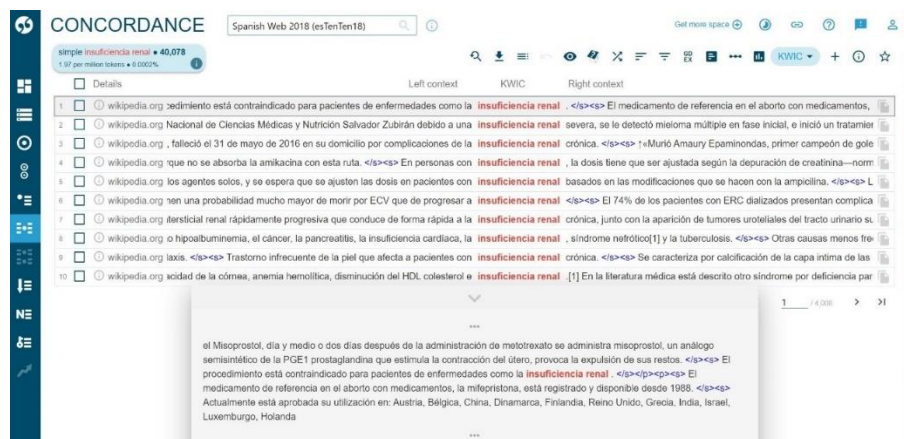


Fig. 6: Contexto extendido de la línea de concordancia número 1 de *insuficiencia renal* en el ‘esTenTen18’.

Otro ejemplo de la utilidad de un corpus en la lengua meta es el siguiente. La UF *lipid metabolism* se traduciría como *metabolismo lípido* según las técnicas uno y dos explicadas más arriba. En otras palabras, el/la traductor/a podría conocer sus traducciones literales y/o podría consultarlas en un diccionario bilingüe; en concreto, las únicas traducciones de los términos en el *Cambridge Dictionary* son *lípido*⁷ y *metabolismo*⁸, respectivamente. Ahora bien, parece que *metabolismo lípido* no es la traducción más frecuente si se consultan sus líneas de concordancia en el ‘esTenTen18’. Para comprender mejor, la Figura 7 revela que *metabolismo lípido* es correcto, pero que también existe la variación *metabolismo lipídico*:

⁷ <https://dictionary.cambridge.org/dictionary/english-spanish/lipid> (Último acceso: 02/03/2022).

⁸ <https://dictionary.cambridge.org/dictionary/english-spanish/metabolism> (Último acceso: 02/03/2022).

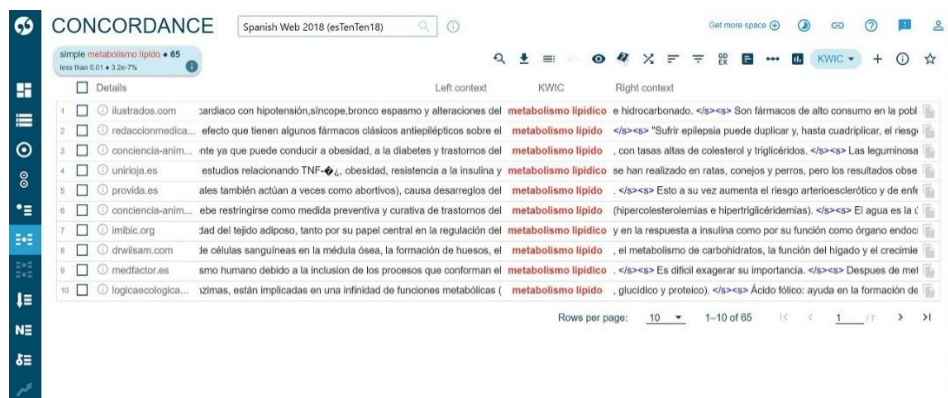


Fig. 7: Líneas de concordancia de *metabolismo lipido* en el ‘esTenTen18’.

Ante dos posibles traducciones de una misma UF especializada, conviene comprobar las líneas de concordancia de ambas en un corpus en la lengua meta para verificar qué opción es más frecuente. Así, por ejemplo, los resultados para *metabolismo lipídico* en la función *Concordance* son más numerosos que para *metabolismo lipido* (1 129 frente a 65), como constata la Figura 8, lo cual se puede considerar un indicador de que la traducción *metabolismo lipídico* sería más apropiada:

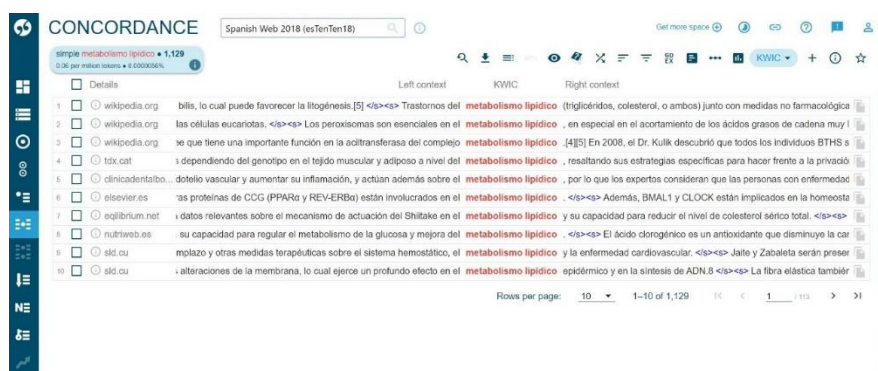


Fig. 8: Líneas de concordancia de *metabolismo lipídico* en el ‘esTenTen18’.

4. Conclusiones

Como conclusión de lo expuesto en el presente capítulo, presentamos un resumen de los dos puntos clave abordados y los resultados alcanzados.

En primer lugar, la extracción de UFs especializadas se puede llevar a cabo a través de dos funciones disponibles en el software *Sketch Engine*, a saber: (1) *Keywords*, la cual arroja UFs típicas del corpus analizado comparando parámetros estadísticos y lingüísticos con los obtenidos en un corpus de referencia, en otras palabras, son combinaciones relacionadas con la temática de los textos, y (2) *Word Sketch*, que produce combinaciones menos específicas y, por ende, menos influenciadas por la temática del corpus.

En segundo lugar, la traducción de UFs especializadas se puede realizar confiando exclusivamente en la experiencia previa que tiene el/la traductor/a sobre la lengua meta, si bien esto podría ser insuficiente. De esta manera, una técnica alternativa podría ser el uso de un diccionario bilingüe o, mejor aún, de un corpus en la lengua meta. Sin duda alguna, la consulta de corpus aseguraría una traducción más adecuada y natural y la inclusión de todos los aspectos lingüísticos y culturales de la UF especializada.

Finalmente, nuestra reflexión nos lleva a creer que el enfoque metodológico adoptado para la extracción y traducción de UFs especializadas en el ámbito científico puede ser extensible a los estudios sobre UFs de la lengua general o de otros lenguajes especializados, y esperamos que resulte útil para distintos tipos de profesionales, por ejemplo, traductores/as, lingüistas o filólogos/as.

Referencias bibliográficas

- BEVILACQUA, C. R. (2004). *Unidades fraseológicas especializadas eventivas: Descripción y reglas de formación en el ámbito de la energía solar* (Tesis doctoral, Universitat Pompeu Fabra, Barcelona, España). <<http://hdl.handle.net/10803/7515>> [02/03/2022]
- BIBER, D., Y REPPEN, R. (2015). Introduction. En D. Biber y R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 1-8). Cambridge: Cambridge University Press.
- CABALLERO ARTIGAS, H. L. (2018). Traducción y equivalencias en la fraseología español-francés. *Anales de Filología Francesa*, 26, 47-60.
- CAMBRIDGE DICTIONARY OF ENGLISH. (2022). Cambridge: Cambridge University Press.
- CORPAS PASTOR, G. (1996). *Manual de fraseología española*. Madrid: Gredos.
- JIMÉNEZ-NAVARRO, E. L. (2019). Nominal collocations in scientific English: A frame-semantic approach. En G. Corpas Pastor y R. Mitkov (Eds.), *Lecture Notes in Artificial Intelligence. Volume 11755. Computational and corpus-based phraseology* (pp. 187-199). Cham: Springer.
- LINDQUIST, H. (2009). *Corpus linguistics and the description of English*. Edimburgo: Edinburgh University Press.
- LORENTE CASAFONT, M. (2002). Terminología y fraseología especializada: Del léxico a la sintaxis. En G. Guerrero Ramos y M. F. Pérez Lagos (Eds.), *Panorama actual de la terminología* (pp. 159-179). Granada: Comares.
- LUZÓN MARCO, M. J. (2000). Collocational frameworks in medical research papers: A genre-based study. *Journal of English for Specific Purposes*, 19, 63-86.
- MEYER, C. F. (2002). *English corpus linguistics. An introduction*. Cambridge: Cambridge University Press.
- PARRA MEMBRIVES, E. (2013). El traductor como mediador cultural. En E. Parra Membrives y Á. García Calderón (Eds.), *Traducción, mediación, adaptación: Reflexiones en torno al proceso de comunicación entre culturas* (pp. 7-17). Berlín: Frank und Timme.
- PAVEL, S. (1993). Vers une méthode de recherche phraséologique en langue de spécialité. *L'Actualité Terminologique*, 26(2), 9-13.
- PÉREZ-LLANTADA, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94.
- PEZIK, P. (2018). *Facets of prefabrication. Perspectives on modelling and detecting phraseological units*. Łódź: Łódź University Press.
- REAL ACADEMIA ESPAÑOLA: *Diccionario de la lengua española*, 23.^a ed., [versión 23.5 en línea]. <<https://dle.rae.es>> [02/03/2022]
- SANZ VICENTE, M. L. (2011). *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español* (Tesis doctoral no publicada). Universidad de Salamanca, Salamanca, España.
- SKETCH ENGINE [SOFTWARE]. (2022). Brno: Lexical Computing Limited.
- TOGNINI-BONELLI, E. (2001). *Corpus linguistics at work*. Ámsterdam/Filadelfia: John Benjamins Publishing Company.
- VARGAS SIERRA, C. (2010). Combinatoria terminológica y diccionarios especializados para traductores. En M. Ibáñez Rodríguez (Ed.), *Lenguas de especialidad y terminología* (pp. 18-49). Granada: Comares.