

Corpus lingüísticos y enseñanza de lenguas

Maribel Serrano Zapata
(Ed.)

Universitat de Lleida
Lleida, 2024

Índice

Introducción	7
<i>Maribel Serrano Zapata</i>	
El potencial inductivo de los corpus lingüísticos en la enseñanza de ELE.....	9
<i>Marta Albelda Marco</i>	
Análisis de errores y etiquetado en los corpus de aprendices de ELE informatizados	31
<i>M^a Ángeles Calero Fernández</i>	
<i>TextAnnot</i> , una herramienta Web para la gestión y anotación de corpus lingüísticos	49
<i>Roberto García González, Albert Berga Gatiús y Juan Manuel Gimeno Illa</i>	
Ausencia de metataxis en la producción escrita de estudiantes chinos de ELE.....	63
<i>Francisco Javier Terrado Pablo, Zhang Peng y Zhang Bohua</i>	
¿ <i>Rabito</i> o <i>rápido</i> ? Análisis de errores en la producción escrita de sinohablantes de ELE.....	77
<i>Demelsa Ortiz Cruz</i>	
Bases para la elaboración de un diccionario para aprendices sinohablantes.....	91
<i>Maria Bargalló Escrivà</i>	
Los índices de disponibilidad léxica como criterio complementario para la selección del leuario de un diccionario de ELE	109
<i>José Antonio Moreno Villanueva y Alicia Madrona Cao</i>	
Léxico disponible y redes semánticas en niños valencianos de 6 años. La construcción del Corpus léxico infantil.....	127
<i>M^a Begoña Gómez Devís</i>	
Creación y análisis de un corpus para el estudio del español (L2) como lengua de instrucción utilizada en los libros de texto.....	139
<i>Ángela Mira Conejero</i>	

Introducción

Maribel Serrano Zapata

Los corpus, conjuntos de datos lingüísticos recopilados para conocer mejor una lengua, cuentan con una larga tradición en los estudios filológicos. Hoy en día, están más vigentes que nunca gracias a las posibilidades que nos confieren las herramientas informáticas. En este libro se pueden encontrar ejemplos diversos de aplicaciones didácticas basadas en corpus lingüísticos. Dado el interés creciente por el aprendizaje del español en China, varios capítulos están dedicados a esta cuestión.

El volumen se abre con una contribución de Marta Albelda sobre las posibilidades que ofrecen los corpus lingüísticos en la didáctica de la lengua española, especialmente desde una perspectiva inductiva. Se muestra y se ilustra con diferentes actividades cómo es posible trabajar con corpus de concordancias, como los corpus de referencia de la RAE. Además, se incide en las ventajas de los corpus de nativos con recuperación completa del texto. En concreto, los corpus de conversaciones orales espontáneas ofrecen opciones interesantes para la enseñanza de cuestiones discursivas y pragmáticas, ya que se basan en muestras de habla contextualizadas.

A continuación, varios capítulos se enmarcan en el proyecto *Elaboración y catalogación de un corpus de textos escritos en ELE producidos por estudiantes sinohablantes*, cuyo objetivo principal fue la construcción de un corpus de aprendices sinohablantes de español. Este proyecto, desarrollado en la Universitat de Lleida durante el trienio 2017-2019, sirvió para diseñar y construir el CINEAS¹. M^a Ángeles Calero, investigadora principal de este proyecto, reflexiona en su contribución sobre las distintas taxonomías empleadas para detectar errores en corpus informatizados y se plantea si es posible establecer una clasificación general del error que pueda aplicarse a cualquier corpus de aprendices digitalizado y cómo hacer frente a las dificultades que conlleva este reto.

En el trabajo conjunto de Roberto García, Albert Berga y Juan Manuel Gimeno se describe el funcionamiento de la herramienta Web que han desarrollado, llamada *TextAnnot*. Esta herramienta informática, disponible en línea, permite gestionar, anotar y etiquetar corpus. A lo largo de este capítulo, además, se explica cómo las anotaciones y etiquetas de errores definidas para el CINEAS pueden reutilizarse y adaptarse a otros corpus lingüísticos.

1. Corpus de INterlengua Española de Aprendices Sinohablantes (<https://cineas.udl.cat>).

El trabajo de los profesores Javier Terrado, Zhang Peng y Zhang Bohua trata de los problemas sintácticos de los estudiantes chinos a la hora de aprender español. Basándose en ejemplos extraídos del CINEAS, los autores demuestran la utilidad de la noción de metataxis de Lucien Tesnière para aludir a la transferencia de conceptos de una lengua a otra, utilizando la expresión más adecuada y teniendo en cuenta las diferencias lingüísticas y las culturales.

Demelsa Ortiz, por su parte, analiza los errores presentes en 40 producciones escritas de estudiantes sinohablantes de español que se encuentran en la Universidad de Zaragoza, por tanto, en un contexto de inmersión. La investigadora lleva a cabo un análisis cuantitativo y cualitativo de los errores hallados en cada uno de los niveles lingüísticos (fonético-ortográficos, morfosintácticos y léxico-semántico). También compara los errores obtenidos según si la tarea de expresión escrita ha sido controlada o libre para acabar concluyendo que apenas se da variación en función del tipo de tarea y que los errores morfosintácticos son los más numerosos.

Maria Bargalló describe las primeras fases del proyecto *Diccionario de español para aprendices sinohablantes* llevado a cabo en la Universitat Rovira i Virgili, cuyo objetivo principal es la elaboración de un diccionario electrónico de nueva planta. En esta primera fase se analizan las necesidades lexicográficas de los aprendices sinohablantes de español de nivel intermedio y se describen los criterios que se siguen en la selección del leuario. La contribución de José Antonio Moreno y Alicia Madrona se enmarca también dentro de este proyecto. En ella, sus autores comparan la disponibilidad léxica de dos grupos de estudiantes (nativos y sinohablantes) en torno a tres centros de interés con la finalidad de establecer qué vocablos podrían figurar en el *Diccionario de español para aprendices sinohablantes* que está elaborando este equipo. Además de los datos que se obtienen a partir de la disponibilidad, también se toman en consideración los índices de frecuencias y la nivelación que se propone en *Plan curricular del Instituto Cervantes*.

Begoña Gómez-Devís aborda la construcción de un corpus de disponibilidad léxica infantil. Tras un repaso de los trabajos dedicados a la aplicación del test de léxico disponible a escolares de 6 años, la autora indaga en las posibilidades psicolingüísticas de este tipo de pruebas, concretamente, a partir de los grafos obtenidos con el programa DispoGrafo en el centro de interés *Colores*.

El volumen se cierra con la contribución de Ángela Mira, quien analiza un corpus formado por libros de texto de enseñanza secundaria utilizados en centros escolares de la provincia de Alicante con una alta proporción de alumnado inmigrante. En esta fase inicial de su estudio, la autora describe los aspectos más relevantes de la lengua empleada en dichos libros, con el fin de identificar las características de la lengua de instrucción, cuyo dominio es fundamental para alcanzar el éxito académico.

En resumen, este libro contribuye a dar una visión panorámica de las distintas investigaciones que pueden realizarse a partir de corpus lingüísticos, lo que demuestra su utilidad y vigencia en el entorno actual y, en específico, en su vertiente didáctica aplicada a la enseñanza de lenguas.