



**UNIVERSIDAD
DE GRANADA**

Facultad de Traducción e Interpretación

GRADO EN TRADUCCIÓN E
INTERPRETACIÓN

MEMORIA DEL TRABAJO FIN
DE GRADO

**THE TRANSLATION OF
COMPLEX NOMINALS IN
THE FIELD OF AIR
QUALITY TREATMENT**

Presentado por:

D.^a Sandra Gutiérrez Bullón

Tutora:

Prof.^a Dra. Pilar León Araúz

El trabajo de fin de grado titulado *The translation of complex nominals in the field of air quality treatment* constituye una aproximación a la traducción de los términos compuestos, también llamados términos multipalabra o unidades poliléxicas (p. ej. *long range transboundary air pollution*). Estos son especialmente frecuentes en el lenguaje científico y constituyen uno de los principales rasgos de la comunicación especializada entre expertos (Sager et al., 1980; Cabré, 1993; Nakov y Hearst, 2013; *inter alia*), de ahí su importancia para la traducción especializada. La descodificación, y por lo tanto comprensión, de estos términos, caracterizados por la yuxtaposición de sus formantes, supone un gran reto para lingüistas, terminólogos y, por supuesto, traductores.

En las últimas décadas, los estudios realizados sobre estos términos se habían centrado fundamentalmente en los términos compuestos formados por dos elementos (Lauer, 1995; Nakov, 2013; Kim y Baldwin, 2013; *inter alia*). No obstante, desde el grupo LexiCon, del Departamento de Traducción e Interpretación de la Universidad de Granada, se han dado varios pasos más allá en el estudio de estos términos en dominios especializados, concretamente el medioambiental (Cabezas-García y Faber, 2017; Cabezas-García y León-Araúz, 2018; Cabezas-García y Faber, 2019, *inter alia*). En este trabajo de fin de grado se continuó con esta línea de investigación desde la perspectiva de la traducción especializada, con el estudio de los términos compuestos formados por tres o más elementos en el ámbito de la contaminación y el tratamiento de la calidad del aire. Uno de los motivos principales que justifican el ámbito de conocimiento elegido es el hecho de que la preocupación por las cuestiones medioambientales ha aumentado considerablemente y, con ella, la investigación sobre estos temas. Más concretamente, la publicación de artículos científicos sobre la contaminación atmosférica ha crecido exponencialmente en el siglo XXI, por lo que resulta fundamental estudiar la terminología del ámbito en profundidad, contribuyendo a facilitar las traducciones de sus textos especializados y mejorando la calidad y eficacia de las mismas. Además, este estudio se realizó dentro del marco de TOTEM: Herramientas terminológicas orientadas hacia la traducción de textos medioambientales (FFI2017-89127-P), proyecto I+D+i en el que trabaja actualmente el grupo LexiCon.

Para la realización del estudio se compilaron manualmente dos corpus, uno en inglés y otro en español, de 1.500.000 palabras cada uno. Si bien ambos contienen textos de la misma temática, el corpus en inglés se compone fundamentalmente de artículos científicos, mientras que el español se nutre principalmente de tesis doctorales. Esto se

debe a la escasez de publicaciones científicas en español (con excepción del ámbito académico), siendo el inglés la *lingua franca* de la comunicación entre expertos a nivel internacional. Estos textos se procesaron con la ayuda de la herramienta de análisis de corpus Sketch Engine, que permite generar concordancias, colocaciones y listas de frecuencia (Killgarrif et al. 2004). Así, se ejecutaron una serie de búsquedas en esta herramienta con el objetivo de (1) extraer una selección representativa de los compuestos del ámbito (en base a criterios de complejidad, frecuencia, propensión a proliferación de variantes, etc.) y (2) decodificar los mismos a nivel estructural y semántico.

Además de estos materiales, se consultaron los siguientes recursos: bases terminológicas en línea (TERMIUMPlus, IATE), corpus paralelos (EurLex, disponible a través de Sketch Engine, y Linguee, disponible en la web) y comparables (Google Scholar). Esto nos permitió comparar con mayor precisión las variantes denominativas que proliferaron en lengua meta, analizadas según criterios estructurales y semánticos (omisiones, transposiciones, permutaciones, etc.), ya que tanto su tipología como su motivación influyeron en el tipo de equivalencia. Así, la variación cognitiva, fruto de la multidimensionalidad y la dinamicidad que caracteriza a los términos estudiados, se constituyó como una de las principales causas de la proliferación de distintas formas en el ámbito de conocimiento de contaminación y el tratamiento de la calidad del aire.

Dada la relevancia de la investigación en términos compuestos para el procesamiento de lenguaje natural (PLN), se hizo además uso de cuatro sistemas de traducción automática (TA) disponibles en línea (DeepL, Google Translate, Apertium y Systran) para analizar las traducciones instantáneas de los términos en la combinación lingüística que nos ocupa (EN>ES). Así, para evaluar el rendimiento de estos motores de TA en la traducción de estas unidades, se establecieron una serie de parámetros estructurados en torno a las nociones de precisión y fluidez, las cuales resultan clave en la evaluación de la traducción tanto humana como automática en la actualidad.

Si bien los aciertos de las traducciones instantáneas proporcionadas por estos sistemas (sobre todo en la traducción de términos de tres formantes) reflejan grandes avances fruto de la investigación en PLN, los resultados obtenidos sacaron a relucir varias de las principales limitaciones de los motores de TA en la actualidad. En efecto, la casuística de errores observada puso de manifiesto la vital importancia de la desambiguación o *bracketing* de los compuestos formados por tres o más términos, así como del

establecimiento de equivalentes que cumplan con las expectativas del dominio y de la lengua meta en cuanto a terminología.

Finalmente, en base a nuestras observaciones, confeccionamos el diseño preliminar de un protocolo de búsqueda basado en corpus y otros recursos orientado hacia la traducción al español de términos compuestos en inglés, del que se pudieran beneficiar traductores tanto profesionales como en formación.

Durante la realización de este trabajo de fin de grado se pusieron en práctica varias de las competencias adquiridas durante la formación en el Grado de Traducción e Interpretación, entre las que se encuentran:

- la competencia lingüística y textual en lengua A (español) y lengua B (inglés);
- la capacidad de aplicar los conocimientos de los procesos de codificación y descodificación lingüísticos asociados a los procesos de traducción;
- la capacidad de aplicar los conocimientos sobre documentación, adquisición y procesamiento de la información;
- la capacidad de aprender en autonomía;
- la competencia tecnológica, con la utilización de diferentes recursos terminográficos en línea (TERMIUM Plus, IATE), editores de texto y código fuente (Notepad++) y software de gestión de corpus y análisis de texto (Sketch Engine);
- la competencia traductora y
- la capacidad de resolver problemas, de analizar y de sintetizar, entre otras.

El presente trabajo de fin de grado se centró en el estudio de la traducción de términos compuestos, tanto desde el punto de vista del traductor humano como desde la óptica de los traductores automáticos. Este doble enfoque permitió constatar que, a pesar de los avances logrados desde distintas disciplinas como la Terminología o el PLN, la traducción de los términos compuestos —en particular, de aquellos formados por más de tres elementos— sigue suponiendo a día de hoy uno de los mayores retos a los que se enfrentan tanto los traductores humanos como los motores de TA.

Las conclusiones extraídas de este estudio experimental abren la puerta a nuevas investigaciones futuras, como la extrapolación de esta metodología a otros dominios especializados y al estudio de los términos compuestos en otras combinaciones lingüísticas o la evaluación del protocolo diseñado en el aula de traducción especializada.

Referencias bibliográficas

- Cabezas García, M. & Faber, P. (2019). Semantic prosody and semantic preference in multi-word terms. *Fachsprache* 41(1-2), 2-21.
- Cabezas García, M. & León Araúz, P. (2018). Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study. In N. Calzolari et al. (eds.), *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, 2511-2518. Miyazaki: ELRA.
- Cabezas-García, M. & Faber, P. (2017). The role of micro-contexts in noun compound formation. *Neologica. La néologie en terminologie* 11, 101-118.
- Cabré, M. T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Antártida, Empúries.
- Kilgarriff, A., Rychlý, P., Smrz, P. & Tugwell, D. (2004). The Sketch Engine. In G. Williams y S. Vessier (eds.), *Proceedings of the 11th EURALEX International Congress* (105-115). Lorient: EURALEX.
- Kim, S. N. & Baldwin, T. (2013). A lexical semantic approach to interpreting and bracketing English noun compounds. *Natural Language Engineering* 19(3), 385-407.
- Lauer, M. (1995). *Designing Statistical Language Learners: Experiments on Noun Compounds*. Sydney: Macquarie University.
- Nakov, P. & Hearst, M. (2013). Semantic interpretation of noun compounds using verbal and other paraphrases. *ACM Transactions on Speech and Language Processing* 10(3), 1-51.
- Nakov, P. (2013). On the interpretation of noun compounds: syntax, semantics, and entailment. *Natural Language Engineering* 19, 291-330.
- Sager, J. C., Dungworth, D. & McDonald, P. F. (1980). *English special languages. Principles and practice in science and technology*. Wiesbaden: Brandstetter Verlag.