

Federico Zanettin, Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies, Manchester & Kinderhook, St. Jerome Publishing, 2012, 244 pp.

María Teresa ORTEGO ANTÓN
Universidad de Valladolid

La obra titulada *Translation-Driven Corpora* (2012) ofrece una aproximación a las herramientas y a los métodos de la lingüística de corpus que pueden aplicarse a los estudios de traducción. Dicha obra, cuya autoría corresponde a Federico Zanettin, se compone de ocho capítulos, cuya sinopsis se recoge en el primero de ellos. La estructura de los capítulos es semejante: la exposición teórica de los contenidos que se ilustra con ejemplos procedentes de la literatura va acompañada de actividades prácticas relacionadas con las cuestiones tratadas y se complementa con lecturas adicionales. Además, el volumen se comercializa junto a un DVD, en el que se recogen tres componentes imprescindibles para la realización de las actividades planteadas: los textos necesarios para compilar los corpus, los corpus propuestos así como los paquetes de *software* para realizar dichas actividades.

En el segundo de los capítulos, Zanettin ofrece una breve panorámica del desarrollo de la lingüística de corpus y de su aplicación metodológica a los estudios de traducción. Seguidamente, procede a centrarse en el área que lleva

por título la presente obra: los «*translation-driven corpora*». En su afán por lograr un acercamiento a esta área de conocimiento, consideramos de especial interés el diagrama de la página 11, en el que quedan resumidos los principales tipos de corpus monolingües y bilingües. El autor señala que se puede emplear una metodología basada en la lingüística de corpus para estudiar fenómenos traductológicos como la simplificación, la explicitación, la estandarización, la traducción de unidades léxicas únicas, las colocaciones atípicas o la interferencia. En este sentido, ofrece ejemplos ilustrativos de estudios de traducción basados en corpus, por ejemplo el *Translational English Corpus (TEC)*, el *Corpus of Translated Finish (CTF)* o el *English-Norwegian Parallel Corpus (ENPC)*. Además, reconoce que las aportaciones de numerosos trabajos en los que se realizan estudios contrastivos con una metodología basada en corpus han contribuido favorablemente al desarrollo de áreas como la formación de los traductores, la enseñanza de lenguas extranjeras o la lingüística computacional y, más precisamente, la traducción asistida por ordenador.

Cómo diseñar un corpus es la cuestión a la que se pretende dar respuesta en el tercero de los capítulos del volumen. Zanettin comienza con la descripción de los diversos parámetros que determinan el diseño de un corpus, tales como el tamaño, la composición, la representatividad y la comparabilidad. Sin embargo, no se limita a ofrecer una perspectiva teórica, sino que también presenta ejemplos prácticos procedentes del proyecto *Corpus of English X Italian (CEXI)*, que se llevó a cabo en la Universidad de Bolonia. Por otro lado, pone de manifiesto que el diseño de un corpus está sujeto a los permisos de los derechos de autor. No obstante, considera que para evitar solicitar dichos permisos, existe la posibilidad de seguir el enfoque empleado por Davies para compilar el *Corpus of Contemporary American English (COCA)*, es decir, utilizar fragmentos de textos. Por último, otra posibilidad es utilizar textos procedentes de internet, cuyo *copyright* no está definido, ya sea empleando la *web* como corpus, explotándola con los distintos programas que han ido emergiendo: *WebConc*, *WebCorp Live*, *WebAsCorpus* o *KWiCFinder*, o como fuente de corpus, tanto para crear corpus generales a través de herramientas como *Boot-Cat*, como para compilar corpus especializados (*DIY corpus*), que pueden analizarse con generadores de concordancias como *WordSmith Tools* o *AntConc*.

El cuarto de los capítulos está dedicado a la codificación y anotación de los corpus. Zanettin comienza definiendo y diferenciando conceptos nucleares, tales como «codificación», «anotación», «marcado» y «etiquetado». Asimismo, expone los distintos tipos de anotación: la que hace referencia a la presentación

así como la descriptiva o estructural, y explica sucintamente los pasos a seguir para anotar textos en formato impreso. A continuación, describe las relaciones existentes entre los estudios de traducción basados en corpus y la anotación de corpus, que acompaña de una introducción a los estándares de codificación de corpus (*XML Corpus Encoding Standard*) así como a los utilizados en las memorias de traducción (*Translation Memory Exchange*). Una vez planteadas las diferencias, incide en las principales aplicaciones de una metodología basada en corpus en la traducción profesional. Por un lado, propone el uso de corpus compuestos por textos planos, aunque también considera adecuado que los traductores utilicen sistemas comerciales como *Sketch Engine*, una herramienta que permite anotar corpus compuestos por textos descargados de internet. En la segunda parte del capítulo señala que la anotación puede hacerse en tres niveles diferenciados, según interese recoger información documental de los textos en su conjunto, que se correspondería con un primer nivel, de cada texto en sí mismo, es decir, un segundo nivel, o incluir información estructural, por ejemplo datos lingüísticos, en un tercer nivel. Para finalizar el capítulo, Zanettin muestra cómo anotar un corpus en el marco del proyecto *TEI* utilizando un programa de anotación llamado *XAIRA*, mediante el cual se pueden introducir los datos relativos a los tres niveles de información que acabamos de mencionar.

En el quinto capítulo se presentan las principales herramientas disponibles en el mercado para crear recursos basados en corpus, así como para recuperar y analizar los datos de los corpus. Paso a paso, Zanettin describe las diversas etapas que hay que tener en cuenta para compilar un corpus. En primer lugar, es necesario obtener los textos, que pueden estar anotados o no. Seguidamente hay que editarlos en *XML* y, si no lo estuvieran, anotarlos utilizando aplicaciones como *SourceForge*, *CNET.com*, *Oxygen*, *Altova XMLSpy®*, *CLAWS* o *TreeTagger*. Asimismo, señala que es necesario utilizar herramientas que permitan gestionar y buscar en los datos del corpus, popularmente conocidas como analizadores de concordancias. A continuación, examina los principales tipos de análisis asistidos por ordenador que se pueden llevar a cabo utilizando las mencionadas herramientas. Aunque se centra principalmente en el análisis de corpus monolingües, tanto las herramientas como las técnicas pueden resultar también relevantes para la compilación y el análisis de corpus multilingües, ya sean paralelos o comparables. El autor ofrece una pequeña introducción teórica de cada uno de los aspectos que pueden analizarse en un corpus y ofrece ejemplos muy interesantes de trabajos recientes basados en el análisis de listados de palabras, de concordancias, de colocaciones, de *clusters*, de gra-

fos, de coligaciones, del perfil de las unidades léxicas e incluso de asociaciones semánticas, a pesar de que reconoce que el análisis de este último aspecto está influenciado por la interpretación subjetiva del investigador.

El contenido del sexto de los capítulos tiene por objetivo mostrar cómo se compila un corpus multilingüe. Zanettin describe conceptos seminales como «corpus paralelo», «corpus comparable» y «corpus de traducción» basándose en las diversas corrientes de la lingüística de corpus. Considera que dos o más subcorpus en diferentes lenguas pueden tener distintos grados de comparabilidad. Por otro lado, emplea la denominación de corpus paralelo para referirse al conjunto de textos que guardan similitud principalmente en la estructura, es decir, dos o más corpus en lenguas diferentes que pueden alinearse en una serie de unidades de traducción bitextuales. Seguidamente explica con detenimiento las diversas fases para compilar y alinear un corpus paralelo: desde el diseño del corpus, pasando por la adquisición de los textos, el codificado, la anotación y la indización de los mismos. Además, ofrece ejemplos de corpus ya compilados que se pueden emplear como corpus comparables siempre y cuando los usuarios que los vayan a explotar se aseguren de que se han empleado técnicas de compilación similares y de que los criterios de equilibrio y representatividad se cumplan. Asimismo, realiza una breve descripción de los programas disponibles para llevar a cabo el alineado así como de la metodología que subyace a dichos recursos, por ejemplo *WordSmith Tools*, *Vanilla aligner*, *EasyAlign 1.0*, *Champollion Tool Kit (CTK)*, *Geometric Mapping and Alignment (GMA)*, *ParaConc Aligner*, *Alinea*, *Translation Corpus Align (TCA)* e *Interactive Sentence Alignment (ISA)*. Para conseguir una mayor comprensión del proceso ofrece un estudio de caso, el corpus *OPUS*. Para finalizar el capítulo, previamente a describir las actividades y las lecturas complementarias, dedica un apartado a mostrar las diferencias entre los corpus paralelos y las memorias de traducción. En línea con este último apartado, expone que el alineado por debajo del nivel de la oración, es decir, las unidades léxicas y los sintagmas, puede ser de interés para establecer correspondencias bidireccionales entre unidades léxicas tanto en corpus paralelos como comparables.

En el penúltimo de los capítulos examina las herramientas y técnicas empleadas para analizar un corpus multilingüe, con especial atención a los estudios de traducción descriptivos y aplicados. También trata cuestiones como las diferentes funciones de los analizadores de concordancias. Una vez analizadas las herramientas que pueden emplearse para realizar búsquedas en corpus paralelos y recuperar concordancias, presenta dos estudios de caso para ilustrar los

diversos tipos de análisis que pueden llevarse a cabo en corpus paralelos según el nivel de anotación y la herramienta empleada para recuperar y mostrar las concordancias. En primer lugar, los diversos métodos que se pueden adoptar para estudiar las características descriptivas de las traducciones se ilustran a través de ejemplos de un corpus paralelo compuesto de varias novelas de Salman Rushdie y de sus traducciones al italiano utilizando *ParaConc*. Después se muestra un análisis contrastivo de las voces «*eye*» y «*occhio*», para el cual Zanettin se ayuda de la interfaz de búsqueda del *Opus Multilingual Word Alignment Database*. Para finalizar, el autor examina el uso de corpus multilingües como recursos para traductores profesionales y concluye que tanto los corpus comparables como los paralelos pueden ser de gran ayuda para estos usuarios, puesto que les asisten a la hora de resolver problemas durante el transvase interlingüístico a los que probablemente no encuentren respuesta en otras fuentes.

El último de los capítulos está dedicado a las conclusiones, en las que vuelve a incidir en que el objetivo que se persigue en un trabajo de investigación en traducción en el que se aplica una metodología basada en corpus depende principalmente de la disponibilidad de los textos que se incluyen. No obstante, con los avances que se están realizando en este campo, la tecnología lingüística tiene a su disposición herramientas cada vez más complejas que permiten una selección de corpus *ad hoc* así como el *crowdsourcing* entre usuarios, de manera que el repositorio de corpus disponibles en internet para la investigación en los estudios de traducción está creciendo exponencialmente. Por tanto, Zanettin prevé que el desarrollo de aplicaciones de la tecnología informática dedicados a la recuperación de información textual y lingüística en textos electrónicos ofrecerá nuevas vías de investigación aplicables a los estudios de traducción.

En suma, la obra recoge ejemplos y aspectos prácticos para la aplicación de los avances de la lingüística de corpus a los estudios de traducción. Las directrices y los consejos que el autor plantea en su obra resultan de gran interés para aquellos investigadores que quieran experimentar con corpus. La estructura progresiva y práctica propuesta por Zanettin nos plantea una obra interesante, de lectura fácil y de gran funcionalidad para adentrarnos en la práctica de la lingüística de corpus aplicada a los estudios de traducción. Al mismo tiempo, ofrece una variada selección de herramientas útiles para realizar trabajos en esta disciplina.