

- Calvino, I. (1979/1980). *Si una noche de invierno un viajero* (E. Benítez, Trad.). Bruguera.
- Campos, H. de (1979). *Deus e o Diabo no Fausto de Goethe*. Perspectiva.
- Cervantes Saavedra, M. de (1605/2004). *Don Quijote de la Mancha* (M. de Riquer, Edición, introducción y notas). Planeta.
- Foer, J. S. (2010). *Tree of codes*. Visual.
- Rothenberg, J. (2020). Three variations on Octavio Paz's "Blanco" and fifteen antiphonals for Haroldo de Campos, with a note on translation, transcreation, and othering. En M. Librandi, J. Pinheiro Dias, y T. Winterbottom (Eds). *Transpoetic exchange. Haroldo de Campos, Octavio Paz, and other multiversal dialogues* (pp. 113-121). Bucknell University Press.

### **Computational and Corpus-Based Phraseology**

GLORIA CORPAS PASTOR  
Y RUSLAN MITKOV (EDS.)

Cham (Suiza), Springer, 2022, 252 pp.

Francisco Javier Lima Florido  
Universidad de Málaga  
ORCID: 0009-0003-2622-3843



El estudio de la fraseología basada en corpus y la fraseología computacional es un área que actualmente posee una importante relevancia para investigadores y profesionales de diversas áreas de la lingüística y la computación. Especialmente para traductores e intérpretes,

así como ingenieros e investigadores que se dedican al desarrollo de tecnologías aplicadas a la traducción e interpretación. La traducción e interpretación de unidades fraseológicas posee es-

pecial dificultad debido a su estructura compleja y su naturaleza idiomática. La fraseología basada en corpus otorga un mejor entendimiento de las unidades fraseológicas que permite a traductores e intérpretes estar más preparados en su trabajo (Corpas Pastor y Gaber, 2021; Ramish, 2020). A nivel computacional, las unidades fraseológicas también suponen una dificultad añadida en tareas de procesamiento de lenguaje natural, sin embargo, las tecnologías y métodos actuales basados en corpus han dado lugar a que surjan nuevas técnicas de procesamiento de unidades y expresiones fraseológicas. En este sentido, se pueden encontrar en la literatura actual publicaciones que abordan el procesamiento de unidades fraseológicas desde un punto de vista puramente computacional (Corpas Pastor y Colson, 2020; Ramish y Villavicencio Aline, 2022); otras cuyo enfoque se centra en traducción automática y tecnologías de la traducción (Bowker y Corpas Pastor, 2022; Mitkov *et al.*, 2018); e incluso volúmenes donde la fraseología computacional es un tema recurrente (Mitkov, 2022). A pesar de este auge, existe aún una necesidad de nuevos métodos y estudios en el área debido, fundamentalmente, a la amplitud y la novedad que caracterizan a sus líneas de investigación. Estas son las principales motivaciones que llevan a la publicación del volumen *Computational and Corpus-Based Phraseology*.

El volumen sobre el que trata esta reseña, *Computational and Corpus-Based Phraseology*, engloba una selección de los artículos que se presentaron en el congreso EUROPHRAS 2022 celebrado durante los días 28 al 30 de septiembre de 2022 en Málaga. Se trata de uno de los congresos internacionales que más destaca en la fraseología computacional y basada en corpus. El volumen resultante incluye 16 capítulos, siendo cada capítulo un artículo de la selección realizada. Los capítulos cubren diversos temas tales como la

290 fraseología en traducción e interpretación, unidades fraseológicas en computación y fraseología en traducción automática, entre muchos otros. Para esta reseña se han identificado cuatro grandes bloques temáticos: Un primer bloque sobre estudios fraseológicos basados en corpus y otros recursos, que engloba seis capítulos; un segundo bloque que incluye tres capítulos en el que se presentan estudios fraseológicos aplicados a la enseñanza de idioma; el tercer bloque cubre dos artículos sobre fraseología aplicada en traducción e interpretación; y un cuarto bloque formado por cinco artículos que tratan sobre recursos y métodos de fraseología computacional aplicados a tecnologías de la traducción e interpretación.

El primer bloque sobre estudios fraseológicos basados en corpus y otros recursos engloba los capítulos que presentan estudios lingüísticos basados en corpus, diccionarios, glosarios o tesauros. El primer capítulo titulado “Constructing the Digital Proverbial Thesaurus: Theoretical and Methodological Implications”, escrito por Melita Aleksa Varga y Kristina Feldvari, presenta una metodología para la creación de un tesauro paremiológico de expresiones idiomáticas y refranes para el idioma croata. De esta forma, las posibles búsquedas en internet relacionadas con refranes que realizan los usuarios se ven enriquecidas con otros posibles resultados relacionados que se les hayan podido pasar por alto.

Otro capítulo basado en el estudio de diccionarios y glosarios es el escrito por Cristiano Furiassi, titulado “Detecting Bajan Phraseology: A Metalexigraphic Analysis”. En este trabajo se presenta un análisis de las unidades fraseológicas idiomáticas más frecuentemente utilizadas en *bajan*, la lengua criolla que se habla en las Islas Barbados. El autor ha realizado un análisis de entradas en diccionarios profesionales específicos sobre inglés del caribe comparándolos con glosarios no profesionales sobre del *bajan* y el

inglés de las Barbados. De este análisis el autor destaca la influencia de las épocas de esclavismo en las expresiones y unidades fraseológicas detectadas.

También relacionado con expresiones fraseológicas en inglés, el capítulo “Make + Adjective Combinations During the Years 1850–1999: A Corpus-Based Investigation”, escrito por Ljubica Leone, presenta un estudio de las construcciones en inglés formadas por *make* y un adjetivo (por ejemplo, *make clear*) desde el año 1850 hasta el 1999. Para realizar este estudio, la autora ha sacado provecho del corpus ARCHER. Para su estudio ha seleccionado un subcorpus que cubra solo el inglés británico durante el período señalado. Como conclusión, la autora destaca que el uso de este tipo de expresiones se estabilizó a partir de 1900 en cuanto a frecuencia, patrones y variabilidad fraseológica.

Siguiendo con estudios en cuanto al uso del lenguaje, el capítulo titulado “Some Insights on a Typology of French Interactional Prefabricated Formulas in Spoken Corpora” escrito por Marie-Sophie Pausé y Agnès Tutin presenta un análisis topológico de expresiones rutinarias o fórmulas de conversación. Las autoras afirman que existe una laguna en cuanto al análisis fraseológico de ciertas expresiones que se utilizan en interacciones tanto orales como escritas. Para cubrir dicha laguna, han combinado métodos basados en corpus para identificar unidades fraseológicas de interacción y clasificarlas, creando una tipología validada mediante un sistema de anotación. Como conclusión, este trabajo ha llevado al descubrimiento de una gran diversidad de fórmulas relacionadas con tres aspectos: el tipo de interacción, la proximidad de los hablantes y el estado de estos en dicha interacción.

En el capítulo escrito por José Luis Rojas Díaz, Juan Manuel Pérez Sánchez y Alejandro Arroyave Tobón que se titula “Specialized Idioms:

From LGP to LSP Phraseological Paradigm” se presenta un estudio centrado en el uso del lenguaje sobre unidades fraseológicas especializadas analizadas mediante métodos aplicados a lenguaje general. Concretamente, se analizan cuatro unidades fraseológicas especializadas del campo de la economía y el comercio en español, inglés y francés. De esta forma, han creado una taxonomía dividida en tres subcategorías que revela la existencia de expresiones idiomáticas en lenguajes especializados con características similares a las que se pueden encontrar en el lenguaje general.

Terminando este bloque con un último capítulo sobre estudios de corpus y uso del lenguaje, en el estudio titulado “Figurative Expressions with Verbs of Ingesting in Croatian” escrito por Jelena Parizoska y Jelena Tušek se investigan expresiones en croata con verbos relacionados con *comer* analizando el corpus web hrWaC. Las autoras han observado una mayor tendencia en cuanto a los verbos referentes a las formas de comer en ser usados de forma figurada que literal, mientras que los verbos sinónimos de *comer* se utilizan frecuentemente de forma literal.

El siguiente bloque versa sobre fraseología aplicada a la enseñanza de idiomas. En este se encuentran aquellos capítulos centrados en la creación de recursos y métodos que puedan aplicar docentes dedicados a enseñar de un segundo idioma. El primer capítulo que trata esta temática es “Lexical Semantic Mind Maps Based on Collocations as a Tool for Teaching Vocabulary: A Case Study” escrito por María Auxiliadora Barrios Rodríguez en el que se presenta una nueva herramienta para enseñanza de vocabulario y colocaciones en el aprendizaje de idiomas a la que han denominado *lexical-semantic mind maps* que establece relaciones funcionales entre diferentes grupos de palabras. En un experimento realizado con este nuevo recurso, demostraron

que al utilizar un *lexical-semantic mind map* diseñado específicamente para el contexto estudiado la capacidad de retención de conceptos de los participantes mejoraba notablemente.

También en este bloque encontramos otro capítulo titulado “Metonymy in Spanish/L2 Teaching: A Cognitive Analysis of Color Idioms and Their Inclusion in the Córdoba Project Database” centrado en la enseñanza de fraseología, vocabulario y expresiones idiomáticas, pero con un enfoque diferente. En este capítulo escrito por Beatriz Martín-Gascón se presenta un análisis cognitivo de la metonimia en español para el desarrollo de recursos didácticos que puedan servir para la enseñanza del idioma. Se centra específicamente en expresiones idiomáticas con colores que ayuden al alumnado apreciar el papel de la metonimia en el idioma que está aprendiendo.

Para finalizar este bloque, el último capítulo incluido trata sobre enseñanza de expresiones idiomáticas. En “Teaching and Learning French Formulas” escrito por Elena Berthemet se presenta un estudio sobre qué tipo de información necesitan los estudiantes de un idioma sobre fórmulas para mejorar sus capacidades pragmáticas en dicho idioma. Concretamente, aquellas expresiones fraseológicas que se utilizan rutinariamente en un idioma como reacciones en conversaciones usualmente acompañadas de gestos y entonaciones concretas. La autora propone un modelo descriptivo de fórmulas en francés que no solo presenta su definición, si no que va acompañado de ejemplos conversacionales con audios y descripciones.

El tercer bloque temático identificado es el que cubre aquellos capítulos relacionados con la aplicación de métodos y técnicas de fraseología aplicados a la traducción e interpretación. En el capítulo titulado “Multi-word Term Translation: A Student-Centered Pilot Study” escrito por Sandra Bullón y Pilar León-Auraúz se

292 presenta un estudio sobre las dificultades de traducir términos multipalabra (MWT de sus siglas en inglés) para estudiantes de traducción. Las autoras han analizado tanto el punto de vista de estudiantes no graduados como de estudiantes de posgrado, concluyendo que se hace necesario incluir las características y peculiaridades de este tipo de términos en las asignaturas de traducción especializada.

Por su parte, Gloria Corpas Pastor presenta una novedosa metodología para mejorar la preparación de una interpretación para la que a menudo no es posible contar con tiempo ni recursos. En el capítulo titulado “Interpreting Tomorrow? How to Build a Computer-Assisted Glossary of Phraseological Units in (Almost) No Time” la autora presenta como solución un método rápido de documentación y creación de glosarios fraseológicos especializados para la terna de idiomas inglés, francés y español utilizando el sistema VIP de asistencia a la interpretación. La metodología presentada permite la creación de recursos avanzados para la preparación de una interpretación en un corto espacio de tiempo, ofreciendo a los intérpretes la oportunidad de obtener un mayor rendimiento en su trabajo.

En el último bloque se sitúan los capítulos que tratan sobre métodos y técnicas aplicados a la creación y/o evaluación de tecnologías de la traducción e interpretación. El primer capítulo que se incluye en este bloque se presenta un experimento sobre traducción automática de refranes del árabe al español. En “Readability and Communication in Machine Translation of Arabic Phraseologisms into Spanish” Mohamed El-Madkouri y Beatriz Soto Aranda han seleccionado un total de 20 refranes en árabe de los cuales algunos cuentan con equivalentes al español y otros no. Los resultados obtenidos muestran que los traductores automáticos aún no cuentan con un buen rendimiento en la traducción de refranes.

Los autores proponen la creación de un corpus específico de refranes en el par de lenguas español-árabe que permita solucionar este problema.

Por su parte, en el capítulo titulado “Author Gender Identification for Urdu Articles” Raheem Sarwar presenta un experimento sobre identificación del género de autores de artículos escritos en urdu, que es también un lenguaje con pocos recursos. La aportación del autor a este problema ha consistido primero en la creación de un corpus específico en urdu para dicha tarea, y segundo la creación de un espacio vectorial a partir de dicho corpus. Para probar la efectividad de su metodología, el autor ha aplicado el corpus creado a entrenamiento de diferentes algoritmos de aprendizaje computacional. Tras el experimento, los resultados revelan que la estrategia de combinar unidades fraseológicas con palabras frecuentes ofrece buenos resultados de detección del género del autor.

También en el capítulo titulado “Bootstrapping a Lexicon of Multiword Adverbs for Brazilian Portuguese” escrito por Izabela Müller, Nuno Mamede y Jorge Baptista se expone el proceso de creación de un recurso con utilidad en tareas de fraseología computacional. Concretamente, se presenta el proceso de *bootstrapping* que se ha seguido para generar un lexicón de adverbios compuestos en portugués brasileño a partir de uno ya existente del portugués de Portugal. Tratándose de un recurso computacional, los autores afirman que puede ser utilizado en tareas relacionadas con la detección automática de variedades de un idioma. La tarea ha resultado en la inclusión de 700 nuevos adverbios en el lexicón que ofrece nuevas capacidades de procesamiento de lenguaje natural.

En cuanto a recursos aplicados a fraseología computacional, en el capítulo “Image Schemas and Image Schematic Complexes: Enhancing Neural Machine Translation Networks” escrito

por Amal Haddad Haddad se presenta una metodología que permite incluir esquemas gráficos de conceptos abstractos en el entrenamiento de redes neuronales para la mejora de la traducción automática. Esta metodología demuestra que la inclusión de esquemas gráficos permite representar dichos conceptos de una forma más correcta, mejorando la precisión de los traductores automáticos.

Terminando este bloque y con el último capítulo que quedaba por tratar en esta reseña, en “Reassessing gApp: Does MWE Discontinuity Always Pose a Challenge to Neural Machine Translation?” Carlos Manuel Hidalgo-Tertero y Xiaqing Zhou-Lian presentan un experimento realizado con la herramienta *gApp* aplicada a traducción automática de somatismos desde el español al inglés y desde el español al chino. El experimento confirma que la herramienta consigue mejorar la precisión de la traducción automática, sin embargo, no siempre es necesario realizar las conversiones de ciertas expresiones idiomáticas.

Como apreciación final, la calidad de los trabajos incluidos en este volumen, sumado al criterio de sus editores para seleccionarlos, dejan notar por qué el congreso EUROPHRAS 2022 es un referente en el ámbito de la fraseología computacional y basada en corpus. Este volumen cumple su objetivo de cubrir en parte la laguna investigadora que aún existe en el área. Los resultados que se presentan en cada capítulo poseen un alto valor añadido a la investigación en temáticas tan punteras como el procesamiento de lenguaje natural, la lingüística de corpus, o el desarrollo de tecnologías lingüísticas aplicadas a traducción e interpretación. Por otro lado, estos trabajos tienen especial valor en traducción e interpretación dado que gran parte de los capítulos dejan constancia de cómo el estudio de la fraseología basada en corpus permite entender

cuestiones como el aprendizaje de conceptos a nivel cognitivo, la importancia de los patrones fraseológicos en el uso del lenguaje, la traducción de expresiones conversacionales o las características de construcciones idiomáticas en lenguaje especializado. Se trata de una obra con gran relevancia en el futuro tanto de la investigación en fraseología computacional y basada en corpus como de sus aplicaciones.

## REFERENCIAS

- Bowker, L., y Corpas Pastor, G. (2022). Translation Technology. En R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics 2nd edition* (pp. 871-905). Oxford: Oxford University Press.
- Corpas Pastor, G., y Colson, J. (2020). *Computational Phraseology*. Amsterdam: John Benjamins.
- Corpas Pastor, G., Gaber, M.: Extracción de fraseología para intérpretes a partir de corpus comparables compilados mediante reconocimiento automático del habla. In: Corpas Pastor, G., Bautista Zambrana, M.R., Hidalgo Tertero, C.M. (eds.) *Sistemas fraseológicos en contraste: enfoques computacionales y de corpus*, pp. 271–291. Granada, Comares (2021). <http://hdl.handle.net/2436/624419>
- Mitkov, R. (Ed.). (2022). *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.
- Mitkov, R., Monti, J., Corpas Pastor, G., y Seretan, V. (2018). Multiword Units in Machine Translation and Translation Technology. *Multiword Units in Machine Translation and Translation Technology*, 1-269. Amsterdam: John Benjamins.
- Ramish, C. (2020). Computational phraseology discovery in corpora with the MWEToolKIT. En Gloria Corpas Pastor; Jean-Pierre Colson (Eds.). *Computational Phraseology* (pp. 111-134). Amsterdam: John Benjamins. <https://hal.science/hal-02739265>
- Ramish, C., y Villavicencio Aline. (2022). Computational treatment of multiword expressions. En R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics* (pp. 649-678). Oxford: Oxford University Press.