Hermenēus
Revista de traducción e interpretación

EDICIONES Universidad Valladolid

Marlén Izquierdo y Zuriñe Sanz-Villar, *Corpus Use in Cross-Linguistic Research: Paving the Way of Teaching Translation and Professional Communication*, Ámsterdam, John Benjamins, 2023, 237 págs.

Reseña de acceso abierto distribuida bajo una Licencia Creative Commons Atribución 4.0 Internacional (CC-BY 4.0). / Open access review under a Creative Commons Attribution 4.0 International License (CC-BY 4.0).
DOI: https://doi.org/10.24197/r24cwr83

En las dos últimas décadas, la investigación lingüística basada en corpus ha vivido un notable crecimiento y diversificación: reflejo de ello es que el número de congresos y publicaciones en este ámbito no ha dejado de aumentar, generando nuevos espacios de colaboración entre investigadores y profesionales de distintas áreas del conocimiento que han permitido refinar los métodos para crear, anotar, procesar y aplicar metodologías basadas en corpus.

Los estudios de corpus, tanto comparables como paralelos, así como la combinación de ambos, han permitido llevar la descripción lingüística a otro nivel, impulsando aplicaciones en ámbitos como la enseñanza de idiomas, la formación de traductores, la accesibilidad o procesamiento de lenguaje natural. Además, en los últimos tiempos se ha puesto el foco no solo en los avances metodológicos y técnicos, sino también en el estudio de dominios y lenguas de menor visibilidad o accesibilidad.

El volumen *Corpus Use in Cross-Linguistic Research: Paving the Way of Teaching Translation and Professional Communication*, editado por Izquierdo y Sanz-Villar, es un buen ejemplo de todo lo anterior: en él se describen proyectos y alianzas que ponen de manifiesto no solo que es necesario construir recursos específicos que fomenten la transferencia de conocimiento, la estandarización y la replicabilidad, sino que también se demuestra que hacerlo es viable, y todo ello combinando trabajos sobre ámbitos y lenguas más estudiados con áreas y pares de lenguas habitualmente menos investigados.

El primer capítulo (pp. 12-33), a cargo de Marco y Bracho Lapiedra, busca la validación empírica de la hipótesis de atracción gravitacional (del inglés *Gravitational Pull Hypothesis*) en traducciones entre inglés, francés, catalán y español. Para ello, los autores aplican un enfoque semántico e inductivo empleando los subcorpus del Corpus Valencià de Literatura

Traduïda (COVALT) (inglés-catalán, inglés-español, francés-catalán, francés-español y los monolingües de catalán y español) para procesar una selección de construcciones con verbos de apoyo (concretamente *fer*/dar + sustantivos predicativos que expresan estados emocionales y eventos dinámicos). Los resultados del estudio confirman la subrepresentación de las construcciones con verbos de apoyo emocionales y, parcialmente, de los dinámicos, sentando una base metodológica replicable para futuras investigaciones. Además, ilustra el potencial de la investigación inductiva mediante corpus en aula de traducción, lo que aporta conciencia contrastiva a los estudiantes sobre las características de la lengua traducida.

En el segundo capítulo (pp. 34-50), Rabadán explora la influencia de las características semánticas y de registro en la traducción al español de cinco construcciones con verbos de apoyo en inglés (*have, take, make, do, give*) mediante el corpus P-ACTRES 2.0, un corpus bidireccional etiquetado gramaticalmente de 4,2 millones de palabras que incluye géneros de ficción y no ficción. El análisis de las concordancias, los patrones de traducción, la asociación «rasgos semánticos-elección traductora» y la variación de registro permiten a la autora realizar una identificación sistemática de muestras representativas para cada verbo y registro que deriva en la descripción de cinco patrones recurrentes, corrobora que los rasgos semánticos de las construcciones con verbos de apoyo condicionan las opciones de traducción y evidencia variaciones según registro. Los resultados del estudio son pertinentes para la mejora de la calidad de distintos tipos de herramientas lingüísticas.

El tercer capítulo (pp. 51-66) corre a cargo de Molés-Cases, quien analiza los verbos que introducen el estilo directo en textos narrativos escritos en español y sus traducciones al alemán. Para ello examina, en el corpus PaGeS (Parallel Corpus German Spanish), 1571 concordancias bilingües para proceder a describir las técnicas de traducción empleadas en este contexto específico, lo cual le permite concluir que la transferencia del verbo de habla es la de uso mayoritario. Este trabajo no solo ayuda a conocer mejor estos usos concretos en una combinación de lenguas poco estudiada para este género, sino que además es un recurso de utilidad para desarrollar la competencia contrastiva en traducción y en la formación de los propios traductores.

Sánchez Nieto realiza en el capítulo cuarto (pp. 67-90) un análisis contrastivo de la gestión del dativo pasivo (verbos *bekommen, kriegen, erhalten*) en traducción alemán-español-alemán. Su interés radica fundamentalmente en que es una estructura que no existe en español, y centra

su análisis en la frecuencia, los roles semánticos y las estrategias de traducción empleadas en el corpus PaGeS anteriormente mencionado. Tras la extracción de datos del corpus, su limpieza manual y su etiquetado y análisis cualitativo, el estudio demuestra que la construcción con *bekommen* es la más frecuente, que suele mantenerse la perspectiva del receptor en la dirección alemán-español, siendo la simplificación la técnica de traducción más empleada, y que en español-alemán suele emplearse el dativo pasivo cuando está presente en el texto origen. La autora completa sus hallazgos con una reflexión sobre las implicaciones prácticas para la formación en traducción.

El capítulo quinto (pp. 91-107) analiza los cuasisinónimos *begin* y *start* en inglés y sus equivalentes en español con el objetivo de identificar diferencias semánticas y traductológicas. Ramón emplea para ello el corpus P-ACTRES 2.0, descrito en el capítulo dos. Tras contrastar los datos obtenidos para cada verbo, la autora presenta un análisis cuantitativo de los patrones sintácticos encontrados tanto en inglés como en las traducciones al español, así como un análisis cualitativo de las técnicas de traducción empleadas. Los resultados muestran que existen diferencias en las estructuras sintácticas empleadas con ambos verbos, pues *begin* tiende a asociarse con oraciones de infinitivo y *start* se asocia con más frecuencia a oraciones intransitivas. Además, también se han detectado diferencias semánticas relevantes: mientras que el primer verbo suele indicar el inicio de una acción, el segundo se asocia al comienzo de una actividad; por su parte, en lo que respecta a su traducción al español, se confirma que existen diferencias sutiles entre ambos cuasisinónimos.

Labrador aborda en el capítulo sexto (pp. 108-123) la complejidad en la traducción al español de un verbo aparentemente sencillo del inglés, *run*, y sus implicaciones para la enseñanza de idiomas y la traducción. Empleando el corpus de ficción y no ficción P-ACTRES 2.0, realiza un análisis intralingüístico de significados, sintaxis y colocaciones en inglés y un análisis interlingüístico de soluciones traductoras sobre las 926 ocurrencias del verbo en el corpus. Los resultados permiten concluir que el verbo *run* es más frecuente en ficción y que sus usos abarcan sentido literal y metafórico, reflejándose en traducciones variadas que incluyen técnicas como el cambio de densidad o la copia de la estructura original entre otras.

Gutiérrez Lanza proporciona información práctica para la industria del doblaje y la formación de traductores y dobladores gracias a las conclusiones alcanzadas en el capítulo séptimo (pp. 124-141). En él explora el uso y adaptación de marcadores conversacionales en el doblaje de diálogos cinematográficos del inglés al español, y realiza una comparación con guiones

no traducidos. El análisis se realiza por un lado mediante el corpus paralelo TRACEci, compuesto por textos origen en inglés, primeras versiones de traducción y versión sincronizada, y por otro lado con una muestra de guiones no traducidos en español extraída de CORPES XXI. La autora confirma que las versiones finales ven reducido el uso de los marcadores conversacionales, lo cual mejora la calidad y naturalidad de los diálogos doblados; confirma, además, que estos se encuentran muy influidos por el impacto de la sincronización y por el reto de conciliar la naturalidad con los requisitos de la sincronización labial.

El capítulo octavo (pp. 142-156), a cargo de Hermosa-Ramírez, trata la caracterización lingüística de las audiodescripciones (AD) y audiointroducciones (AI) de ópera en catalán y español, ejemplo del interés de este volumen por incluir temáticas y lenguas habitualmente menos exploradas. Tras la creación de cuatro subcorpus propios (AD_ES, AD_CAT, AI_ES, AI_CAT) a partir de guiones del Teatre del Liceu y del Teatro Real, la autora mide los parámetros de densidad léxica, variación léxica, longitud media de palabra y oración y el índice de legibilidad Flesch-Szigriszt. Tras la extracción y cuantificación de los valores mediante varias herramientas, el estudio demuestra que tanto las AD como las AI combinan rasgos de discurso planeado y espontáneo, pero que las AI tienden a acercarse más al primero (mayor longitud media y legibilidad). En última instancia, este trabajo ofrece un protocolo replicable para estudiar dominios audiovisuales poco explorados y sugiere posibles aplicaciones didácticas.

El noveno capítulo (pp. 157-178) cierra la primera parte del volumen, orientada predominantemente a la descripción lingüística. En él, Li estudia la construcción de la imagen nacional de China en la traducción periodística del *New York Times* al español y al chino mediante un corpus paralelo multilingüe (New York Times Multilingual Parallel Corpus, NYTMPC) compuesto por 753 textos alineados por párrafos y anotados manualmente. El análisis revela que los enfoques temáticos y las estrategias de reformulación de titulares difieren en las versiones en español y chino, configurando imágenes nacionales divergentes, lo cual pone de manifiesto el poder de la traducción y los medios en la construcción e interpretación de la imagen de un país. El capítulo proporciona un modelo pionero de imagología aplicada a los estudios de traducción y ofrece pautas para mejorar la práctica de la traducción periodística y la objetividad informativa.

La segunda parte del volumen, más orientada a la compilación de corpus y ejemplos de aplicaciones concretas, comienza con el capítulo décimo (pp. 179-194) a cargo de De Camillis y Contarino, quienes exploran la adaptación

y evaluación de un sistema de traducción automática (ModernMT) para textos jurídico-administrativos en italiano y la variante del alemán de Tirol del Sur. La sofisticada metodología incluye la compilación del corpus paralelo LEXB, compuesto por textos legislativos y administrativos en italiano y sus traducciones al alemán, la adaptación del motor ModernMT con los textos del corpus y la evaluación automática de la calidad de la traducción mediante métricas estándar (BLEU, chrF3) y el desarrollo de un marco de evaluación terminológica automática para medir la precisión de la terminología jurídica traducida. Se observa con este estudio una mejora en la calidad general de la traducción automática tras la adaptación al campo concreto estudiado, aunque la traducción de la terminología específica sigue revelándose como un reto a mejorar. Cabe destacar especialmente que las herramientas empleadas son de libre acceso, lo cual supone una ayuda de gran valor para los usuarios de esta variante del alemán; como valor añadido, el estudio supone un modelo replicable de investigación y adaptación de la traducción automática para lenguas o variantes lingüísticas minoritarias o con escasos recursos.

Politova, Bonetskaya, Dolgov, Frolova y Pyrkova presentan en el capítulo undécimo (pp. 195-215) la creación de un *gold standard dataset* (conjunto de datos de referencia) anotado manualmente para mejorar la alineación de palabras en la combinación de lenguas ruso-chino, de gran dificultad al tratarse de lenguas muy diferentes entre sí. En el capítulo, los investigadores de diversas disciplinas describen los procesos de elaboración del corpus paralelo ruso-chino RuZhCorp (3,5 millones de palabras) y de anotación manual de una parte para construir el *gold standard*. Tras definir de forma detallada las pautas y reglas de alineación, los autores entrenaron y evaluaron dos modelos de aprendizaje automático, comparando su rendimiento con y sin el *gold standard dataset*. Los resultados demuestran que la inclusión de ejemplos alineados manualmente mejora sustancialmente la calidad de los modelos. Las pautas de alineación y la exhaustiva metodología de este estudio se revelan como un modelo valioso para optimizar sistemas de traducción automática dada su replicabilidad con otros pares de lenguas tipológicamente distantes.

Por último, Ortego Antón describe en el capítulo duodécimo (pp. 216-233) el desarrollo de GEFEM 2.0©, una herramienta de ayuda a la redacción en inglés de fichas descriptivas de embutidos a partir de textos en español. Para su construcción se ha realizado un análisis multinivel de dos corpus, paralelo y comparable, de este género textual, describiendo en detalle su forma (estructura retórica) y contenido (términos clave), la cual ha permitido la identificación de frases prototípicas del género y la elaboración de una base

de datos terminológica bilingüe clasificada por campos semánticos. Como aportación singular a este volumen, también se describe en detalle la interfaz de la herramienta y sus posibilidades de utilización. La autora argumenta que el resultado de este trabajo es una herramienta fiable que garantiza corrección, coherencia y adecuación terminológica. No menos importante es el hecho de que supone un claro ejemplo de transferencia de conocimiento al ámbito agroalimentario, pues su desarrollo pretende facilitar la internacionalización de pymes del sector.

En conclusión, las editoras, Izquierdo y Sanz-Villar, han reunido doce estudios empíricos que muestran la versatilidad de los estudios basados en corpus en ámbitos como la traducción, la enseñanza de idiomas y la comunicación profesional. Destaca por su cobertura tanto teórica como práctica, y su rigor metodológico y variedad de ejemplos ilustrativos aportan soluciones reales. Y no menos importante, es un compendio que atiende a lenguas y dominios habitualmente poco representados. En suma, la integración de descripciones y aplicaciones convierten este libro en un manual de consulta de referencia para traductores, docentes, profesionales del procesamiento del lenguaje natural y cualquier profesional interesado en las últimas tendencias de la investigación interlingüística basada en corpus.

LETICIA MORENO PÉREZ
Universidad de Valladolid
leticia.moreno@uva.es
https://orcid.org/0000-0001-9211-7166