

Fertilidad léxica en la formación de términos compuestos

MELANIA CABEZAS-GARCÍA

Profesora Ayudante Doctora
Universidad de Granada
Facultad de Traducción e Interpretación
c/Buenucesos, 11
18002 Granada
melaniacabezas@ugr.es

FERTILIDAD LÉXICA EN LA FORMACIÓN DE TÉRMINOS COMPUESTOS

RESUMEN: Los términos compuestos son un tipo de unidad léxica especialmente habitual en el discurso especializado. En ellos se condensa el conocimiento científico-técnico, por lo que su correcto tratamiento es fundamental para la transmisión de la información. Sin embargo, su análisis presenta complicaciones. En este artículo se utilizan técnicas de corpus para investigar la formación conceptual y pragmática de términos compuestos en inglés pertenecientes al ámbito de la energía eólica. En primer lugar, se desambiguó la estructura de los términos compuestos, a cuyos componentes se les asignaron posteriormente categorías semánticas. Tras analizar la relación interna en los términos compuestos, estos se anotaron con roles semánticos. Dicho análisis reveló el microcontexto de cada término, que explica la fertilidad en la formación de los términos compuestos y facilita distintas tareas cognitivas y discursivas, como la traducción.

PALABRAS CLAVE: término compuesto; traducción; combinación conceptual; terminología; corpus.

SUMARIO: 1. Introducción. 2. Formación de términos compuestos. 3. Materiales y métodos. 4. Microcontextos. 5. Conclusiones.

LEXICAL FERTILITY IN THE FORMATION OF MULTIWORD TERMS

ABSTRACT: Multiword terms are a type of lexical unit that is particularly common in specialized discourse. They condense scientific and technical knowledge, that is why its correct treatment is essential for information dissemination. However, their analysis is not an easy task. This article presents a corpus-based study, which examined the conceptual and pragmatic formation of English multiword terms belonging to the field of wind power. Firstly, the analysis involved the structural disambiguation of multiword terms. Then, the constituents of multiword terms were assigned semantic categories and their internal relation was identified. Finally, multiword terms were labelled with semantic roles. This analysis revealed the microcontext of each term, which explains fertility in the formation of multiword terms and facilitates different cognitive and discursive tasks, such as translation.

KEYWORDS: multiword term; translation; conceptual combination; terminology; corpus.

SUMMARY: 1. Introduction. 2. Multiword term formation. 3. Materials and methods. 4. Microcontexts. 5. Conclusions.

FERTILITÉ LEXICALE DANS LA FORMATION DES TERMES COMPLEXES

RÉSUMÉ: Les termes complexes sont un type d'unité lexicale particulièrement fréquente dans le discours spécialisé. Ils condensent les connaissances scientifiques et techniques, leur traitement correct étant donc essentiel pour la transmission de l'information. Cependant, leur analyse est rarement facile. Dans cet article, des techniques de corpus sont utilisées pour étudier la formation conceptuelle et pragmatique de termes complexes en anglais appartenant au domaine de l'énergie éolienne. L'analyse a d'abord consisté à désambigüer la structure des termes complexes, dont les for-mants ont ensuite été annotés avec des catégories sémantiques. Après avoir analysé la relation interne dans les termes complexes, ils ont été annotés avec des rôles sémantiques. Cette analyse a ré-vélé le microcontexte de chaque terme, qui explique la formation fertile des termes complexes et facilite plusieurs tâches cognitives et discursives, dont la traduction.

MOTS-CLÉS: terme complexe ; traduction ; combinaison conceptuelle ; terminologie ; corpus.

SOMMAIRE: 1. Introduction. 2. Formation des termes complexes. 3. Matériel et méthodes. 4. Microcontextes. 5. Conclusions.

Fecha de Recepción: 01/12/2020

Fecha de Revisión: 27/07/2021

Fecha de Aceptación: 09/08/2021

Fecha de Publicación: 01/12/2021

DOI: <http://dx.doi.org/10.25267/Pragmalinguistica.2021.i29.04>

Fertilidad léxica en la formación de términos compuestos

MELANIA CABEZAS-GARCÍA

1. INTRODUCCIÓN

Los términos compuestos (TC, p. ej. *fixed-pitch wind turbine* o *aerogenerador de paso fijo*) son muy habituales en el discurso especializado, debido a sus múltiples posibilidades de combinación conceptual, lo que explica también su frecuente uso para denominar nuevos conceptos (M. T. Díaz Hormigo y E. Vega Moreno 2018). En los TC se materializan los sistemas conceptuales especializados, de modo que contribuyen a la transmisión del conocimiento experto. Sin embargo, su análisis engloba aspectos problemáticos como su identificación o desambiguación estructural, su análisis semántico o la especificación de la relación semántica entre los formantes. Estos factores repercuten en diferentes tareas cognitivas y discursivas, como la comprensión y producción de textos, o su traducción. A pesar del interés que han recibido cuestiones como las relaciones semánticas de los TC o su desambiguación estructural, existen aspectos que aún requieren de una respuesta satisfactoria. Es el caso, por ejemplo, de su formación en inglés y en otras lenguas, cuyo conocimiento facilita parte de los problemas que plantean los TC.

En este artículo se analiza el desarrollo conceptual y pragmático de los TC, que ofrece múltiples beneficios de cara a distintas tareas cognitivas, como la interpretación o la traducción de estos términos. Para ello, compilamos un corpus de textos especializados en inglés sobre energía eólica, formado por 3 025 237 palabras. La investigación se centró, por tanto, en los TC del inglés, que constituye la *lingua franca* en la comunicación, si bien se emplearon ejemplos en español para ilustrar los resultados relativos a la traducción.

Para extraer los TC se utilizaron expresiones regulares en el corpus. A continuación, el análisis semántico de los TC conllevó los siguientes pasos: (i) desambiguación estructural; (ii) asignación de categorías semánticas; (iii) descodificación de la relación semántica interna mediante paráfrasis; y (iv) etiquetado con roles semánticos. Gracias a este análisis pragmático y semántico, se descubrió el *microcontexto* de los TC, que da cuenta de su formación y productividad, y facilita la comprensión y traducción de estos términos.

2. FORMACIÓN DE TÉRMINOS COMPUESTOS

Los compuestos se caracterizan por la unión de dos o más lexemas que constituyen una unidad de significado (P. Downing 1977; J. Levi 1978). Estos lexemas se pueden unir y formar una palabra, dando lugar a compuestos gráficos, como *sacacorchos*, o bien mantenerse separados, siendo entonces compuestos sintagmáticos, como *pez espada*. Aunque muchos autores han considerado compuestos únicamente a aquellos que presentan unión ortográfica (G. Corpas Pastor 1996; P. Štekauer *et al.* 2012), otros muchos también aplican esta

denominación a los que cuentan con un espacio entre los lexemas. En este trabajo consideramos compuestos tanto a aquellos que presentan unión ortográfica como a los que no, si bien nos interesamos por estos últimos. En concreto, nos centramos en los términos compuestos (TC), que son combinaciones del lenguaje especializado cuyo núcleo está modificado por uno o varios complementos. Por ejemplo, en *floating offshore wind turbine* (aerogenerador flotante), el núcleo *turbine* está modificado por *floating*, *offshore* y *wind*. Los TC son muy frecuentes en inglés y en español (P. Nakov 2013; J. Fernández Domínguez 2019), y son especialmente habituales en el discurso especializado (K. Kageura 2002; P. Nakov 2013), gracias a sus diversas posibilidades de combinación conceptual. Así, constituyen una herramienta primordial para la transmisión del conocimiento especializado.

Los TC se han estudiado desde la Lingüística General (P. Downing 1977; J. Levi 1978), la Terminología (L. Bowker 1998; U. Heid 2001; K. Kageura 2002), la Lexicografía (I. Meyer y K. Mackintosh 1996; S. Montero Martínez 2008), la Lingüística Computacional (M. Lauer 1995; B. Rosario *et al.* 2002) o la Psicolingüística (G. L. Murphy 1988; C. L. Gagné 2000; P. Maguire *et al.* 2010), entre otros campos. Identificarlos correctamente en el texto a veces resulta complicado, pues pueden estar formados por varios términos o incluso palabras de la lengua general. Las dependencias estructurales de los TC de tres o más formantes también plantean problemas. Además, el análisis semántico de los TC resulta complejo, ya que se omite la relación interna entre los formantes. Su traducción tampoco es tarea sencilla, debido a los diferentes patrones de formación de términos y a su descripción poco sistemática en los recursos terminológicos. Sin embargo, el estudio de la formación de los TC puede facilitar muchas de estas tareas.

Para comprender la generación y conceptualización de los TC, es necesario abordar primero su formación estructural. Así, los TC pueden ser endocéntricos o exocéntricos (J. Levi 1978; P. Nakov 2013). En los TC endocéntricos, uno de los componentes es el núcleo y el otro es el modificador, que añade características a dicho núcleo. Por ejemplo, en *offshore wind farm* (parque eólico marino), *offshore* modifica al núcleo, *wind farm*. La estructura binaria de los TC endocéntricos destaca su similitud con otro concepto, pues el núcleo indica la categoría a la que pertenece, mientras que los modificadores especifican una o varias características distintivas (L. Bowker 1998; C. L. Gagné 2000; P. Maguire *et al.* 2010). Así, el nuevo concepto resultante de la combinación suele ser un hipónimo del núcleo (P. Downing 1977; M. Cabezas-García y P. Faber 2017). Por ejemplo, *offshore wind farm* es un tipo de *wind farm*.

Los TC exocéntricos son menos frecuentes en el lenguaje general y aún menos en el discurso especializado. Carecen de núcleo (J. Levi 1978; P. Nakov 2013), como se observa en *saber tooth* (dientes de sable), en el que ningún componente designa la categoría a la que pertenece el TC (un felino). Esto explica su naturaleza idiomática y poco transparente. Existe también un tercer grupo de TC coordinados, en los que ambos constituyentes están al mismo nivel (p. ej. *secretary-treasurer* [secretario-tesorero]) (J. Fernández Domínguez 2019). En este artículo

nos centramos en los TC endocéntricos, que son, con diferencia, los más comunes en diferentes lenguas y en el discurso especializado.

Las categorías gramaticales que intervienen en los TC, así como la posición de los formantes, varían según los patrones de formación de términos de cada lengua (J. Levi 1978; J. Fernández Domínguez 2019). En inglés, los TC suelen formarse mediante la premodificación (P. Nakov 2013). Es decir, el núcleo se sitúa a la derecha, precedido por los modificadores, que pueden ser sustantivos (*speed ratio* [*relación de velocidad*]), adjetivos (*reactive power* [*potencia reactiva*]) o participios que actúan como adjetivos (*rated power* [*potencia nominal*]). A diferencia de las lenguas germánicas, como el inglés, el encadenamiento de sustantivos no es posible en las lenguas romances. Así pues, los TC en estas lenguas, como el español, suelen formarse por postmodificación mediante adjetivos (*potencia reactiva*) o sintagmas preposicionales (*hueco de tensión*). Como es de esperar, estas diferencias estructurales suelen causar problemas de comprensión y traducción.

En cuanto a la conceptualización subyacente en los TC, que influye en su formación, varios estudios señalan la existencia de proposiciones subyacentes, que cuentan con un predicado. J. Levi (1978) afirma que los TC surgen a partir de la supresión del predicado (p. ej. *virus causes disease* > *viral disease* [*enfermedad vírica*]) o de su nominalización (p. ej. *person that analyzes finance* > *financial analyst* [*analista financiero*]). Dado que las proposiciones subyacentes son conceptuales, es posible asignar roles semánticos a los argumentos del predicado (p. ej. AGENT, PATIENT) y especificar las relaciones semánticas (p. ej. *affects*, *causes*), que transmiten con mayor precisión la formación conceptual de los TC.

Varias teorías psicolingüísticas han tratado de explicar la combinación conceptual presente en estos términos. G. L. Murphy (1988) señala que el modificador rellena uno de los *slots* o huecos abiertos por el núcleo. Por ejemplo, en *apartment dog* (*perro del piso*), *apartment* rellena el hueco de HABITAT de *dog*. Por tanto, el significado del núcleo determina los huecos que pueden abrirse, así como los modificadores que pueden ocuparlos. Esta noción de apertura de huecos es similar a la de W. Martin (1992) y B. Rosario *et al.* (2002), y constituye la base de los microcontextos que presentamos en este estudio (Sección 4). Por el contrario, C. L. Gagné (2000) propone que los TC se forman mediante el establecimiento de una relación semántica entre ambos conceptos (p. ej. *snowman* [*muñeco de nieve*] > *man made of snow*). Sin embargo, como comentaremos a continuación, la apertura de huecos y las relaciones semánticas no son incompatibles, sino más bien complementarias.

En su estudio de los patrones semánticos de los TC, P. Maguire *et al.* (2010) sostienen que los núcleos que son semánticamente similares abren huecos que se rellenan con el mismo tipo de modificadores. Es decir, defienden que los miembros de una categoría semántica suelen combinarse del mismo modo (p. ej. AREA+ANIMAL > ANIMAL *located_at* AREA > *water mammal* [*mamífero acuático*], *freshwater fish* [*pez de agua dulce*]). El estudio de P. Maguire *et al.* (2010) resulta fundamental para nuestro enfoque de la formación de TC, pues las categorías semánticas pueden revelar patrones regula-

res en estos términos, facilitando así su comprensión y traducción.

No obstante, estas teorías de combinación conceptual se han aplicado sobre todo a las palabras de la lengua general. En la Terminología y el lenguaje especializado, la combinación conceptual se ha estudiado a modo de *colocación conceptual* en W. Martin (1992) y se ha aplicado posteriormente en los trabajos de U. Heid (2001), M. C. L'Homme (2000) y en los marcos conceptuales de la Terminología Basada en Marcos (Sección 3) (S. Montero Martínez 2008; M. Cabezas-García y P. Faber 2018; P. Faber y M. Cabezas-García 2019). W. Martin (1992) defiende que los conceptos especializados presentan una estructura argumental que resulta visible en su definición. Esta estructura argumental está formada por huecos, que se rellenan con otros conceptos de características semánticas específicas. Por tanto, se aplica la idea de la apertura de huecos a las combinaciones especializadas, si bien estas no constituyen siempre TC.

Se observa, así, que la combinación de conceptos que da lugar a los TC no se da de forma aleatoria, sino que existen restricciones. Las entidades suelen combinarse con atributos (material, color, ubicación, etc.), pero no con causas o resultados, que son típicos de los procesos. Por ejemplo, una inundación produce un resultado, pero una mesa, no. Estas preferencias conceptuales se reflejan a menudo en forma de patrones recurrentes, que empleamos para estructurar el mundo (J. Levi 1978; K. Kageura 2002; B. Rosario *et al.* 2002; P. Maguire *et al.* 2010). Dichos patrones pueden definirse mediante categorías y roles semánticos.

Por ejemplo, en lugar de afirmar que *cake* (*pastel*) suele combinarse con *carrot* (*zanahoria*) (*carrot cake*), resulta más adecuado señalar que la categoría conceptual de FOOD (COMIDA) suele coocurrir con la categoría de INGREDIENT (INGREDIENTE). Dado que tales categorías no son específicas de un idioma, este patrón se encuentra tanto en inglés (p. ej. *meat pie* [*pastel de carne*], *grilled chicken salad* [*ensalada de pollo a la parrilla*], *mascarpone tart* [*tarta de mascarpone*]) como en español (p. ej. *empanada de atún*, *flan de huevo*, *pastel de berenjenas*). J. Levi (1978), B. Rosario *et al.* (2002) y K. Kageura (2002), entre otros, han analizado la formación de TC basándose en combinaciones recurrentes de categorías. Por ejemplo, patrones como PATIENT/VIRUS DISEASE (PACIENTE/ENFERMEDAD VÍRICA) (*influenza patient* [*paciente de gripe*], *aids survivor* [*superviviente del sida*]) son comunes en el dominio de la biomedicina (B. Rosario *et al.* 2002).

Los roles semánticos también pueden emplearse para captar regularidades en los TC. Este enfoque es menos transparente que el anterior debido a la frecuente omisión del predicado, ya que de él dependen los roles semánticos de los argumentos. Sin embargo, estos roles pueden revelar regularidades, como en *wind energy system* (*sistema de energía eólica*) o *wind power plant* (*central eólica*), ambos formados por un INSTRUMENTO (*system, plant*) y el PRODUCTO de su acción (*wind energy, wind power*).

Los patrones recurrentes de categorías y roles semánticos facilitan la inferencia de la relación semántica interna en los TC (B. Rosario *et al.* 2002; P. Maguire *et al.* 2010). Por ejemplo, B. Rosario *et al.* (2002: 3) defienden que las categorías BODY REGION (ZONA DEL CUERPO) y CARDIOVAS-

CULAR SYSTEM (SISTEMA CARDIOVASCULAR) coocurren en TC como *scalp artery* (arteria del cuero cabelludo), *heel capillary* (capilar del talón) y *limb vein* (vena de la extremidad), que codifican la relación *located_at*. Por tanto, se puede asumir que *pulmonary vein* (vena pulmonar) también codifica esta relación, dado que sus componentes son similares. Este enfoque facilita la comprensión y la traducción de TC desconocidos porque, como señala J. Levi (1978), estas combinaciones frecuentes no solo se producirán en inglés, sino probablemente también en otros idiomas.

La combinación de categorías semánticas y roles permite también explicar la formación de los TC. Uno de los pocos estudios que integra ambos enfoques es el de P. Bouillon *et al.* (2012), quienes analizan TC en francés e italiano, marcando sus componentes con categorías (ARTIFACT, TIME PERIOD, etc.) y roles (AGENT, INSTRUMENT, etc.). Además, especifican la existencia de un predicado interno. Por ejemplo, en *freight train* (tren de carga), el predicado subyacente es *transport*, de forma que se explicita la relación semántica.

No obstante, aunque las categorías y los roles semánticos son útiles para la caracterización de los TC, las relaciones semánticas también son necesarias. Estas relaciones suelen estar ocultas (p. ej. *offshore wind farm* [parque eólico marino] > *wind farm located offshore*). Para especificarlas se han empleado generalmente inventarios de relaciones (M. Casas Gómez 2020, p. ej. *cause*, *affect*, etc.), que pueden presentar una mayor o menor granularidad. Sin embargo, según O. Jespersen (1942) y P. Downing (1977), la asociación entre los constituyentes del TC no puede caracterizarse con un inventario cerrado de relaciones. Para ello, M. Lauer (1995) utiliza paráfrasis preposicionales (*olive oil* [aceite de oliva] > *oil from olives*) y P. Nakov y M. Hearst (2006) emplean paráfrasis verbales (*malaria mosquito* [mosquito de la malaria] > *mosquito that carries/spreads/transmits malaria*). Estas son actualmente muy populares entre los lingüistas computacionales.

Más que una alternativa a las taxonomías tradicionales, las paráfrasis pueden entenderse como un método para acceder a estas relaciones. De este modo, las relaciones semánticas y las paráfrasis son enfoques complementarios: las paráfrasis ayudan a representar la combinación de conceptos de un TC, que puede etiquetarse con una relación semántica.

3. MATERIALES Y MÉTODOS

Para la extracción y el análisis de los términos, compilamos un corpus de textos especializados en inglés sobre energía eólica, un área de indudable actualidad debido a la creciente preocupación por el cuidado del medio ambiente. Dicho corpus contenía 3 025 237 palabras y estaba formado por textos especializados (informes, libros y artículos de investigación) 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 inglés, pues este estudio no persigue criterios diacrónicos ni diatópicos. Los textos se seleccionaron en función de su autoridad (reputación y experiencia del autor), contenido (cober-

tura temática, objetividad, actualidad y destinatarios) y el diseño de las fuentes (accesibilidad y facilidad de tratamiento). La herramienta empleada para la compilación y el análisis del corpus fue Sketch Engine (<https://www.sketchengine.eu/>, A. Kilgarriff *et al.* 2014).

Para la extracción de TC, partimos de 5 de los conceptos más relevantes del dominio de la energía eólica (*wind* [viento], *turbine* [turbina], *power* [potencia/energía], *generator* [generador] y *voltage* [tensión]), que seleccionamos debido a su elevada frecuencia en el corpus, su amplia capacidad de generación de TC y el consiguiente interés en el estudio de la fertilidad léxica. Estos conceptos se emplearon en expresiones regulares mediante Corpus Query Language (CQL) para extraer TC que incluyesen dichos conceptos de partida. En la Tabla 1 se muestra un ejemplo de una consulta CQL para la extracción de TC:

```
[tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,}[lemma="generator"][[tag!="N.*|JJ.*"]
```

Tabla 1: Ejemplo de consulta CQL para extraer TC que incluyan un término concreto (*generator*). Fuente: Autora

Mediante la consulta de la Tabla 1 buscamos *generator* ([lemma="generator"]), que podía ir precedido de sustantivos, adjetivos, adverbios y participios pasados o presentes, que apareciesen una o más veces ([tag="N.*|JJ.*|RB.*|VVN.*|VVG.*"]{1,}). *Generator* no podía ir seguido por sustantivos o adjetivos ([tag!="N.*|JJ.*"]), para evitar que los TC extraídos formasen parte de una secuencia mayor. Así, obtuvimos TC como *synchronous generator* (*generador síncrono*) o *doubly fed generator* (*generador doblemente alimentado*).

Nos centramos en los 40 TC más frecuentes generados por cada uno de los 5 términos de partida, de modo que obtuvimos una muestra de 200 TC. Tras revisar sus concordancias y descartar los términos erróneos o duplicados, la lista se redujo a 182 TC. Su análisis implicó los siguientes pasos: (i) desambiguación estructural, (ii) asignación de categorías semánticas, (iii) descodificación de la relación semántica interna y (iv) asignación de roles semánticos. Se trata de un análisis semántico en contexto, que emplea tanto el cotexto inmediato como el contexto oracional y comunicativo, como se observará a continuación.

En primer lugar, fue necesario determinar la estructura de aquellos términos formados por más de dos elementos, ya que este paso es necesario antes de realizar las siguientes tareas. Por tanto, debe desambiguarse la estructura de los TC, con vistas a reducir aquellos que cuentan con más de dos formantes (61 de los 182) a su estructura de núcleo y modificador. Para ello, utilizamos el protocolo de M. Cabezas-García y P. León-Araúz (2019), que incluye diferentes consultas de corpus basadas en las propiedades lingüísticas de los TC y en los recientes avances en el Procesamiento del Lenguaje Natural. Por ejemplo, *wind turbine rotor* (*rotor del aerogenerador*) se señaló como [*wind turbine*] rotor porque *wind turbine* era más frecuente en el corpus que la otra combinación posible, *turbine rotor*. Además de este tipo de búsqueda, el protocolo contempla otras consultas para resolver la estructura de los TC.

Seguidamente, asignamos categorías semánticas a los TC, para lo que empleamos un inventario de categorías elaborado a partir de los conceptos incluidos en EcoLexicon (<https://ecolexicon.ugr.es/>), una base de conocimiento terminológica sobre el medio ambiente (M. Cabezas-García 2020). Las categorías de este inventario se basan en las definiciones de los conceptos y en la información contextual extraída del corpus de EcoLexicon. Tras determinar las características compartidas por los conceptos, el inventario de categorías semánticas se organizó jerárquicamente en cinco niveles. El nivel más general incluye las categorías ontológicas iniciales de ENTITY (objetos mentales y físicos), PROCESS (eventos con duración) y ATTRIBUTE (características de las entidades o procesos). El resto de niveles son más específicos. Estas categorías se asignaron a los grupos resultantes de la desambiguación estructural y se determinaron consultando las definiciones y la información contextual en el corpus de energía eólica. En la Tabla 2 se presenta la anotación de *generator torque control* (*control del par del generador*). Dado que la estructura interna de este término era [*generator torque*] *control*, al modificador, *generator torque*, se le asignó la categoría ENTITY>FORCE>STRESS, y al núcleo, *control*, se le asignó la categoría PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY.

<i>generator torque control</i>	
[generator torque]	ENTITY>FORCE>STRESS
control	PROCESS>CHANGE>CHANGE IN SIZE/INTENSITY

Tabla 2: Asignación de categorías semánticas a *generator torque control*. Fuente: Autora

A continuación, analizamos la relación semántica interna en los TC. Para ello, buscamos en primer lugar paráfrasis verbales y, cuando no se obtuvieron resultados, extrajimos paráfrasis libres. Al mostrar la proposición subyacente en contexto, ambas facilitan la identificación de la relación interna en los TC. La siguiente expresión CQL, que busca los dos elementos o grupos principales del TC unidos por un verbo, se utilizó para buscar paráfrasis verbales (Tabla 3):

<pre>([lemma="power"]{0,10}{tag="V.*"}{0,10}[lemma="curve"] within <s/> ([lemma="curve"]{0,10}{tag="V.*"}{0,10}[lemma="power"] within <s/>)</pre>

Tabla 3: Ejemplo de consulta CQL para extraer paráfrasis verbales. Fuente: Autora

Con esta expresión, que ilustramos con *power curve* (*curva de potencia*), se busca uno de los elementos del TC ([lemma="power"]), seguido de un verbo ([tag="V.*"]), que representaría la relación entre los componentes del TC. Este verbo podría ir precedido o seguido de 0 a 10 elementos ({0,10}). También se busca el otro constituyente del TC ([lemma="curve"]). Además, las paráfrasis debían aparecer dentro de la misma oración (within <s/>), ya que no nos interesaba obtener oraciones diferentes. También incluimos el símbolo | para indicar la disyunción e indicamos la misma expresión en orden inverso, con vistas a encontrar los constituyentes a ambos lados del verbo. Con esta

expresión se obtuvieron líneas de concordancia como las de la Figura 1, en las que *power* y *curve* aparecen unidos por verbos como *plot* (*trazar*) y *represent* (*representar*), que señalan la proposición conceptual *curve represents power*.

characteristics, including power curve (the **curve which plots the power** output of a turbine as a function of the wind
respectively. In Figures 3 and 4, the green **curve represents practical output power**; the blue curve is for the prediction result of
Wind turbine manufacturers provide power **curves representing turbine power** output as a function of wind speed (see Chapter I

Figura 1: Muestra de las paráfrasis verbales de *power curve*. Fuente: Sketch Engine

Sin embargo, no siempre se encontraron verbos que ilustraran claramente la relación interna de los TC. En esos casos, se utilizaron paráfrasis libres (es decir, ocurrencias de los constituyentes del TC en el corpus). Estas pueden emplearse para analizar el contexto en el que los elementos del TC coocurren y, por tanto, para extraer características semánticas y pragmáticas. Por ejemplo, la siguiente consulta se utilizó para extraer paráfrasis libres de *nominal voltage* (*tensión nominal*) (Tabla 4):

```
([lemma="nominal"][lemma!="voltage"]][0,10][lemma!="nominal"][lemma="voltage"]
within <s/>|
([lemma!="nominal"][lemma="voltage"])[0,10][lemma="nominal"][lemma!="voltage"]
within <s/>
```

Tabla 4: Ejemplo de consulta CQL para extraer paráfrasis libres. Fuente: Autora

Con la expresión regular de la Tabla 4 se busca uno de los constituyentes del TC ([lemma="nominal"]), al que no podía seguir inmediatamente el otro constituyente ([lemma!="voltage"]) para evitar extraer el propio TC. A continuación se introdujo un margen de 0 a 10 palabras ([0,10]), seguido del otro elemento del TC ([lemma="voltage"]), al que tampoco podía preceder el constituyente previamente incluido ([lemma!="nominal"]). La consulta debía aparecer en la misma oración (within <s/>). Finalmente, introdujimos el símbolo de disyunción |, seguido de la misma secuencia en orden inverso. Se obtuvieron paráfrasis libres que aludían a un elemento que no estaba presente en el TC (*value* [*valor*]), lo que probablemente complicó la extracción de paráfrasis verbales:

- (1) It is essential that the *voltage* is kept close to the *nominal value*, in the entire power system.
- (2) The large-scale centralised power plants *keep the node voltages within the allowed deviation from their nominal value* and the number of dedicated voltage control devices is limited.
- (3) The voltage balance control scheme is reactivated and input capacitor *voltages quickly converge to nominal values*.

Cuando no se pudo acceder a la semántica del TC por ninguno de estos medios, se combinó la búsqueda del propio TC en el corpus con el uso de recursos terminológicos externos. Las relaciones se asignaron desde el núcleo hacia el modificador, ya que su direccionalidad normalmente

puede alterarse (*has_function* vs. *function_of*). Empleamos el inventario de relaciones de EcoLexicon (*type_of*, *part_of*, *made_of*, *delimited_by*, *located_at*, *takes_place_in*, *phase_of*, *affects*, *causes*, *attribute_of*, *opposite_of*, *studies*, *measures*, *represents*, *result_of*, *effected_by* y *has_function*), además de la relación específica del dominio *uses_resource*.

Tras identificar las proposiciones conceptuales, se asignaron roles semánticos a los grupos resultantes de la desambiguación estructural. Aunque estos roles se han atribuido tradicionalmente a los argumentos de un predicado, nos basamos en las proposiciones conceptuales subyacentes en los TC. Estas proposiciones incluían varios conceptos, que estaban vinculados por una relación semántica. La asignación de roles se basó, por tanto, en la relación que vinculaba los conceptos y no necesariamente en la presencia explícita de un predicado (p. ej. en *power station* [*central eléctrica*], *station* es el AGENTE que *causa* un PACIENTE, *power*). Se utilizó el inventario de roles semánticos de EcoLexicon, que se inspira en los roles propuestos en FrameNet (C. J. Fillmore 1982) y VerbNet (K.ipper *et al.* 2004). Los roles incluidos fueron AGENT, THEME, POSSESSOR, PATIENT, INSTRUMENT, LOCATION y DESCRIPTIVE. Por ejemplo, *wind park* (*parque eólico*) se anotó como *wind*[INSTRUMENT] *park*[AGENT] porque la proposición conceptual era *park uses_resource wind*. En conclusión, este análisis semántico y pragmático arrojó luz sobre la formación de los TC.

4. MICROCONTEXTOS

El análisis de corpus permitió investigar la formación de los TC desde una perspectiva conceptual y pragmática, y proporcionó información útil para diferentes tareas cognitivas y discursivas. En primer lugar, al explorar cómo se combinaban las categorías semánticas dentro de los TC, se identificaron patrones recurrentes, que están directamente relacionados con los microcontextos que se presentarán a continuación. Por ejemplo, la combinación de las categorías ATTRIBUTE y ENTITY>CREATION produjo 32 TC, como *asynchronous generator* (*generador asíncrono*) y *commercial wind turbine* (*aerogenerador comercial*). También se descubrieron variantes terminológicas (p. ej. sinónimos de *wind farm* [*parque eólico*] como *wind plant* o *wind power station*) y variantes conceptuales (p. ej. *wind power* como atributo del viento [*potencia del viento*] o la electricidad resultante de su uso [*energía eólica*]).

Para comprender de manera eficiente un texto, resulta fundamental identificar las variantes terminológicas con las que puede nombrarse un concepto. En cuanto a la traducción, la variación terminológica debe reconocerse en el idioma de origen, pues ello facilita la identificación de las variantes en el idioma de destino. Además, también debería considerarse la preferencia por una u otra variante, así como sus posibles consecuencias cognitivas o estilísticas. En este sentido, al traducir las variantes terminológicas, que suelen ser TC (P. León-Araúz y M. Cabezas-García 2020), se debe considerar si respetar esta variación (porque cumple funciones cognitivas o comunicativas [S. Fernández Silva y K. Kerremans 2011]) o evitarla en aras de la coherencia.

La variación conceptual puede dificultar la comprensión de un TC y, por tanto, de su potencial combinatorio. Por ejemplo, cuando *power* se entiende como un atributo (*potencia*), se combina con verbos como *extract* (*extraer*) o *absorb* (*absorber*). Por el contrario, cuando *power* se entiende como electricidad, tiende a coocurrir con predicados como *store* (*almacenar*) o *consume* (*consumir*). Sin embargo, la coincidencia de significados en un mismo término no siempre se reproduce en otras lenguas. En español, *power* se suele traducir como *energía* o *electricidad* cuando significa energía eléctrica. Por el contrario, su equivalente habitual cuando se refiere a un atributo es *potencia*. Los TC derivados también siguen este patrón. Por tanto, calcar automáticamente el término en el idioma de destino puede ser peligroso, ya que puede llevar al uso de equivalentes que no transmitan el mismo significado.

Otro tipo de variación conceptual está relacionado con la multidimensionalidad, es decir, la descripción del mismo concepto desde diferentes perspectivas. Ello suele dar lugar a variantes terminológicas resultantes de las diferentes conceptualizaciones. La multidimensionalidad puede afectar a la forma del TC, ya que la dimensión que más se destaca en el contexto inmediato es la que aparece en primera posición en el TC (L. Bowker 1998: 490). Por ejemplo, en un contexto relacionado con el color, se preferirá *colour flatbed scanner* (*escáner de cama plana a color*) a *flatbed colour scanner* porque *colour* se nombra en primer lugar. Este también es el caso en las variantes españolas *densidad espectral de potencia* y *densidad de potencia espectral*, en las que el término más cercano al núcleo destaca cognitivamente. Sin embargo, sus correspondencias en inglés (*power spectrum* y *power spectral density*) no permiten esta variación formal, lo que subraya las diferencias entre las lenguas de origen y de destino.

Las relaciones semánticas también tienen un papel central en los microcontextos. El uso de paráfrasis para encontrar esta relación permitió extraer verbos que caracterizan la relación entre los constituyentes del TC, lo que aclaró su significado (p. ej. *control* [*controlar*], *adjust* [*ajustar*], *fluctuate* [*fluctuar*] o *vary* [*variar*] especifican la relación *affects*). Además, las paráfrasis verbales revelaron información semántica y contextual adicional, lo que facilita en gran medida la comprensión. Ello puede observarse en la siguiente paráfrasis de *wind penetration* (*penetración eólica*), que desvela el lugar donde la energía del viento *penetra* o se integra: «Wind has already made solid steps forward, penetrating national transmission systems by as much as 10 per cent in several markets and as much as 21 per cent in Denmark».

Las paráfrasis también sacaron a la luz elementos que no figuraban de forma explícita en los TC. Por ejemplo, las paráfrasis de *floating turbine* (*turbina flotante*) mostraron que la turbina *is installed on a floating foundation/ is set on a floating column/ is installed on a floating platform*. Por tanto, la turbina flota porque está situada sobre elementos de soporte como bases, columnas o plataformas. Sin embargo, estos elementos no se mencionan explícitamente en el TC. Las paráfrasis libres también desvelaron argumentos adicionales, elementos ocultos del TC y variantes.

Una vez identificada la relación semántica interna, se analizaron los roles semánticos de los componentes del TC en la proposición conceptual subyacente, lo que permitió captar regularidades. Por ejemplo, el esquema de roles POSSESSOR-THEME generó 56 TC, como *line-to-neutral voltage (tensión de fase)* o *wind turbine blade (pala del aerogenerador)*. Estas restricciones son fundamentales en los microcontextos.

Para analizar la formación de TC en el dominio de la energía eólica, partimos de la premisa de que las unidades semánticamente similares abren huecos que se rellenan por otras unidades que también comparten características conceptuales, lo que da lugar a combinaciones especializadas. Estos huecos son una especie de estructura argumental y pueden percibirse en la definición del concepto. Así, agrupamos las unidades semánticamente similares (es decir, TC cuyos núcleos pertenecían a la misma categoría semántica). Las categorías semánticas de estos núcleos se consideraron *categorías nucleares*. Por ejemplo, *turbine (turbina)*, *converter (convertidor)*, *grid (red)*, *plant (central)* o *farm (parque)*, entre otros términos, se refieren a conceptos específicos de la categoría nuclear ENTITY>CREATION. Para explorar si estos diferentes términos presentaban las mismas pautas de combinación, comparamos las categorías semánticas y los roles con los que coocurrían, así como las relaciones que codificaban.

El análisis del potencial combinatorio de las categorías nucleares mostró que los conceptos o términos pertenecientes a una determinada categoría nuclear presentan patrones de combinación semejantes. Esta similitud en las combinaciones de los TC nos llevó a hablar de *microcontextos*, que explican la productividad en la formación de los TC. Los microcontextos son una especie de estructura argumental del núcleo del TC. En concreto, las categorías semánticas nucleares abren uno o varios huecos en función de su significado, que determinan los TC que pueden formarse para especificar esa categoría nuclear. El análisis de la formación de TC a través de sus microcontextos facilita distintas tareas cognitivas, como la comprensión y la traducción de los TC.

Ilustramos este proceso con la categoría de ENTITY>CREATION, que fue muy productiva en el estudio. En la Tabla 5 se muestran los huecos abiertos por esta categoría (columna de la izquierda), etiquetados con su rol en la proposición conceptual del TC. A continuación, se muestran las categorías semánticas que pueden rellenar estos huecos, así como la relación semántica codificada entre el núcleo y el modificador. Finalmente, la columna de la derecha presenta una muestra de los TC que exhiben ese patrón de combinación.

ENTITY>CREATION			
Hueco (rol)	Categoría	Relación	Ejemplos
PATIENT	PROCESS>MOVEMENT>ENERGY MOVEMENT	<i>causes</i>	AC generator
		<i>affects</i>	voltage source converter

	ENTITY>MATTER		steam generator
INSTRUMENT	ENTITY>FORCE	<i>uses_resource</i>	wind farm
	PROCESS>MOVEMENT		
	ATTRIBUTE	<i>has_attribute</i>	vertical axis wind turbine
DESCRIPTIVE	ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	<i>has_function</i>	large wind farm
		<i>has_type</i>	offshore wind turbine
LOCATION	ATTRIBUTE>LOCATION	<i>located_at</i>	onshore wind farm generator rotor
POSSESSOR	ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	<i>part_of</i>	wind turbine generator
	ENTITY>MATTER>SOLID MATTER>MATERIAL		permanent magnet generator
THEME (PART)	ENTITY>CREATION>ARTIFACT> INSTRUMENT>TRANSFORMING INSTRUMENT	<i>has_part</i>	wound rotor induction generator

Tabla 5: Huecos abiertos por la categoría ENTITY>CREATION. Fuente: Autora

La categoría ENTITY>CREATION abre los siguientes huecos: PATIENT, INSTRUMENT, DESCRIPTIVE, LOCATION, POSSESSOR y THEME (PART). El hueco PATIENT alude al concepto que sufre la acción del núcleo y siempre se rellenó con la categoría PROCESS>MOVEMENT>ENERGY MOVEMENT, como se muestra en *AC generator (alternador)* y *voltage source converter (convertidor en fuente de tensión)*. Sin embargo, PATIENT puede tener dos conceptualizaciones diferentes, según la relación que codifique con el núcleo.

Por un lado, PATIENT puede ser un concepto que se convierte o transporta cuando esta es la acción llevada a cabo por la categoría ENTITY>CREATION. Así, la relación interna es *affects* (converter *affects* voltage source). En nuestro estudio, este patrón se dio en los TC derivados de los núcleos *converter (convertidor)*, *invertir (inversor)* y *grid (red)*. Por otro lado, la categoría ENTITY>CREATION también puede enfatizar la fase de producción. En estos casos, se entiende por PATIENT el concepto que resulta de esta producción y la relación codificada es *causes*. Este patrón fue recurrente en los TC cuyo núcleo era *generator (generador)*, *plant o station (central)* (*generator causes* AC). Como se puede observar, la categoría de ENTITY>CREATION puede especificarse indicando el concepto que sufre la acción (PATIENT).

El hueco INSTRUMENT especifica el concepto empleado por ENTITY>CREATION para realizar la actividad. Son dos los tipos de conceptos que suelen rellenar este hueco: (i) conceptos referidos al recurso externo que se emplea (p. ej. *steam generator [generador de*

vapor], *wind farm* [*parque eólico*]), que pertenecen a las categorías de ENTITY>MATTER o PROCESS>MOVEMENT; (ii) conceptos referidos a la fuerza o los materiales empleados para generar movimiento (p. ej. *induction generator* [*generador de inducción*], *impulse turbine* [*turbina de acción*]), que pertenecen a las categorías de ENTITY>FORCE y ENTITY>MATTER. Los recursos utilizados son muy diversos: pueden ser fluidos, químicos, gases, etc. (ENTITY>MATTER), además de procesos como el viento o la inducción. Como es de esperar, las características de los conceptos pertenecientes a ENTITY>CREATION (*generator* [*generador*], *turbine* [*turbina*], *farm* [*parque*], etc.) varían en función del recurso utilizado. La relación codificada siempre es *uses_resource*. Los diferentes conceptos pertenecientes a ENTITY>CREATION abrieron este hueco, excepto *grid* (*red*), que se centra en el concepto transportado (*power grid* [*red eléctrica*]) más que en el instrumento empleado para el transporte.

El hueco DESCRIPTIVE permite especificar características de ENTITY>CREATION. Dado que estas características son atributos, la relación semántica interna es *has_attribute* (p. ej. en *vertical axis wind turbine* [*aerogenerador de eje vertical*], *wind turbine has_attribute* vertical axis). Sin embargo, en dos casos se codificó otra relación. Por un lado, en *commercial wind turbine* (*aerogenerador comercial*) se activa la relación *has_function*, ya que el concepto es una turbina eólica para uso comercial. No obstante, *commercial* sigue siendo un rasgo descriptivo del núcleo, por lo que desempeña el rol DESCRIPTIVE. Por otro lado, *wind turbine generator* (*generador del aerogenerador*) codifica la relación *has_type*, ya que un aerogenerador es un tipo de instrumento generador. En este caso, la categoría semántica que rellena este hueco no es ATTRIBUTE, sino ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT. Sin embargo, *wind turbine* mantiene su rol (DESCRIPTIVE). En definitiva, aunque estos dos conceptos no son atributos, aluden a funciones y subtipos, que también constituyen características descriptivas.

Otro de los huecos que se abrió fue el de LOCATION, que evoca el lugar donde se encuentra ENTITY>CREATION. Este hueco se relleno con la categoría ATTRIBUTE>LOCATION y la relación es *located_at* (p. ej. en *offshore wind turbine* [*aerogenerador marino*], *wind turbine located* offshore).

Además, el hueco de POSSESSOR indica la categoría más general a la que pertenece la categoría nuclear. En nuestra muestra, fueron los instrumentos de transformación (p. ej. *generator* [*generador*], *turbine* [*turbina*]) los que activaron este hueco, que se relleno con conceptos de la misma categoría, de modo que dos conceptos similares (instrumentos de transformación) se combinaron para formar un TC. La relación es *part_of* (p. ej. en *generator rotor* [*rotor del generador*], *rotor part_of* generator). Por tanto, los TC en los que se especifica este hueco suelen designar las partes de un instrumento transformador que se encargan de la transformación (p. ej. *rotor generator* [*rotor del generador*], *wind turbine generator* [*generador del aerogenerador*]).

Finalmente, la categoría nuclear de ENTITY>CREATION también abrió el hueco THEME (PART), que se activa cuando ENTITY>CREATION se entiende como POSSESSOR. Así pues, este hueco hace referencia a sus partes. Solo los instrumentos transformadores abrieron este hueco en nuestra

muestra, que se rellenó con la categoría semántica de ENTITY>MATTER>SOLID MATTER>MATERIAL cuando se trataba de los materiales de la entidad, como *permanent magnet* en *permanent magnet generator* (*generador de imanes permanentes*). También puede rellenarse con la categoría ENTITY>CREATION>ARTIFACT>INSTRUMENT>TRANSFORMING INSTRUMENT, como se refleja en los componentes de la entidad encargados de la transformación (p. ej. *wound rotor* en *wound rotor induction generator* [*generador de inducción de rotor bobinado*]). En todos los casos, la relación interna es *has_part*.

El análisis de los huecos abiertos por ENTITY>CREATION destacó la alta productividad de esta categoría en la formación de TC (67 TC siguieron estas pautas de combinación). Además, varios de estos huecos coincidieron en los diferentes conceptos pertenecientes a la categoría ENTITY>CREATION (p. ej. *generator* [*generador*], *grid* [*red*], *farm* [*parque*]), lo que recalca la semejanza en sus pautas de combinación. En concreto, estos huecos fueron los de PATIENT, INSTRUMENT, DESCRIPTIVE y LOCATION, que se activaron en prácticamente todos los conceptos. Así pues, estos huecos forman la *estructura argumental* o microcontexto de la categoría nuclear ENTITY>CREATION.

En definitiva, los microcontextos constituyen una nueva visión del procedimiento de *slot filling* (apertura y ocupación de huecos), que refleja el mecanismo conceptual por el que se generan los TC en la lengua especializada. De este modo, pueden entenderse como perfiles semánticos de la categoría en cuestión y resultan útiles a diferentes niveles.

Por ejemplo, pueden utilizarse para recrear el mapa conceptual del dominio especializado, que constituye el fondo contextual que estructura todo el conocimiento. Dado que el análisis de los microcontextos requiere la exploración las proposiciones conceptuales que subyacen a los TC, se proporciona una visión del sistema conceptual, que es independiente de la lengua. Por ejemplo, el análisis de los huecos abiertos por *turbine* (*turbina*) revela algunas de las relaciones que este concepto codifica con otros: (i) *uses_resource* (activada en TC como *gas turbine* [*turbina de gas*], *hydraulic turbine* [*turbina hidráulica*] o *wind turbine* [*aerogenerador*]); (ii) *has_attribute* (activada en *fixed-speed wind turbine* [*aerogenerador de velocidad fija*] y *downwind turbine* [*aerogenerador a sotavento*]); (iii) *located_at* (activada en *onshore turbine* [*turbina terrestre*] y *offshore turbine* [*turbina marina*]); y (iv) *has_part* (activada en *turbine rotor* [*rotor de la turbina*] y *wind turbine generator* [*generador del aerogenerador*]). En este sentido, conocer los conceptos relacionados y, con ellos, el dominio especializado, resulta vital cuando se traslada un texto a otra lengua.

Asimismo, los microcontextos y las definiciones están interrelacionados. Por una parte, los microcontextos pueden inferirse a partir de las definiciones de los conceptos. Por otra parte, el análisis de los TC generados a partir de un núcleo permite delimitar las características definitorias de este concepto nuclear. Por ejemplo, algunos de los TC formados a partir de *plant* (*central*) en el ámbito de la energía eléctrica son *power plant* (*central eléctrica*), *wind power plant* (*central eólica*), *wind plant* (*central eólica*), *hydropower plant* (*central hidroeléctrica*) y *thermal power plant* (*central térmica*). En estos TC, el núcleo (*plant*)

pertenece a la categoría de ENTITY>CREATION>STRUCTURE y abre los huecos de PATIENT (*power*) e INSTRUMENT (*wind, hydro-, thermal*). Así pues, gracias a esta información podemos definir *plant* del siguiente modo: *plant structure that converts some form of energy into electrical energy*.

La aclaración del concepto en forma de definición resulta también de gran utilidad para la traducción, ya que proporciona las características conceptuales a partir de las cuales surgirá el equivalente, además de poder utilizarse para verificar la correspondencia conceptual.

Además, los TC pueden surgir a partir de la ocupación de varios huecos. Esto ocurre a menudo en los TC con tres o más formantes. Por ejemplo, en *wound rotor induction generator* (*generador de inducción de rotor bobinado*), *generator*, que es el núcleo, activa primero el hueco de INSTRUMENT (*induction*) y luego el hueco de THEME (PART) (*wound rotor*). Dado que este tipo de TC es más extenso y cognitivamente más complejo, puede dificultar la comprensión y la traducción. Por ello, el análisis conceptual de los microcontextos permite aclarar este tipo de términos.

Igualmente, el análisis de los microcontextos y, por tanto, de los esquemas regulares de combinación de los TC facilita la inferencia de nuevos TC al compararlos con los patrones ya conocidos. Por ejemplo, si en *voltage source converter* (*convertidor en fuente de tensión*), el núcleo *converter* rellena el hueco de PATIENT con la categoría de PROCESS>MOVEMENT>ENERGY MOVEMENT (mediante el modificador *voltage source*) y codifica la relación *affects*, es de esperar que exista el mismo patrón en *voltage source inverter* (*inversor en fuente de tensión*). Esto también se aplica a las correspondencias interlingüísticas, como *convertidor en fuente de tensión* o *inversor en fuente de tensión*, que también se pueden inferir fácilmente comparándolas con los patrones conocidos. Este proceso reduce el tiempo dedicado a la comprensión del concepto, lo que a su vez acelera el proceso de traducción.

Por otro lado, al conocer los huecos que abre el núcleo de un TC en la lengua origen, resulta también más sencillo reconocer las combinaciones de este núcleo en el idioma de destino. Dado que los TC rara vez se traducen literalmente, deben identificarse los diferentes huecos, pues su materialización puede variar en cada lengua. Por ejemplo, en Termium Plus (<https://www.btb.termiumplus.gc.ca/>), *generating station* se define del siguiente modo: «unit that converts some form of energy into electrical energy». El término en inglés (*generating station*) alude a la generación de energía, mientras que su equivalente en español (*central eléctrica*) especifica la electricidad resultante. Como puede observarse, un hueco puede omitirse en una lengua y explicitarse en otra. Por ejemplo, *aerogenerador* evoca específicamente la actividad de generación, algo que no ocurre en su equivalente *wind turbine*. En consecuencia, un mismo hueco puede plasmarse en dos lenguas mediante categorías o roles semánticos distintos (*rated power, rated output* > *potencia nominal, potencia nominal de salida*). Por tanto, un buen dominio de los microcontextos permite acceder a las posibles combinaciones en las lenguas origen y meta, y agiliza el proceso de búsqueda.

En definitiva, los microcontextos permiten explicar la generación

conceptual de los TC. Así, profundizar en la formación de estos términos simplifica el proceso de comprensión, así como la identificación de equivalentes de traducción.

5. CONCLUSIONES

En este capítulo se ha profundizado en la formación conceptual y pragmática de los TC mediante un estudio basado en corpus. La noción de *microcontexto* parte de las investigaciones sobre expresiones poliléxicas desarrolladas desde disciplinas como la Lingüística General, la Terminología, la Lingüística Computacional y la Psicolingüística. En traducción, los términos equivalentes suelen estar formados por un número y tipo diferente de elementos. En el caso de los TC, las traducciones literales no suelen ser la norma (p. ej. *freno aerodinámico* > *flap*). Por este motivo, la traducción de los TC debe basarse en un profundo análisis conceptual y contextual como el que proporcionan los microcontextos, gracias a los cuales se dispone de los instrumentos conceptuales necesarios antes de establecer equivalencias.

Por consiguiente, los microcontextos permiten acceder al sistema conceptual subyacente, elaborar definiciones, aclarar términos complejos e inferir nuevos TC tanto en la lengua de origen como en la de destino. Asimismo, el análisis semántico por medio de categorías, roles y relaciones también resulta de utilidad para la comprensión de textos y su traducción. En particular, este análisis reveló la presencia de variación terminológica y conceptual, que repercuten en distintas tareas discursivas, así como la utilidad de las paráfrasis para la extracción de información conceptual y contextual.

AGRADECIMIENTOS

Investigación financiada con fondos del proyecto PID2020-118369GB-I00: «Integración transversal de la cultura en una base de conocimiento terminológico medioambiental», TRANSCULTURE, concedido por el Ministerio de Ciencia e Innovación de España.

REFERENCIAS

- BOUILLON, P., JEZEK, E., MELLONI, C. & PICTON, A. (2012): "Annotating qualia relations in Italian and French complex nominals", Calzolari, N. *et al.* (eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Estambul: ELRA, pp. 1527-1532. Disponible en: <https://archive-ouverte.unige.ch/unige:30915> (Fecha de consulta: 17/11/2020).
- BOWKER, L. (1998): "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study", *Meta*, 43(4), pp. 631-651. Disponible en: <https://www.erudit.org/fr/revues/meta/1998-v43-n4-meta169/002134ar.pdf> (Fecha de consulta: 09/11/2020).
- CABEZAS-GARCÍA, M. (2020): *Los términos compuestos desde la Terminología y la Traducción*. Berlín: Peter Lang. <https://doi.org/10.3726/b17354>.

- CABEZAS-GARCÍA, M. & FABER, P. (2017): "A Semantic Approach to the Inclusion of Complex Nominals in English Terminographic Resources", Mitkov, R. (coord.): *Computational and Corpus-Based Phraseology*, Lecture Notes in Computer Science, 10596, Cham: Springer, pp. 145-159. https://doi.org/10.1007/978-3-319-69805-2_11.
- CABEZAS-GARCÍA, M. & FABER, P. (2018): "Phraseology in specialized resources: an approach to complex nominals", *Lexicography*, 5(1), pp. 55-83. <https://doi.org/10.1007/s40607-018-0046-x> (Fecha de consulta: 16/11/2020).
- CABEZAS-GARCÍA, M. & LEÓN-ARAÚZ, P. (2019): "On the Structural Disambiguation of Multi-word Terms", Corpas Pastor, G. y Mitkov, R. (coords.): *Computational and Corpus-Based Phraseology*, Lecture Notes in Computer Science, 11755, Cham: Springer, pp. 46-60. https://doi.org/10.1007/978-3-030-30135-4_4 (Fecha de consulta: 17/11/2020).
- CASAS GÓMEZ, M. (2020): "Conceptual relationships and their methodological representation in a dictionary of the terminological uses of lexical semantics", *Fachsprache*, 42(1-2), pp. 2-26. <https://doi.org/10.24989/fs.v42i1-2.1789> (Fecha de consulta: 16/11/2020).
- CORPAS PASTOR, G. (1996): *Manual de fraseología española*. Madrid: Gredos.
- DÍAZ HORMIGO, M. T. & VEGA MORENO, É. (2018): "Algunas de las aplicabilidades actuales de las investigaciones en neología y sobre los neologismos", *Pragmalingüística*, 26, pp. 54-68. DOI: 10.25267/Pragmalingüística.2018.i26.03 (Fecha de consulta: 10/10/2020).
- DOWNING, P. (1977): "On the creation and use of English compound nouns", *Language*, 53, pp. 810-842. <https://doi.org/10.2307/412913> (Fecha de consulta: 05/11/2020).
- FABER, P. & CABEZAS-GARCÍA, M. (2019): "Specialized Knowledge Representation: From Terms to Frames", *Research in Language*, 17(2), pp. 197-211. <https://doi.org/10.2478/rela-2019-0012> (Fecha de consulta: 05/11/2020).
- FERNÁNDEZ DOMÍNGUEZ, J. (2019): "Compounds and multi-word expressions in Spanish", Schlücker, B. (coord.): *Complex Lexical Units. Compounds and Multi-Word Expressions*, Berlin, Boston: De Gruyter, pp. 189-219.
- FERNÁNDEZ SILVA, S. & KERREMANS, K. (2011): "Terminological variation in source texts and translations: a pilot study", *Meta. Journal des Traducteurs*, 56(2), pp. 318-335. Disponible en: <https://www.erudit.org/en/journals/meta/2011-v56-n2-meta1821316/1006179ar/> (Fecha de consulta: 04/11/2020).
- FILLMORE, C. J. (1982): "Frame semantics", The Linguistic Society of Korea (coord.): *Linguistics in the Morning Calm*, Hanshin, pp. 111-137.
- GAGNÉ, C. L. (2000): "Relational-Based Combinations Versus Property-Based Combinations: A Test of the CARIN Theory and the Dual-Process Theory of Conceptual Combination", *Journal of Memory and Language*, 42, pp. 365-389. doi:10.1006/jmla.1999.2683 (Fecha de consulta: 02/11/2020).
- HEID, U. (2001): "Collocations in sublanguage texts: Extraction from corpora", Wright, S. E. y Budin, G. (eds.): *Handbook of Terminology Management. Volume 2: Applications-Oriented Terminology Management*, Amsterdam/Filadelfia:

- John Benjamins, pp. 788-808. <https://doi.org/10.1075/z.htm2.31hei> (Fecha de consulta: 11/11/2020).
- JESPERSEN, O. (1942): *A modern English grammar: On historical principles*. Copenhagen: Munksgaard.
- KAGEURA, K. (2002): *The dynamics of terminology: A descriptive theory of term formation and terminological growth*. Amsterdam: John Benjamins.
- KILGARRIFF, A., BAISA, V., BUŠTA, J., JAKUBÍČEK, M., KOVÁŘ, V., MICHELFEIT, J., RYCHLÝ, P. & SUCHOMEL, V. (2014): "The Sketch Engine: ten years on", *Lexicography*, 1(1), pp. 7-36. Disponible en: <https://link.springer.com/article/10.1007/s40607-014-0009-9> (Fecha de consulta: 27/10/2020).
- KIPPER, K., SNYDER, B. & PALMER, M. (2004): "Extending a verb-lexicon using a semantically annotated corpus", Calzolari, N. et al. (eds.), *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisboa: ELRA, pp. 1557-1560. Disponible en: <http://www.lrec-conf.org/proceedings/lrec2004/pdf/84.pdf> (Fecha de consulta: 10/11/2020).
- LAUER, M. (1995): *Designing Statistical Language Learners: Experiments on Noun Compounds*. Tesis doctoral. Sidney: Macquarie University.
- LEÓN-ARAÚZ, P. & CABEZAS-GARCÍA, M. (2020): "Term and translation variation of multiword terms", *MonTI. Monografías de Traducción e Interpretación*, número especial 6, pp. 210-247. <https://doi.org/10.6035/MonTI.2020.ne6.7> (Fecha de consulta: 04/11/2020).
- LEVI, J. (1978): *The Syntax and Semantics of Complex Nominals*. Nueva York: Academic Press.
- L'HOMME, M. C. (2000): "Understanding Specialized Lexical Combinations", *Terminology*, 6(1), pp. 89-110. <https://doi.org/10.1075/term.6.1.06hom> (Fecha de consulta: 04/11/2020).
- MAGUIRE, P., WISNIEWSKI, E. J. & STORMS, G. (2010): "A corpus study of semantic patterns in compounding", *Corpus Linguistics and Linguistic Theory*, 6(1), pp. 49-73. <https://doi.org/10.1515/cilt.2010.003> (Fecha de consulta: 01/10/2020).
- MARTIN, W. (1992): "Remarks on Collocations in Sublanguages", *Terminologie et Traduction*, 2(3), pp. 157-164.
- MEYER, I. & MACKINTOSH, K. (1996): "Refining the terminographer's concept-analysis methods: How can phraseology help?", *Terminology*, 3(1), pp. 1-26.
- MONTERO MARTÍNEZ, S. (2008): "A constructional Approach to Terminological Phrases", Bernal, E. y DeCesaris, J. (eds.), *XIII EURALEX International Congress*, Barcelona: Institut Universitari de Lingüística Aplicada/Documenta Universitaria, pp. 1015-1022. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5065321> (Fecha de consulta: 04/10/2020).
- MURPHY, G. L. (1988): "Comprehending complex concepts", *Cognitive Science*, 12, pp. 529-562. https://doi.org/10.1207/s15516709cog1204_2 (Fecha de consulta: 07/10/2020).
- NAKOV, P. (2013): "On the interpretation of noun compounds: syntax, semantics, and entailment", *Natural Language Engineering*, 19, pp. 291-330. <https://doi.org/10.1017/S1351324913000065> (Fecha de consulta: 15/10/2020).
- NAKOV, P. & HEARST, M. (2006): "Using Verbs to Characterize Noun-Noun Relations", Euzenat, J. y Domingue, J. (eds.):

- Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2006*, Berlin: Springer, pp. 233-244. https://doi.org/10.1007/11861461_25 (Fecha de consulta: 04/10/2020).
- ROSARIO, B., HEARST, M. & FILLMORE, C. J. (2002): "The descent of hierarchy, and selection in Relational Semantics", Isabelle, P. (ed.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Filadelfia: ACL, pp. 247-254. DOI: 10.3115/1073083.1073125 (Fecha de consulta: 10/10/2020).
- ŠTEKAUER, P., VALERA, S. & KÖRTVÉLYESSY, L. (2012): *Word-Formation in the World's Languages*. Cambridge: Cambridge University Press.