

Metodología para la traducción de términos compuestos mediante corpus¹



Melania Cabezas-García
melaniacabezas@ugr.es
<https://orcid.org/0000-0002-8622-1036>
Universidad de Granada, España

Resumen

Los términos compuestos son una de las principales dificultades en la traducción de los textos especializados. Dado que estos términos no siempre figuran en los recursos terminológicos o lo hacen de forma poco exhaustiva, los traductores y terminólogos deben dominar diversas técnicas para resolver sus dificultades. Tradicionalmente, se han utilizado textos paralelos para extraer terminología y adquirir conocimientos especializados. Estos textos pueden compilarse en forma de corpus comparables (es decir, textos escritos originalmente en la lengua de origen y de llegada), que pueden utilizarse para explotar un mayor número de textos de manera más eficiente. No obstante, a menudo hay reticencia a su utilización, en algunos casos debido a la falta de conocimiento de las técnicas de consulta. En este estudio se describe una metodología para identificar en corpus comparables los equivalentes en inglés de los términos compuestos en español, con vistas a: 1) facilitar esta tarea en la labor cotidiana de los traductores y terminólogos; y 2) proporcionarles técnicas de extracción manual en corpus. Con este fin, analizamos un corpus comparable de textos especializados sobre energía eólica escritos originalmente en español e inglés. La metodología propuesta se basa en la semántica distribucional y consiste en la identificación y la comparación de elementos contextuales de los conceptos en ambos idiomas. Los resultados muestran que este procedimiento simplifica la identificación de equivalentes de los términos compuestos en corpus.

Palabras clave: corpus comparable; semántica distribucional; término compuesto; terminología; términos equivalentes; traducción de textos especializados.

A Method for Translating Multiword Terms Using Corpora

Abstract

Multiword terms (MWTs) are one of the main stumbling blocks in specialized translation. Translators and terminologists must master a variety of techniques to find their interlinguistic equivalents, because they are not always well described or even included in most terminological resources. Parallel texts, which are very useful for terminology extraction and knowledge acquisition, are usually employed for these purposes. They can be compiled in the form of comparable corpora, which are collections of texts originally written in the source and target languages. To make the most of them,

1 El presente trabajo se ha realizado en el marco del proyecto «Herramientas terminológicas orientadas hacia la traducción de textos medioambientales» (FFI2017-89127-P), financiado por el Ministerio de Economía y Competitividad.

translators must master advanced corpus query techniques. Unfortunately, this is not always the case. This paper presents a corpus-based methodology that facilitates the identification of English equivalents of Spanish MWTS and provides terminologists and translators with manual corpus query techniques that simplify the translation of MWTS. A comparable corpus formed by original Spanish and English texts on wind power was analyzed for the purposes of the study. We proposed a translation methodology based on distributional semantics, which entailed the obtention and comparison of contextual elements of the concepts in the source and target languages. Our results showed that this technique facilitates the identification of MWT translation equivalents in corpus.

Keywords: comparable corpus; distributional semantics; multiword term; terminology; equivalent terms; specialized translation.

Méthodologie pour la traduction des termes complexes à partir de corpus

Résumé

La qualité d'une traduction spécialisée dépend en grande partie de la bonne gestion des termes complexes. Étant donné que ceux-ci n'apparaissent pas toujours dans les ressources terminologiques et qu'ils sont décrits de manière non exhaustive, les traducteurs et les terminologues doivent maîtriser diverses techniques pour résoudre leurs difficultés. Traditionnellement, les textes parallèles ont été utilisés pour extraire la terminologie et acquérir des connaissances spécialisées. Ces textes peuvent être compilés sous la forme de corpus comparables (c'est-à-dire, des textes écrits à l'origine dans chaque langue), ce qui permet d'exploiter plus efficacement un plus grand nombre de textes. Cependant, on hésite souvent à les utiliser, probablement en raison d'un manque de connaissance des techniques de requête. Cette étude décrit une méthodologie permettant d'identifier dans un corpus comparable les équivalents anglais des termes complexes espagnols, en vue (i) de faciliter cette tâche dans le travail quotidien des traducteurs et des terminologues, et (ii) de leur fournir des techniques de requête manuelle de corpus. À cette fin, nous avons compilé un corpus comparable de textes spécialisés sur l'énergie éolienne, rédigés à l'origine en espagnol et en anglais. La méthodologie proposée s'inspire de la sémantique distributionnelle et consiste à identifier et à comparer les éléments contextuels des concepts dans les deux langues. Les résultats ont montré que cette procédure simplifie l'identification des équivalents des termes complexes dans les corpus.

Mots-clés : corpus comparable ; sémantique distributionnelle ; terme complexe ; terminologie ; termes équivalents ; traduction spécialisée.

1. Introducción

Los términos compuestos (TC) son característicos del discurso especializado (Arroyave Tobón y Quiroz Herrera, 2012; Bowker, 2015; Maniez, 2008); por ejemplo, *aerogenerador de paso de pala fijo* en el campo de la energía eólica.

Este tipo de términos destaca por los problemas que plantean, cuya adecuada gestión resulta fundamental para la producción y la traducción de textos especializados: por un lado, la identificación de los TC puede resultar compleja, ya que están constituidos por un número variable de elementos. Existen TC con dos formantes, como *central eólica*, y otros que pueden tener cinco o más elementos, como *aerogenerador de regulación pasiva por pérdida aerodinámica*. Además, sus formantes a veces pertenecen a la lengua general, lo que complica su identificación como parte del TC. Por otro lado, su análisis es complejo, debido a factores como la compactación de formantes, la desambiguación estructural, la omisión de la relación interna entre los elementos del TC o su elevada especialización. Como es de esperar, estas cuestiones también dificultan la comprensión de los TC, un requisito fundamental para la traducción.

El primer paso para traducir los TC suele ser la consulta de recursos terminológicos. No obstante, el usuario observa que, en muchas ocasiones, los TC no figuran en los diccionarios especializados y bases de datos terminológicas. Cuando lo hacen, su descripción queda lejos de ser exhaustiva y, por tanto, no responde a las necesidades de documentación terminológica (Cabezas-García y Faber, 2017). Por esta razón, los traductores deben recurrir a fuentes alternativas, como los corpus.

Desde hace tiempo, los traductores vienen utilizando *textos paralelos* para extraer terminología y adquirir el conocimiento especializado (Bowker, 2004; Gallego-Hernández, 2015).

Los textos paralelos son textos originales en la lengua meta, que siguen las mismas convenciones (tema, función, género, etc.) que el texto original y, así, proporcionan información sobre las convenciones terminológicas, fraseológicas, sintácticas, etc., de un determinado campo de especialización. No deben confundirse, por tanto, con los *corpus paralelos*, que están formados por textos en una lengua, alineados con sus traducciones en otra lengua.

En este artículo nos interesamos por los textos paralelos como fuente de documentación terminológica, que pueden compilarse en forma de corpus comparables (es decir, textos escritos originalmente en las lenguas de origen y de llegada) y permiten explotar, de manera mucho más eficiente, un gran volumen de textos. Sin embargo, a pesar de que los traductores cada vez recurren más al análisis de corpus, se observa cierta reticencia a su uso, probablemente debido al desconocimiento de las técnicas avanzadas de consulta (Bowker, 2004; Gallego-Hernández, 2015).

Por otro lado, la indudable presencia del inglés como lengua internacional de la comunicación especializada redundante en un volumen cada vez mayor de textos que deben traducirse o producirse en esta lengua, para facilitar la comunicación internacional. En este artículo describimos una metodología de consulta de corpus que permite identificar los equivalentes en inglés de TC en español, con el objetivo de: 1) facilitar su traducción y 2) proporcionar técnicas de extracción manual en corpus que los traductores puedan integrar en su labor diaria.

El procedimiento descrito no se basa en la traducción de un texto específico, sino en la propuesta de técnicas para facilitar la traducción de los TC. Para ello, compilamos un corpus comparable de textos especializados en energía eólica (p. ej. libros, informes, artículos de investigación, etc.). Dicho corpus estaba formado por

textos escritos originalmente en español e inglés, y contenía aproximadamente 3 millones de *tokens* en cada lengua.

La metodología de obtención de equivalentes se funda en la semántica distribucional y consiste en la identificación y la comparación de elementos contextuales de los conceptos en las lenguas de origen y de llegada. Las consultas en el corpus se realizaron por medio de expresiones en Corpus Query Language (CQL) —un código especial de consulta que permite realizar búsquedas avanzadas— y otras funcionalidades de la herramienta de análisis de corpus Sketch Engine (Kilgarriff *et al.*, 2004; Sketch Engine, s. f.), como la opción Filter context.

En el apartado que sigue describimos el estado de la cuestión en lo que concierne a la traducción de los TC. A continuación, realizamos un estudio de caso para ahondar en la metodología de identificación de equivalentes interlingüísticos de los TC, mediante el análisis de un corpus comparable compilado para este estudio. En la última sección presentamos las conclusiones del estudio y futuras líneas de investigación.

2. La traducción de términos compuestos. Estado de la cuestión

Los TC se hallan tanto en la lengua general como en la lengua especializada, pero se caracterizan por su elevada presencia en el discurso especializado. Han recibido diferentes denominaciones (“término compuesto”, “término poliléxico”, “compuesto nominal”, etc.), que no siempre aluden al mismo concepto. No obstante, la más extendida es la de “término compuesto” (Cabezas-García, 2020), que empleamos en este trabajo.

Este tipo de términos constituye uno de los principales problemas de traducción de los textos especializados (Daille *et al.*, 2004; Fernández Silva y Kerremans, 2011; Maniez, 2008;

Montero Martínez y García de Quesada, 2003). A ello contribuyen su complejidad estructural, semántica y cognitiva, además de los diferentes patrones de codificación de cada lengua.

Normalmente, el primer paso para decidir la traducción de un TC es consultar recursos terminológicos. Sin embargo, estos a menudo no aparecen descritos, debido a su frecuente carácter neológico o la falta de afianzamiento en una lengua (Fernández Silva y Kerremans, 2011; Sanz Vicente, 2011). Esto hace que los terminólogos y traductores, además de conocer cómo suelen codificarse dichos términos en las distintas lenguas, necesiten emplear recursos adicionales, como los corpus.

La identificación de equivalentes de los TC pasa por un exhaustivo análisis del término,² el concepto y el sistema conceptual en el que se enmarca. Gracias a este proceso de profundización conceptual, el traductor podrá proponer equivalentes distintos a la traducción literal del TC, que a menudo desemboca en errores de traducción (Daille *et al.*, 2004; Shahzad *et al.*, 2000). Por ejemplo, el equivalente del TC en castellano *perfil aerodinámico* es, en inglés, *airfoil*, un término simple. Por tanto, para traducir los TC, se deben considerar factores como la asimetría en la formación de términos en cada lengua. En las lenguas romances, como el español, los TC están formados por un núcleo, al que le siguen modificadores adjetivales (*energía nuclear*) o preposicionales (*generación de energía*). En cambio, en las lenguas germánicas, como el inglés, los modificadores preceden al núcleo y suelen ser sustantivos (*voltage drop*) o adjetivos (*electrical power*).

2 Cabe recordar la diferencia entre “concepto” (unidad de pensamiento) y “término” (designación de un concepto). Por ejemplo, el término *viento* sirve para nombrar al concepto relativo a la corriente de aire producida en la atmósfera por causas naturales.

Como señalábamos anteriormente, los TC en la lengua de origen y de llegada pueden contar con un número diferente de formantes. Cuando presentan más de dos elementos, es necesaria la desambiguación estructural o *bracketing* (Cabezas-García, 2019), mediante la cual se aclaran las dependencias internas entre los formantes (p. ej. *control del [par del generador]*). Cuando dichas dependencias no se analizan o se comprenden erróneamente, se pueden producir errores de traducción. Es el caso en *onshore wind farm*, donde *onshore* modifica a *wind farm*, como se observa en su *bracketing*: *onshore [wind farm]*. Por tanto, se alude a un parque eólico que se sitúa *onshore* y no a un tipo de viento *onshore*. Gracias a esta aclaración, se pueden proponer los equivalentes en español *parque eólico terrestre* o *parque eólico en tierra*. No obstante, una mala interpretación de las dependencias internas podría llevar a pensar que *onshore* modifica a *wind* y, con ello, el *bracketing* sería **[onshore wind] farm* y la traducción del término sería *parque de viento terrestre* o *parque de viento en tierra*, lo cual constituye claramente un error de traducción. Además, cuando existe más de un equivalente de traducción, se debe elegir la variante terminológica más adecuada al contexto (Cabezas-García, 2020).

Por otra parte, el concepto de *equivalencia* resulta ineludible al abordar cuestiones de traducción. Sin embargo, la traducción y la terminología persiguen objetivos distintos; por ello, a menudo entienden la *equivalencia* de modos diferentes. Por un lado, los terminólogos suelen centrarse en la equivalencia entre términos (entendidos como unidades que representan el mismo concepto), pues su objetivo es a menudo la compilación y la descripción de términos en recursos, o la normalización del discurso especializado. Por otro, en cambio, los traductores buscan la correspondencia a nivel oracional o textual. Sus traducciones pretenden ser funcionales y, por tanto, cumplir con su objetivo en

la lengua y la cultura de llegada (Nord, 1997; Reiss y Vermeer, 1984), velando siempre por la precisión y adecuación del texto y, con él, su terminología (Cabré, 2003). Así, a nivel oracional o textual, cuando no se dispone de equivalentes terminológicos, se pueden considerar no solo aquellas unidades o expresiones que tienen estrictamente el mismo significado, sino también otras variantes terminológicas, como los hiperónimos. Ello explica que la noción de *equivalencia* sea más amplia para los traductores que para los terminólogos. A pesar de que los equivalentes en forma de hiperónimos o variantes terminológicas con distintas conceptualizaciones no se incluyen normalmente en los recursos terminológicos, suelen ser válidas a nivel textual.

La complejidad de la equivalencia también radica en los diferentes niveles en los que puede observarse, como en la forma del mensaje, su significado, su estilo, su función o su efecto (Reiss y Vermeer, 1984). En este artículo nos centramos en la equivalencia entre términos (Arntz y Picht, 1995; Felber, 1984; Wüster, 1979) en un contexto de traducción. Para establecer dichas equivalencias, se debe determinar el grado de correspondencia conceptual entre los posibles equivalentes (Sanz Vicente, 2011).

No obstante, no siempre se da una equivalencia total. Por este motivo, Felber (1984) y Arntz y Picht (1995) señalaron diferentes grados de equivalencia: 1) equivalencia total (coinciden todas las características conceptuales); 2) intersección (coinciden la mayoría de características conceptuales); 3) inclusión (el concepto en la lengua de llegada es más amplio y abarca al concepto en la lengua de origen), y 4) falta de equivalencia (el concepto no existe en la lengua de llegada). Estas equivalencias se basan en distintos grados de correspondencia conceptual (Arntz y Picht, 1995; Felber, 1984). Pese a ello, como hemos mencionado anteriormente, al traducir un texto se puede adoptar una visión más amplia de la equivalencia.

Volviendo a la traducción de los TC, se observan dos líneas de investigación: 1) la comparación y el análisis de TC en dos o más lenguas, desarrollada principalmente en la terminología (Cabezas-García, 2019; Fernández Silva y Kerremans, 2011; Salager-Meyer, 1985, *inter alia*); y 2) la extracción terminológica bilingüe, llevada a cabo especialmente desde la lingüística computacional (Daille *et al.*, 2004; Déjean y Gaussier, 2002; Shahzad *et al.*, 2000, *inter alia*).

En la primera vertiente, encontramos estudios como el de Salager-Meyer (1985), quien compara la sintaxis y la semántica de TC en inglés y español. Establece las principales estructuras morfológicas de los TC en inglés y sus correspondencias habituales en español (p. ej. N1+N2 > N2+*de*+N1). Igualmente, señala que no se puede generalizar en torno a la traducción de los TC, pues esta depende, en gran medida, del traductor y del contexto de la traducción.

Por otro lado, Montero Martínez y García de Quesada (2003) emplean un corpus comparable en inglés y español para analizar los TC con vistas a su traducción. Exploran las diferencias y similitudes entre los equivalentes terminológicos y defienden que los patrones semánticos suelen coincidir en estas dos lenguas.

Además, Fernández Silva y Kerremans (2011) investigan la variación terminológica en textos escritos en gallego y en sus traducciones al inglés. Aunque su objeto de estudio no son los TC, gran parte de los términos analizados pertenecen a esta tipología. En su estudio, crean un índice de variación interlingüística, para reflejar la distancia entre los equivalentes.

Sanz Vicente (2011) investiga la traducción de los TC del inglés al español. Emplea un corpus comparable para analizar la influencia de la primera lengua en los equivalentes en la segunda y, a continuación, los clasifica según

el procedimiento empleado (préstamo, calco, paráfrasis, adaptación, etc.). Logra identificar alrededor del 80 % de los equivalentes de los TC en dicho corpus. Asimismo, Sanz Vicente (2011) realiza una extracción terminológica bilingüe, por lo que retomaremos su trabajo cuando abordemos la segunda corriente de estudio.

Por su parte, Tercedor Sánchez *et al.* (2013) también exploran la traducción de los TC del inglés al español, prestando especial atención al uso de calcos. Subrayan que se suele recurrir a ellos cuando se dispone de poco tiempo para traducir, ya que se consideran una opción válida. No obstante, cuando los traductores cuentan con más tiempo para sopesar varias posibilidades de traducción, se suelen decantar por otras opciones.

Por último, Maniez (2013) estudia la traducción al francés de los adjetivos compuestos del inglés (p. ej. *decay-corrected*) y defiende que los recursos como Linguee deben complementarse con consultas en corpus paralelos y comparables.

La segunda corriente de estudio, la extracción terminológica bilingüe, se ha investigado principalmente desde la lingüística computacional, debido a su importancia para los sistemas de traducción automática. En este enfoque, los corpus, ya sean paralelos o comparables, resultan fundamentales para la extracción de equivalentes. De este modo, se deben implementar técnicas de consulta de corpus para obtener resultados precisos.

La *semántica distribucional* se emplea a menudo en corpus comparables para establecer equivalencias (Daille *et al.*, 2004; Daille y Morin, 2005; Déjean y Gaussier, 2002; Shahzad *et al.*, 2000). Esta se basa en la descripción del significado de un término en función de su

distribución en un conjunto de contextos. De este modo, la coincidencia contextual indica la semejanza de significados, como sucede en la sinonimia (Déjean y Gaussier, 2002, p. 4). Los equivalentes de traducción se extraen mediante un análisis del contexto del término en la lengua de origen, que se suele representar por medio de vectores. Se entiende que el análisis del mismo contexto en la lengua de llegada permite descubrir correspondencias interlingüísticas, puesto que estas comparten los mismos contextos (Daille *et al.*, 2004; Daille y Morin, 2005; Déjean y Gaussier, 2002; Shahzad *et al.*, 2000).

Como apuntan Shahzad *et al.* (2000), los TC no se traducen siempre de forma literal. Estos autores utilizan un corpus comparable, en el que aplican un algoritmo que se encarga de traducir el contexto del TC a la lengua de llegada, con vistas a obtener posibles equivalentes. Déjean y Gaussier (2002) también analizan la coincidencia de los elementos contextuales en un corpus comparable inglés-alemán, con el fin de hallar coincidencias de significado. Su objetivo principal consiste en extraer traducciones de términos simples y TC mediante la identificación de elementos contextuales de cada formante. Sin embargo, solo emplean los vectores que coinciden en los contextos de todos los formantes del TC. Cuando se identifican varias equivalencias posibles, se selecciona la que presenta una mayor semejanza conceptual. Estos estudios preliminares sobre la traducción de los TC sentaron las bases para nuevas investigaciones, como la que presentamos en este artículo.

Daille *et al.* (2004) y Daille y Morin (2005) aplican técnicas de la semántica distribucional para extraer los equivalentes de términos simples y TC en inglés, francés y español. Inicialmente, se seleccionan los términos en cada corpus y se agrupan las variantes terminológicas. Tras comprobar la correlación de sus elementos contextuales, se alinean los

términos en cada lengua, con sus correspondientes traducciones. En estos trabajos, Daille *et al.* (2004) y Daille y Morin (2005) tuvieron en cuenta características específicas de los TC, como su falta de composicionalidad o las diferencias estructurales con otros TC y sus traducciones.

Asimismo, Maniez (2008) propone soluciones manuales para los traductores, como el uso de Google para encontrar los equivalentes en francés de los TC del inglés. Para ello, emplea los formantes del TC cuya traducción se conoce, junto con operadores como el asterisco. Por ejemplo, para traducir *gluten-sensitive enteropathy*, realizó la consulta “entéropathie * gluten”, que arrojó los siguientes resultados: *entéropathie au gluten*, *entéropathie d’intolérance au gluten*, *entéropathie induite par le gluten* y *entéropathie de sensibilité au gluten* (Maniez, 2008, p. 163). Las contribuciones de este autor resultan de indudable interés para los traductores y terminólogos, gracias al desarrollo de técnicas de extracción manual que pueden aplicar en sus tareas cotidianas. Asimismo, el trabajo traductológico y terminológico puede beneficiarse de otras consultas avanzadas en corpus.

También, mediante técnicas de extracción manual, Sanz Vicente (2011, p. 457) identifica los equivalentes del 80 % de los TC del inglés de su muestra, empleando estrategias como las siguientes: 1) generar listas de los TC en inglés, para detectar posibles anglicismos en español; 2) consultar las posibles traducciones en español del TC o de sus formantes; 3) producir listas de los elementos que suelen combinarse con el TC, y 4) buscar los términos o raíces que se escriben de forma similar en ambas lenguas.

Arroyave Tobón y Quiroz Herrera (2012) describen técnicas para la traducción de los TC del inglés al español. Estas se basan en la correcta identificación del TC y de su relación interna. Seguidamente, observan la traducción al español

de otros términos similares y proponen una traducción para el TC en cuestión.

Harastani *et al.* (2013) investigan la traducción al inglés de TC del francés cuyo núcleo está modificado por un adjetivo relacional (p. ej. *cancer pulmonaire*). Tras sustituir este adjetivo por el sintagma preposicional equivalente (*cancer du poumon*), traducen el TC de forma literal e investigan su ocurrencia en un corpus comparable en inglés. Aunque alcanzan una precisión del 86 %, este enfoque no será aplicable a muchos TC, ya que todos los modificadores adjetivales no pueden transformarse en un sintagma preposicional (p. ej. en español se dice *freno hidráulico*, pero no **freno de agua*).

Como puede observarse, la lingüística computacional ha realizado significativas contribuciones al campo de la extracción terminológica bilingüe, aunque estas técnicas computacionales no siempre resultan de fácil aplicación a las tareas prácticas de traducción y terminología. No obstante, es indudable que dichas técnicas pueden ser de gran utilidad en la identificación de equivalentes de traducción. En este sentido, Pimentel (2013) subraya la importancia de la coincidencia de significados, el contexto colocacional, el conocimiento enciclopédico y el *tertium comparationis* en la búsqueda de equivalentes por parte de traductores y terminólogos.

La *coincidencia de significados* se suele analizar comparando sistemas conceptuales. Aunque estos sistemas deben ser independientes de la lengua, los traductores a menudo los elaboran en la lengua de origen y, a continuación, los comparan con el sistema desarrollado en la lengua de llegada.

El *contexto colocacional* también ayuda a identificar equivalentes, especialmente en el caso de términos polisémicos, como *bunch* (p. ej. *bunch of flowers* [IN] > *bouquet* [FR]; *bunch of hair* [IN] > *touffe* [FR]) (Pimentel, 2013).

El *conocimiento enciclopédico* determina las consultas que pueden realizarse para localizar el equivalente: desde búsquedas específicas, cuando se intuye la posible correspondencia, hasta consultas más amplias, cuando no se dispone de indicios.

Por último, mediante el *tertium comparationis*, los traductores y terminólogos comparan el uso y los patrones colocacionales de los posibles equivalentes con los del término en la lengua de origen (Pimentel, 2013).

Como se desprende de los estudios presentados, la traducción de los TC no es una tarea sencilla, debido a las características de estos términos en las lenguas de origen y de llegada, y a la frecuente necesidad de aplicar técnicas de corpus para trazar equivalencias de forma satisfactoria. Como describimos en este trabajo, dichas técnicas permiten identificar equivalentes terminológicos, que pueden responder a problemas particulares de traducción o bien incluirse en recursos especializados.

3. Metodología para la traducción de los términos compuestos: un estudio de caso en el campo de la energía eólica

En esta sección describimos una metodología que permite identificar los equivalentes interlingüísticos de los TC mediante la consulta de corpus e ilustramos dicha metodología a partir de un estudio de caso. El procedimiento presentado se aplicó a la traducción de TC del español al inglés, si bien puede emplearse en cualquier lengua y direccionalidad, pues se basa en cuestiones semánticas y pragmáticas independientes del idioma.

Las técnicas propuestas se inspiran en la semántica distribucional, según la cual el significado de un término puede abstraerse a partir de su distribución en varios contextos. De tal modo, la coincidencia de contextos indica la

semejanza semántica entre variantes terminológicas de una misma lengua y entre equivalentes interlingüísticos. En este estudio se analizaron dos tipos de contexto: 1) el cotexto o contexto lingüístico inmediato de un término, y 2) el contexto oracional. Se establecieron correspondencias entre términos, pues esta investigación adopta una perspectiva terminológica de la equivalencia, en la que las características conceptuales de los posibles equivalentes deben ser semejantes. Nuestro objetivo no es, por tanto, la traducción de un texto concreto, en cuyo caso podrían aplicarse técnicas más amplias, como el uso de hiperónimos.

Se analizó un corpus comparable de textos especializados sobre energía eólica, formado por 3 618 296 *tokens* en español (Corpus Eólica ES) y 3 619 300 *tokens* en inglés (Corpus Eólica EN). El corpus estaba compuesto de textos especializados (informes, libros, artículos de investigación, etc.) en el campo de la energía eólica, extraídos de revistas, libros y sitios web específicos de este dominio temático. Se trata de textos publicados entre los años 2000 y 2019, y pertenecientes a distintas variedades del español y del inglés, pues este estudio no persigue criterios diacrónicos ni diatópicos.

Los corpus comparables, redactados originalmente en la lengua de estudio, se recomiendan en las investigaciones terminológicas y traductológicas (Bowker, 2015), pues las traducciones presentes en los corpus paralelos pueden contener expresiones no idiomáticas, construcciones que no son propias de una lengua o incluso errores. Además, aunque existen corpus paralelos de temática especializada (Arce Romeral y Seghiri, 2018; Pérez Carrasco y Seghiri, 2021), estos son menos frecuentes que los corpus comparables, que existen en muchos casos y, además, pueden crearse más rápidamente.

El análisis se realizó mediante la herramienta de gestión de corpus Sketch Engine (Kilgarriff *et al.*, 2004; Sketch Engine, s. f.). Para ilustrar la metodología propuesta, se emplea un TC habitual del discurso de la energía eólica y la energía eléctrica: *tensión en la etapa de continua*. Este término resulta de interés por su especialización y opacidad semántica, que pueden complicar su traducción.

El primer paso del procedimiento es el análisis conceptual, cuya profundidad puede variar en función de la tarea que se realice y el tiempo del que se disponga. Así, se pueden elaborar sistemas conceptuales *ad hoc* para tareas concretas de traducción o bien sistemas más profundos para labores terminológicas. En cualquier caso, este análisis inicial del dominio conceptual es importante, porque permite a los traductores o terminólogos adquirir el conocimiento especializado que se necesita para la traducción de los textos especializados (Hurtado Albir, 2011).

El siguiente fragmento es tomado de un texto especializado extraído del Corpus Eólica ES.

En la Figura 4.9 se muestra la evolución de la tensión de la etapa de continua del convertidor de frecuencia cuando se produce el hueco de tensión lento y tanto los aerogeneradores individuales como las máquinas equivalentes de los modelos agregados cuentan con una resistencia de frenado como sistema de protección a sobretensiones en la etapa de continua.

En este fragmento pueden observarse algunas de las relaciones del sistema conceptual. Por ejemplo, se aprecia que la *etapa de continua* es una parte integrante de los convertidores de frecuencia, que son máquinas que cambian la frecuencia de una corriente alterna. Además, los huecos de tensión afectan a la tensión de la etapa de continua, que puede controlarse mediante la instalación de una resistencia de frenado en los aerogeneradores.

No obstante, no siempre resulta sencillo encontrar textos como el del fragmento citado, que aclaren parte del sistema conceptual. En esos casos, pueden emplearse patrones de conocimiento, es decir, estructuras léxico-sintácticas que expresan las relaciones semánticas de forma explícita en el lenguaje natural (Meyer, 2001). Por ejemplo, para descubrir los conceptos a los que afecta la *tensión en la etapa de continua*, se puede buscar en el corpus este término junto con patrones de conocimiento que expresen la relación *afecta*, como en el siguiente ejemplo: *tensión en la etapa de continua afecta a/ tiene efectos sobre/ influye en/ actúa sobre*, etc. Este tipo de análisis permite obtener una panorámica general del campo especializado.

A continuación, se debe profundizar en el significado del propio término (Cabezas-García, 2019). En el caso de los TC, es recomendable aclarar primero las dependencias estructurales que se dan en el interior del TC, en un procedimiento que se conoce como *bracketing* (Cabezas-García, 2019) y que permite delimitar las dos partes esenciales del TC: el núcleo y el modificador. En el caso de *tensión de la etapa de continua*, el *bracketing* sería [*tensión*] [*de la etapa de continua*].

Una vez que se dispone de esta estructura de núcleo y modificador, se pueden llevar a cabo consultas de patrones de conocimiento o de paráfrasis, que aclaran el significado del TC. Las *paráfrasis* son elementos contextuales, como verbos o preposiciones, que unen al núcleo con el modificador, como ocurre en *aerogenerador marino* > *aerogenerador situado en/ ubicado en/ instalado en* el mar, de forma que proporcionan más detalles sobre el significado del TC (Cabezas-García, 2017). Así, se observó que la *tensión en la etapa de continua* es la tensión que presenta un circuito que forma parte de los convertidores de frecuencia, denominada *etapa de continua* o *bus de corriente continua*,

entre otras formas, y cuya función es rectificar una corriente alterna en corriente continua que después se suministra a un inversor.

Tras conocer el significado del TC, el procedimiento de traducción parte de la identificación de elementos contextuales que sean característicos del término en la lengua de origen. Estos elementos contextuales se traducirán y se emplearán en las consultas en el corpus en inglés, pues la coincidencia de contextos suele revelar términos equivalentes en las lenguas de origen y de llegada. Estas pistas conceptuales pueden ser elementos del concepto (p. ej. partes de la definición), formantes del TC, unidades léxicas que suelen coocurrir con el término, conceptos que codifiquen una relación directa con este, cantidades específicas, nombres de leyes, nombres propios, etc. Es decir, cualquier tipo de término o información que suela aparecer junto al TC en la lengua de origen o incluso que forme parte de él.

Dichos elementos deben aparecer en diferentes textos del corpus; de lo contrario, podrían aludir a cuestiones propias de un único texto.

No existe un número fijo de elementos contextuales que deban seleccionarse. Por un lado, si se incluye un gran número, hay más probabilidades de encontrar ocurrencias del posible equivalente, pero también más ruido. Por otro, un número reducido de elementos puede producir silencio. Por tanto, la clave consiste en identificar aquellos que codifican una relación fuerte y frecuente con el TC, para así garantizar la obtención de ocurrencias, además de aquellos cuyo vínculo con el TC es distintivo o específico. Por supuesto, en estas consultas influye el conocimiento enciclopédico de la persona que las realice, que determinará el número y tipo de elementos que se incluyan.

Tras analizar las concordancias de *tensión en la etapa de continua*, se seleccionaron los siguientes elementos contextuales: *convertidor*, *condensador*, *UDC*, *VDC*, *fluctuación* y *resistencia de frenado*. Estos se observaron frecuentemente junto al término, con el que codificaban relaciones directas y características: 1) la etapa de continua es un componente de algunos *convertidores*; 2) el aumento de la capacidad del *condensador* disminuye la variación de la tensión en la etapa de continua, que siempre se intenta controlar; 3) *UDC* y *VDC* son las siglas con las que se alude a este tipo de tensión; 4) *fluctuación* es un término que suele coocurrir con el TC en cuestión, pues a menudo se habla de *fluctuaciones en la tensión en la etapa de continua*; y 5) la *resistencia de frenado* evita que la tensión de la etapa de continua alcance valores no deseados.

En la Figura 1 se presenta una muestra de las concordancias en las que estos elementos contextuales coocurren con *tensión en la etapa de continua* y sus variantes terminológicas (p. ej. *tensión en el bus de continua*, *tensión del bus DC*).

A este análisis en la lengua de origen le sucedió la traducción de los elementos contextuales, para así poder emplearlos en las búsquedas en inglés. Para ello, consultamos recursos terminológicos como TERMIUM Plus (Gobierno de

Canadá, 2021) e Interactive Terminology for Europe (IATE) (2021), dada su naturaleza multilingüe y la descripción de términos especializados. De este modo, los elementos contextuales que empleamos en inglés fueron los siguientes: *converter* (convertidor), *capacitor* (condensador), *UDC* y *VDC* (ambas siglas se suelen emplear en las dos lenguas), *fluctuation* (fluctuación) y *brake resistor* (resistencia de frenado).

Para localizar los equivalentes de los TC, partimos de que se conoce la traducción de parte del TC (ya sea el núcleo o el modificador) o bien que esta ha podido identificarse en recursos terminológicos o en TC similares. La búsqueda de equivalentes de los TC que proponemos se basa en la consulta combinada de este formato conocido y los elementos contextuales. Sin embargo, no basta con la simple coocurrencia de estos términos. Deben estar relacionados del mismo modo que lo hacen en la lengua original, ya que los términos de consulta también pueden coocurrir en otros contextos, expresando relaciones distintas que no son las que se dan en el concepto que nos interesa. De este modo, se espera que, junto con la parte del TC cuya traducción se conoce, se encuentre el resto de la traducción del TC. En el caso de *tensión en la etapa de continua*, *tensión* suele traducirse por *voltage* en inglés. La consulta en Sketch

cargo, esta solución causa mayores fluctuaciones de	tensión en la etapa de continua	cuando se perciba un hueco de tensión. </s><s> Conc
capacidad del condensador disminuye la variación de la	tensión en el bus de continua	. </s><s> Igualmente, en Belchior et al. </s><s> (2003,
Este consumo de potencia desde el rotor, hace que la	tensión del bus DC	del convertidor descienda demasiado y la máquina pu
s> En el caso de que los buses estén separados y las	tensiones de los buses de continua	de cada convertidor sean diferentes, la tensión impue
del valor de cada una de las fases y (Vdc) el valor de	tensión en el bus de corriente continua	. </s><s> 1.5.2. </s><s> Modulación por anchura de p
as tensiones de línea. </s><s> Dada una determinada	tensión en la etapa de continua	, UDC , las tensiones de línea (uab, ubc, uca) y las tens,
según la siguiente ecuación, </s><s> donde udc es la	tensión de la etapa de continua	, C es el valor del condensador, Idc-R es la corriente de
imiento del generador. </s><s> • Mantener constante la	tensión de la etapa de continua	del convertidor de frecuencia. </s><s> • Transferir la p
cumulada en el condensador , con lo que evita que la	tensión de la etapa de continua	alcance valores no deseados y facilita la recuperación
neutro máxima que puede generar es del 35,4 % de la	tensión del bus de continua	, Vdc , o bien, considerando un rectificador ideal, del 8
a etapa chopper se realiza a partir de la medida de la	tensión en bus DC	(vDC). </s><s> La energía generada en exceso, es de

Figura 1. Elementos contextuales de *tensión en la etapa de continua*. Fuente: Corpus Eólica ES.

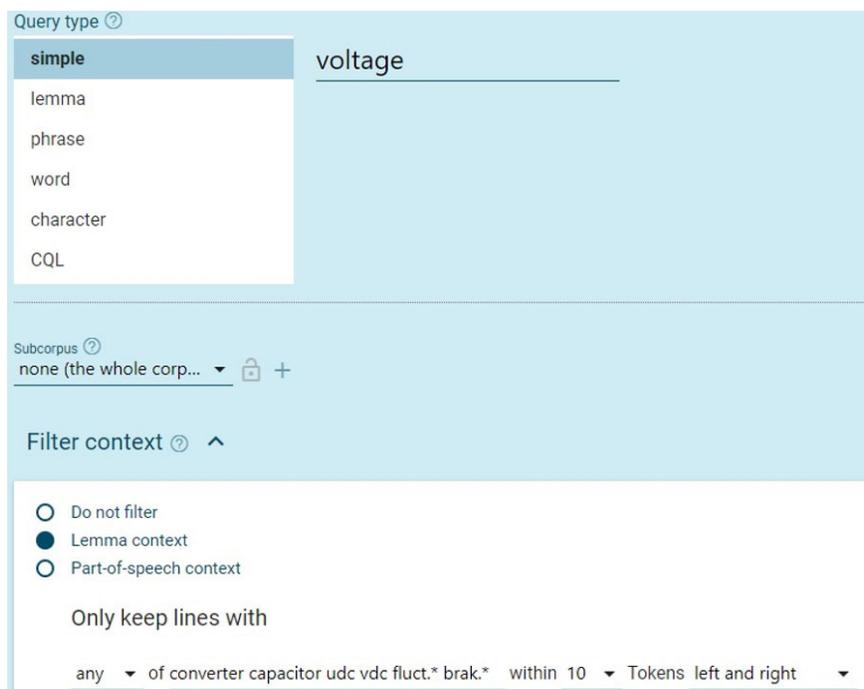


Figura 2. Consulta de los elementos contextuales de *tensión en la etapa de continua* en el corpus en inglés. Fuente: Corpus Eólica EN.

Engine fue, por tanto, la que se muestra en la Figura 2.

Para realizar una consulta combinada, empleamos dos métodos de búsqueda disponibles en la opción *Concordance* de Sketch Engine: 1) la búsqueda simple o en CQL, y 2) el filtro de contexto (*Filter context*).

La búsqueda simple o en CQL puede utilizarse para introducir la parte del TC que se conoce, que aparecerá en el centro de las concordancias si se selecciona el modo de visualización *Key Word in Context* (KWIC). De esta forma, los posibles equivalentes del TC se mostrarán también en el centro de las concordancias.

La búsqueda simple o en CQL puede usarse igualmente para introducir los elementos contextuales. Sin embargo, en ese caso, estos se agruparían en el centro de las líneas de concor-

dancia, lo cual complicaría la identificación de los equivalentes, de modo que en este estudio no empleamos esta opción.

Por otro lado, la búsqueda simple permite incluir uno o varios términos consecutivos, mientras que la búsqueda en CQL ofrece opciones más avanzadas, como el empleo de disyunciones para buscar uno u otro término, sin la presencia necesaria de ambos. Dado que la parte del TC que conocíamos estaba formada únicamente por un término (*voltage*), utilizamos la búsqueda simple.

Para indicar los elementos contextuales hicimos uso de la función *Filter context*, dentro de la cual seleccionamos la opción *Lemma context*. En ella, cuando los elementos se separan con espacios, se identifican como lemas diferentes.

Asimismo, Sketch Engine permite hacer búsquedas truncadas, cuando se desea obtener,

tem in the IFSA microgrid by utilizing the measurement of the dc-link	voltage	of the converter	and its derivatives, not by the system frequency.
DFIG is very non-linear.	voltage	is almost kept constant when the DC braking	chopper operates.
he capacitor current (Icap), the converter power (Pconv), and DC bus	voltage	(Udc) can be calculated by 61–64, respectively.	The control b
DC-link voltage control loop [30,31].	voltage	udc is compared with a constant reference value and the error is proc	
monitors the DC link voltage and orders the converter to block if the DC	voltage	exceeds a given range.	Third, the damping characteristics of
irce converter (NSC) [14].	voltages	of generator-rectifier units by varying the switching ratio so that the o	
grid, which leads to significantly increased fluctuations of the dc-link	voltage	.	In order for the DFIG based WTs to fulfill the low-voltage ridi
decoupled control.	voltage	of the back-to-back converter and reactive power.	The rotor \
ie MTDC system applies voltage droop control.	voltage	control can be shared among different converters with voltage droop	
g duration.	voltage	also experience high fluctuations, resulting in DFIG to be disconnect	
nk capacitor can be given as follows, where vdc and C are the dc-link	voltage	and capacitance, respectively.	Also, Pb and Pg are the avera
7 m/s and is kept at this speed until 2.5 s. The initial dc-link capacitor	voltages	are chosen differently with Vc1 (0) = 350 V and Vc2 (0) = 400 V to inv	
rents, the grid-side current iag and voltage eag of phase a, the dc-link	voltage	Vdc and the capacitor voltages Vc1 and Vc2 are plotted for the three	
approach is to control the generator converter to maintain the DC link	voltage	at a constant value and then use the network converter to control the	

Figura 3. Muestra de las concordancias obtenidas en la búsqueda de equivalentes en inglés de tensión en la etapa de continua. Fuente: Corpus Eólica EN.

por ejemplo, distintas terminaciones de una raíz (p. ej. *fluct.* > fluctuate, fluctuation*). Estas formas constituyen palabras diferentes y no se obtendrían, por tanto, con la búsqueda por lemas. Esta función puede resultar de gran utilidad, dado que, en ocasiones, se producen cambios en la categoría gramatical de los equivalentes en dos lenguas. Por ello, la utilizamos en el caso de *fluctuation* y *brake resistor*, en los que indicamos *fluct.** y *brak.**, para permitir un mayor margen de resultados derivados de estas raíces.

Al seleccionar la opción *Filter context*, es posible elegir si se desean buscar todos (*all*), alguno (*any*) o ninguno (*none*) de los elementos introducidos. Seleccionamos *any* para indicar que no era necesario que todos los términos figurasen a la vez con *voltage*, sino que era suficiente con su aparición por separado.

Del mismo modo, la ventana contextual puede abarcar 1, 2, 3, 4, 5, 7, 10 o 15 *tokens* a la izquierda, a la derecha o a ambos lados del término de partida (*voltage*). En este caso, seleccionamos una ventana contextual de 10 elementos a ambos lados del término. En la Figura 3 se presentan algunas de las concordancias obtenidas, entre las que identificamos posibles equivalentes de *tensión en la etapa de continua*.

Gracias a este análisis distribucional, se obtuvieron varios TC en inglés que podrían corresponder a *tensión en la etapa de continua*: *DC-link voltage*, *DC bus voltage*, *DC voltage*, *DC capacitor voltage* y *DC-link capacitor voltage*, además de distintas variantes ortográficas de los mismos. Como se observa, se trata de formas que guardan similitud entre sí y que podrían ser equivalentes en inglés de *tensión en la etapa de continua* o quizás de otros términos.

Para determinar la correspondencia conceptual entre estas posibles equivalencias y *tensión en la etapa de continua*, llevamos a cabo un análisis conceptual y contextual más profundo. Este arrojó concordancias desambiguadoras (véase Tabla 1), que permitieron descartar los términos *DC voltage*, *DC-link capacitor voltage* y *DC capacitor voltage* como equivalentes de *tensión en la etapa de continua*.

Como se aprecia en el ejemplo 1 de la Tabla 1, *DC voltage* alude a un tipo de tensión, la tensión continua, pues se muestra en oposición a la tensión alterna (*AC voltage*). No se trata, por tanto, de la tensión presente en un punto específico (la etapa de continua). Además, el ejemplo 2 aclara que se trata de conceptos diferentes, que además se designan con símbolos distintos.

Tabla 1. Ejemplo de concordancias desambiguadoras

1. A DC/AC inverter converts the DC voltage of the energy storage system to an AC voltage to drive the electric machine.
2. V_d is the diode rectifier output DC voltage, and v_{dc} is the DC-link voltage, as the boost converter output voltage.
3. The initial DC-link capacitor voltages are chosen differently with $V_{c1}(0) = 350$ V and $V_{c2}(0) = 400$ V to investigate the DC-link voltage.
4. In Fig. 6, the corresponding machine-side currents, the grid-side current i_{ag} and voltage e_{ag} of phase a, the DC-link voltage V_{dc} and the capacitor voltages V_{c1} and V_{c2} are plotted for the three different control schemes.

Fuente: Corpus Eólica EN.

Por otro lado, el ejemplo 3 refleja que *DC-link capacitor voltage* y *DC-link voltage* son conceptos diferentes. Igualmente, en el ejemplo 4 se percibe que *capacitor voltage* y *DC-link voltage* tampoco aluden al mismo concepto.

En concreto, *capacitor voltage* se designa con el mismo símbolo que *DC-link capacitor voltage* (V_c), de modo que parece tratarse de variantes terminológicas.

DC capacitor voltage también parece ser otra denominación posible de estos términos, si bien solo presenta una ocurrencia en un único texto, por lo que descartamos esta opción, al tratarse probablemente de un uso particular de un autor.

Por tanto, las correspondencias de *tensión en la etapa de continua* que se obtuvieron fueron *DC-link voltage* y *DC bus voltage*, con sus correspondientes variantes ortográficas, que alternan el uso del guion y de las mayúsculas en DC. Se establecieron estas correspondencias al comprobar que los términos en las lenguas de origen y de llegada aludían al mismo concepto. Asimismo, los términos en la lengua de llegada aparecieron frecuentemente junto a los elementos contextuales, con los que codificaban la misma relación que se observó en la lengua de origen.

En el corpus en inglés, se indica la pertenencia de la etapa de continua (*DC link*) a un convertidor (*converter*), se menciona la importancia de los sistemas de frenado (*DC braking chopper*) y del condensador (*capacitor*) para controlar la *tensión en la etapa de continua*, se mencionan las fluctuaciones de este tipo de tensión (*fluctuations of the DC-link voltage*) y se emplean las siglas que también son frecuentes en español (V_{DC} y U_{DC}).

Con el fin de validar estas correspondencias, es preciso comprobar que no se emplean únicamente en un texto o por un mismo autor. Para ello, analizamos las concordancias de estos términos sin incluir los elementos contextuales. Gracias a estas concordancias adicionales, confirmamos que los términos en inglés se utilizan para nombrar el mismo concepto que *tensión en la etapa de continua*.

En conclusión, en la Tabla 2 se presentan los términos que designan la *tensión en la etapa de continua* en español y en inglés, junto con la frecuencia de uso de cada variante. No obstante, conviene señalar que en las traducciones no se empleará siempre la variante más frecuente, sino la que mejor se adapte al contexto y a la situación comunicativa.

En la Tabla 3 se presenta un resumen de la metodología descrita, en el que se distinguen las tres etapas de la búsqueda de equivalentes: 1) análisis conceptual, 2) traducción y 3) validación de equivalentes en la lengua de llegada, así como los pasos internos de cada una de estas tareas.

Por medio del análisis de *tensión en la etapa de continua* se aprecia que, a menudo, la similitud formal de los TC no indica semejanza semántica, como hemos observado con las concordancias desambiguadoras. Para evitar errores, su traducción debe reposar sobre un profundo análisis conceptual y contextual (véase Tabla 1), como hemos presentado en este artículo.

Tabla 2. Términos en español e inglés que designan la tensión en la etapa de continua

	Término	Número de ocurrencias
En español	1. tensión en la etapa de continua	33
	2. tensión de la etapa de continua	23
	3. tensión de bus dc	46
	4. tensión de bus de continua	27
	5. tensión de bus de corriente continua	14
En inglés	6. dc-link voltage	66
	7. DC-link voltage	54
	8. DC link voltage	32
	9. DC link voltage	8
	10. DC bus voltage	25
	11. DC bus voltage	9

Tabla 3. Resumen de la metodología propuesta

1. Análisis conceptual
 - 1.1 Análisis del dominio conceptual
 - 1.2 Análisis del concepto
 - 1.2.1 Desambiguación estructural del TC (*bracketing*)
 - 1.2.2 Análisis de la relación semántica interna
2. Traducción
 - 2.1 Identificación de elementos contextuales en la lengua origen
 - 2.2 Traducción de los elementos contextuales a la lengua de llegada
 - 2.3 Consulta combinada en corpus: elementos contextuales y formante conocido del TC
 - 2.4 Análisis conceptual y contextual de los posibles equivalentes en la lengua de llegada
3. Validación de equivalentes en la lengua de llegada

A pesar de que este procedimiento solo permite obtener equivalentes que también constituyen TC en la lengua de llegada, no se trata de una limitación, ya que la mayoría de TC del español se traducen también por TC en inglés, y viceversa (Sanz Vicente, 2011). En cambio, este procedimiento es recomendable, porque permite obtener correspondencias que no cuentan necesariamente con el mismo número de formantes que el término en la lengua de origen. Esto reviste una especial importancia, pues los TC no siempre se traducen de forma literal (p. ej. *tensión de bus de corriente continua* > *DC-link voltage*).

Para llevar a cabo este procedimiento es necesario conocer el equivalente de alguno de los

formantes del TC en la otra lengua, algo que, por otro lado, suele ocurrir cuando nos enfrentamos a la traducción de estos términos. En efecto, es habitual que se conozca la traducción de parte del TC, pero no la forma completa del término en la lengua de llegada. En cualquier caso, este procedimiento debe complementarse con el análisis de las concordancias de los posibles equivalentes. Además, este segundo análisis de concordancias permitirá localizar otras variantes terminológicas que no se hubieran obtenido, como variantes monoléxicas o aquellas que no incluyan el formante del TC con el que se inició la búsqueda (p. ej. *voltage*). En otras palabras, mediante esta metodología se obtienen equivalentes que difieren formalmente del término en la lengua de origen. Este

dato es significativo, pues la variación terminológica es un aspecto propio de los TC (Fernández Silva y Kerremans, 2011), que puede dificultar tanto la comprensión previa como la búsqueda de equivalentes.

Por otra parte, a pesar de que se genera cierto ruido, esta técnica proporciona equivalentes de traducción válidos en casos en los que los recursos terminológicos no muestran resultados. Se trata, por tanto, de un procedimiento aplicable en la labor cotidiana de traductores y terminólogos, que subraya la utilidad de los corpus en las tareas de trasvase lingüístico.

4. Conclusiones

Los corpus ofrecen múltiples posibilidades, como se ha demostrado en este artículo, en el que se presenta una metodología que facilita la traducción de los TC mediante la semejanza contextual en las lenguas de origen y de llegada. Las técnicas de corpus simplifican la traducción de los TC y permiten evitar errores de traducción, que a menudo se derivan de una falta de información o del escaso conocimiento del campo especializado. Además, estas técnicas de corpus compensan la descripción poco sistemática de los TC en los recursos terminológicos y resultan de utilidad para los terminólogos y traductores.

La metodología propuesta incluye técnicas de la semántica distribucional, adaptadas para su uso por parte de traductores o terminólogos. Dicha metodología, que ilustramos por medio de un estudio de caso en el campo de la energía eólica, incluye una fase de análisis conceptual, otra fase de traducción y un paso final de validación de correspondencias interlingüísticas.

Inicialmente, se proponen técnicas de corpus para el análisis conceptual del dominio (por ejemplo, mediante el empleo de patrones de conocimiento). A continuación, se debe analizar

el concepto en cuestión, que debe desambiguarse estructuralmente para poder explorar su relación interna (por medio de patrones de conocimiento o paráfrasis). Una vez aclarado el significado del concepto, la obtención de equivalentes interlingüísticos requiere, en primer lugar, de la identificación de elementos contextuales del término en lengua origen, que después se traducirán y se emplearán como elementos de búsqueda en el corpus de la lengua de llegada. Para garantizar la obtención de equivalentes interlingüísticos, se deberá verificar que los elementos contextuales codifican las mismas relaciones semánticas tanto en la lengua origen como en la lengua meta. Por último, estas correspondencias interlingüísticas deberán validarse a través de un análisis de concordancias.

No obstante, este procedimiento queda lejos de ser la única técnica posible. Al contrario, se presenta como un ejemplo de la utilidad del contexto en la traducción de los TC. En concreto, esta metodología permite identificar equivalentes de los TC que no constituyen traducciones literales, pues estas derivan a menudo en errores de traducción. Evidentemente, los pasos propuestos también pueden complementarse con otras consultas, como la búsqueda de la traducción del núcleo del TC junto con los posibles modificadores, que igualmente puede ofrecer equivalentes del TC. En definitiva, estas sencillas consultas de corpus ofrecen un gran abanico de posibilidades para los traductores y terminólogos, de forma que cada usuario pueda idear nuevas consultas en función de sus necesidades.

Entre las líneas de investigación que se pueden derivar del presente estudio se encuentra el análisis de la utilidad de esta metodología en la traducción de otras unidades, como las colocaciones verbales, y la evaluación de su implementación en la docencia en Traducción.

Referencias

- Arce Romeral, L. y Seghiri Domínguez, M. (2018). Booth-friendly term extraction methodology based on parallel corpora for training medical interpreters. *Current Trends in Translation Teaching and Learning E*, 5, 1-46. http://www.cttl.org/uploads/5/2/4/3/5243866/cttl_e_2018_1.pdf
- Arntz, R. y Picht, H. (1995). *Introducción a la terminología*. Fundación Germán Sánchez Ruipérez.
- Arroyave Tobón, A. y Quiroz Herrera, G. (2012). Consideraciones didácticas para la enseñanza de sintagmas nominales con premodificación compleja del inglés al español. *Núcleo*, 24(29), 179-215. <http://ve.scielo.org/pdf/nu/v24n29/art08.pdf>
- Bowker, L. (2004). Corpus resources for translators: Academic luxury or professional necessity? *TradTerm*, 10, 213-247. <https://doi.org/10.11606/issn.2317-9511.tradterm.2004.47178>
- Bowker, L. (2015). Terminology and translation. En H. Kockaert y F. Steurs (Eds.), *Handbook of terminology* (pp. 304-323). John Benjamins. <https://doi.org/10.1075/hot.1.ter5>
- Cabezas-García, M. (2019). *Los compuestos nominales en terminología: formación, traducción y representación* [Tesis doctoral]. Universidad de Granada, España.
- Cabezas-García, M. y Faber, P. (2017). A semantic approach to the inclusion of complex nominals in english terminographic resources. En R. Mitkov (Ed.), *Computational and corpus-based phraseology* (pp. 145-159). Springer. https://doi.org/10.1007/978-3-319-69805-2_11
- Cabré, M. T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology*, 9(2), 163-200. <https://doi.org/10.1075/term.9.2.03cab>
- Daille, B., Dufour-Kowalski, S. y Morin, E. (2004). *French-English multi-word term alignment based on lexical context analysis* [Congress session]. Fourth International Conference on Language Resources and Evaluation (LREC 2004) (pp. 919-922). ELRA, Lisboa, Portugal. <http://www.lrec-conf.org/proceedings/lrec2004/pdf/563.pdf>
- Daille, B. y Morin, E. (2005). *French-English terminology extraction from comparable corpora* [Conference session]. International Conference on Natural Language Processing. ijcnlp 2005, Lecture Notes in Computer Science 3651. Springer. https://doi.org/10.1007/11562214_62
- Déjean, H. y Gaussier, É. (2002). Une nouvelle approche à l'extraction de lexiques bilingues à partir de corpus comparables. *Lexicometrica*, 1-22.
- Felber, H. (1984). *Terminology manual*. Infoterm.
- Fernández Silva, S. y Kerremans, K. (2011). Terminological variation in source texts and translations: A pilot study. *Meta*, 56(2), 318-335. <https://doi.org/10.7202/1006179ar>
- Gallego-Hernández, D. (2015). The use of corpora as translation resources: A study based on a survey of Spanish professional translators. *Perspectives*, 23(3), 375-391. <https://doi.org/10.1080/0907676X.2014.964269>
- Gobierno de Canadá (2021). *TERMIUM Plus*. <https://www.btb.termiumplus.gc.ca/>
- Harastani, R., Daille, B. y Morin, E. (2013). *Identification, alignment, et traductions des adjectifs relationnels en corpus comparables*. En *Vingtième conférence du Traitement Automatique du Langage Naturel 2013 (taln 2013)* (pp. 313-326). atala. <https://www.aclweb.org/anthology/F13-1023/>
- Hurtado Albir, A. (2011). *Traducción y traductología. Introducción a la traductología*. Cátedra.
- Interactive Terminology for Europe (IATE) (2021). European Union Terminology. <https://iate.europa.eu/>
- Kilgarriff, A., Rychlý, P., Smrz, P. y Tugwell, D. (2004). *The Sketch Engine* [Congress session]. 11th euralex International Congress (pp. 105-115). Euralex, Lorient, France. <https://euralex.org/publications/the-sketch-engine/>
- Maniez, F. (2008). Using the Web and computer corpora as language resources for the translation of complex noun phrases in medical research articles. *Panacea@*, 8(26), 162-167. https://www.tremedica.org/wp-content/uploads/n26_tribuna-Maniez.pdf

- Maniez, F. (2013). The translation into French of adjectives formed with a noun and a past participle in English-language medical articles. *Panacea*, 14(38), 240-247.
- Meyer, I. (2001). Extracting knowledge-rich contexts for terminography. A conceptual and methodological framework. En D. Bourigault, C. Jacquemin y M. C. L'Homme (Eds.), *Recent advances in computational terminology*. (pp. 279-302). John Benjamins. <https://doi.org/10.1075/nlp.2.15mey>
- Montero Martínez, S. y García de Quesada, M. (2003). Terminological analysis for translation. *Perspectives*, 11(4), 293-314. <https://doi.org/10.1080/0907676X.2003.9961482>
- Nord, C. (1997). *Translating as a purposeful activity: Functionalist approaches explained*. St. Jerome Publishing.
- Pérez Carrasco, M. y Seghiri Domínguez, M. (2021). 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) 5, 123-148. <https://revistas.uncu.edu.ar/ojs3/index.php/analeslinguistica/article/view/4592>
- Pimentel, J. (2013). Methodological bases for as-signing equivalents: A contribution. *Terminology*, 19(2), 237-257. <https://doi.org/10.1075/term.19.2.04pim>
- Reiss, K. y Vermeer, H. J. (1984). *Grundlegung einer allgemeinen Translationstheorie*. Niemeyer.
- Salager-Meyer, F. (1985). Syntax and semantics of compound nominal phrases in medical English literature: A Comparative Study with Spanish. *English for Specific Purposes*, 95, 6-12.
- Sanz Vicente, L. (2011). *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español* [Tesis doctoral]. Universidad de Salamanca. <https://infoling.org/repository/PhDdiss-Infoling-54-11-2012.pdf>
- Shahzad, I. Ohtake, K., Masuyama, S. y Yamamoto, K. (2000). Identifying translations of compound nouns using non-aligned corpora [Congress sessions]. Workshop MAL'99 (pp. 108-113). <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.3533&rep=rep1&type=pdf>
- Sketch Engine (s. f.). [Página web]. <https://www.sketchengine.eu/>
- Tercedor Sánchez, M., López Rodríguez, C. I. y Alarcón Navío, E. (2013). Identifying features of translation through multiword lexical units. *Belgian Journal of Linguistics*, 27(1), 87-109. <https://doi.org/10.1075/bjl.27.05ter>
- Wüster, E. (1979). *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Springer in Komm.

Cómo citar este artículo: Cabezas García, M. (2021). Metodología para la traducción de términos compuestos mediante corpus. *Mutatis Mutandis. Revista Latinoamericana de Traducción*, 14(2), 451-468. <https://doi.org/10.17533/udea.mut.v14n2a08>