



Reseñas / *Book reviews*

Corpus lingüísticos y enseñanza de lenguas (Maribel Serrano, ed.), Lleida: Editorial Universitat de Lleida, 2024, 150 págs. ISBN: 978-84-9144-469-5. PRECIO 18 €. Reseñado por M^a Teresa Miramón Escribano, *Universidad Rey Juan Carlos*. <https://orcid.org/0000-0001-7412-7280>

Reseña recibida el / *Review received*: 2024-06-10

Reseña aceptada el / *Review accepted*: 2025-01-23

Corpus lingüísticos y enseñanza de lenguas aúna una colección de artículos en torno al tema de los corpus en la lingüística española y cómo estos conjuntos de textos pueden ayudar al lingüista a encontrar datos empíricos y pertinentes con los que identificar distintos fenómenos lingüísticos objeto de estudio. Dada la necesidad, en el campo de la enseñanza del español como lengua extranjera (ELE) de investigar distintos aspectos de la interlengua de los estudiantes, en este volumen se pueden encontrar varios artículos enfocados a la descripción de corpus lingüísticos de aprendices de ELE, especialmente dirigidos a identificar características del habla de estudiantes sinohablantes.

La enseñanza de ELE en China ha observado un crecimiento exponencial en los últimos años, y hay una necesidad de contar, en primer lugar, con estudios rigurosos basados en datos y, en segundo lugar, con materiales de enseñanza contruidos a partir de dichos estudios. Por eso, el libro presenta un tema actual y de especial interés. Este

monográfico cuenta con una introducción general y nueve capítulos. Cada uno de ellos se centra en un aspecto diferente del campo de estudios de los corpus lingüísticos.

La primera contribución que se muestra es de Marta Albelda, perteneciente al Grupo de investigación Val.Es.Co. (Universitat de València), y trata el tema de la explotación de los corpus lingüísticos y discursivos para perseguir objetivos didácticos en la enseñanza de lenguas extranjeras. La autora propone dos caminos para explotar la utilidad de estos conjuntos de textos. Por un lado, utilizando el método *data-driven learning* (Johns, 1991), que se nutre de corpus de concordancias para enseñar aspectos gramaticales y léxicos y, por otro lado, usando corpus de acceso directo al texto para mostrar cuestiones pragmático-discursivas. Albelda señala la falta de modelos que abran camino para desarrollar actividades que permitan una explotación docente para la enseñanza y aprendizaje de lenguas extranjeras, y su trabajo ofrece claves para ejecutar ejercicios que permitan el

aprendizaje inductivo a partir de la exposición a datos. La autora explica el potencial del aprendizaje inductivo con el método *data-driven learning* y presenta ejemplos de actividades usando corpus de concordancia, ofreciendo al lector orientación sobre este asunto. Por último, también señala la importancia en la enseñanza de ELE de los corpus de nativos para desarrollar cuestiones discursivas y pragmáticas. Para ello, pone el foco en la explotación de corpus de conversaciones espontáneas debido a su importancia contextual y por ser expresiones naturales, primarias y prototípicas de la comunicación humana.

Después de esta aportación, los siguientes tres capítulos están enmarcados en el proyecto *Elaboración y catalogación de un corpus de textos escritos en ELE producidos por estudiantes sinohablantes*, desarrollado por el grupo LECCEL (Lenguas en Contacto y Enseñanza de Lenguas). El fruto de este proyecto fue la creación de un corpus lingüístico textual digitalizado, de amplias dimensiones y representativo, especializado en aprendices de ELE sinohablantes. Este conjunto de textos se materializó en la creación del CINEAS¹, y M^a Ángeles Calero, investigadora principal del proyecto, abre esta serie de contribuciones que se sitúan en torno al citado corpus.

En este capítulo, Calero hace una revisión sobre distintas taxonomías de errores existentes en el contexto del aprendizaje de ELE en lingüística de corpus, para identificar criterios básicos y comunes que abarquen las dificultades esenciales y que permitan establecer una clasificación generalizable.

Además de esto, repasa las diferentes aplicaciones informáticas existentes para la anotación de errores en

corpus lingüísticos de aprendices de ELE, y describe la herramienta *TextAnnot*, creada sobre la base del CINEAS y a disposición de la comunidad científica para anotar corpus lingüísticos informatizados. Calero hace hincapié en la importancia de los programas informáticos de etiquetado de errores para avanzar “1) en la descripción de la interlengua de estudiantes de ELE, 2) en la identificación de las áreas de dificultad [...] y 3) en el desarrollo de estrategias y materiales [...]” (p. 44).

El tercer capítulo pertenece a Roberto García González, Albert Berga Gatus y Juan Manuel Gimeno Illa, y en él se profundiza sobre la mencionada herramienta *Web TextAnnot*. Como señalan los autores, este programa permite la posibilidad de adaptar la anotación dependiendo de las necesidades de estudio, tanto de manera individual como de forma colaborativa, y posibilita la recuperación de la información anotada para establecer investigaciones. Además, cuenta con un diseño planteado para que pueda adaptarse a otros proyectos de creación de corpus lingüísticos. Los autores explican la metodología seguida para desarrollar la herramienta, y facilitan las claves de su implementación, definiendo los perfiles de usuarios, los campos de anotación de muestras, las cargas de ficheros de muestras con anotaciones, la configuración de etiquetado, y el etiquetado y uso del corpus. Por último, recuerdan que *TextAnnot* se ha probado exitosamente en el CINEAS y ya se puede consultar en línea.

La cuarta y última contribución en torno al CINEAS tiene como autores a F. Javier Terrado, Zhang Peng y Zhang Bohua, y toma ejemplos del citado corpus para representar el proceso

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

metatáctico de Tesnière (1959) para realizar transformaciones de estructuras lingüísticas de la L1 a la L2. Esta transformación se denomina *metataxis*: “el cambio estructural al que el traductor debe someter una frase en el proceso de versión a otra lengua” (p.65). Los autores señalan la importancia de que los estudiantes, especialmente al comienzo de su aprendizaje, lleven a cabo procesos de metataxis para evitar verter las estructuras sintácticas de su lengua materna al español. Se hace hincapié en la necesidad debido a que el chino mandarín y el español son lenguas de tipología muy alejada. Para ilustrar esto, se utilizan ejemplos del CINEAS teniendo en cuenta las diferencias lingüísticas y culturales de ambas lenguas.

Por su parte, Demelsa Ortiz continúa en su contribución con el tema de la enseñanza de ELE a estudiantes sinohablantes, tomando como punto de partida un conjunto de cuarenta textos producidos por este tipo de aprendices, con actividades controlada (en una primera fase) y tareas libres (en una segunda fase), de nivel B1 y en contexto de inmersión lingüística. Ortiz presenta un análisis de errores de 9 estudiantes sinohablantes que formaron parte del estudio, dividido en las dos fases mencionadas. Para ello, clasifica los errores en tres niveles: fonético-ortográfico, morfosintáctico y léxico-semántico. Entre las conclusiones a las que llega, plantea que los errores en actividades controladas, en general, son más cuantiosos que en ejercicios libres, y hace hincapié en la necesidad de la continua investigación para la mejora de la enseñanza-aprendizaje de estudiantes sinohablantes de ELE.

Los siguientes dos capítulos se enmarcan bajo el paraguas del proyecto *Diccionario de español para aprendices*

sinohablantes (DEAS), de la Universitat Rovira i Virgili.

María Bargalló explica en su capítulo las primeras etapas para la confección del diccionario electrónico, dirigido a usuarios de nivel B1 y B2 de referencia según el MCER. Bargalló repasa los diccionarios existentes, bilingües o monolingües, enfocados a ELE, y establece tres objetivos para el proyecto: “1) Analizar las necesidades [...] 2) Determinar el leuario del diccionario [...]; 3) Establecer los ítems para el diseño de la base de datos lexicográfica” (p. 93). Para desarrollar el primer objetivo, usan cuestionarios dirigidos a estudiantes sinohablantes. Para determinar el leuario, acuden al análisis de manuales, la disponibilidad léxica, una selección de unidades pluriverbales (recopilaciones y materiales publicados con selecciones ya realizadas) y a datos culturales (análisis en diccionarios para aprendices de ELE y de manuales, entre otros). Por último, para la construcción de la base de datos lexicográfica, se tiene en cuenta una serie de ítems propuestos por Bergenholtz y Nielsen (2014) de 39 categorías, una base de datos bilingüe (español-chino) y una serie de ilustraciones.

Por su parte, José Antonio Moreno y Alicia Madrona profundizan sobre el tema de la disponibilidad léxica como criterio para la selección del leuario del *Diccionario*. Describen en su contribución el léxico disponible en tres centros de interés: *Educación, Política y Justicia*, y establecen comparaciones entre dos grupos (25 estudiantes nativos y 25 estudiantes chinos). La identificación de los términos de mayor interés permite una selección de léxico que a su vez sirve para establecer correlaciones con el léxico que muestran los inventarios de *Nociones específicas* del PCIC (Instituto Cervantes, 2006) y

los índices de frecuencias del Corpes XXI-Enclave RAE y Davies y Davies (2018). Los autores señalan que cuanto más específico es el centro de interés, más se acerca la media del léxico disponible entre ambos grupos: el centro de interés *Educación* presenta la mayor cifra de voces distintas actualizadas por parte de los estudiantes sinohablantes, alejada de *Política y Justicia*. Éste último centro es el que presenta mayor déficit léxico.

Begoña Gómez-Devís aborda en su capítulo el tema de la construcción de un corpus de léxico disponible de niños de 6 años (al inicio de la educación obligatoria). Para ello, utiliza el marco de la disponibilidad léxica infantil, tomando como punto de referencia la ciudad y área de Valencia y enfocando su centro de interés en el campo *Colores*. Para el análisis de la vertiente cognitiva del centro de interés y la capacidad asociativa de las palabras, utiliza la herramienta informatizada DispoGrafo, diseñada por el equipo de la Universidad de Concepción (Chile), que genera automáticamente grafos “cuyos nodos representan palabras y cuyas aristas simbolizan las relaciones entre ellas” (p. 129), así se pueden establecer redes semánticas estables. A lo largo de su contribución, Gómez-Devís muestra nuevas perspectivas con posibilidad de aplicación en la enseñanza.

Por último, el libro se cierra con la contribución de Ángela Mira, que se centra en estudiantes jóvenes

inmigrantes en edad de escolarización, cuya lengua materna no es el español. El movimiento migratorio supone un reto educativo para los docentes y la Administración, y en las últimas décadas se han establecido una serie de medidas para regular esta situación, pero la realidad del aula de acogida y la atención a la diversidad lingüística y cultural presentan unos obstáculos que generan cierta problemática, como señala la autora. Para ello, Mira se propone crear un corpus lingüístico formado por libros de texto para analizar los aspectos más relevantes de la lengua utilizada en ellos, con el objetivo de identificar las características básicas de la lengua de instrucción asociada con cada materia. De esta manera, se puede analizar la adecuación de este tipo de lenguaje a la realidad de los aprendices.

En conclusión, en este volumen temático se pueden encontrar valiosas aportaciones teóricas, avaladas por una bibliografía actualizada, además de prácticos análisis de fenómenos lingüísticos y ejemplos didácticos. Constituye una referencia indispensable para investigadores interesados/as en la enseñanza de lenguas, especialmente enfocados/as en el estudio de la interlengua de estudiantes sinohablantes, al igual que para aquellos/as que quieran profundizar en herramientas Web de utilidad docente e investigadora, como diccionarios, corpus lingüísticos, o herramientas de anotación y etiquetado de corpus.

REFERENCIAS BIBLIOGRÁFICAS

- Bergenholtz, Henning y Nielsen, Sandro (2014). What is a Lexicographical Database?, *Lexicos*, 23, 77-87.
- Davies, Mark y Davies, Kathy Hayward (2018). *A frequency dictionary of Spanish. Core vocabulary for learners*. London y New York: Routledge.

- Instituto Cervantes (2006). *Plan curricular del Instituto Cervantes. Niveles de referencia del español*. Madrid: Instituto Cervantes. https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/
- Johns, Tim (1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. En Tim John y Philip King (Eds.), *Classroom Concordancing. English Language Research Journal*, 4, 27-45.
- Tesnière, Lucien (1959). *Éléments de syntaxe structural*. París: Klincksieck [1994. *Elementos de sintaxis estructural*. Madrid: Gredos. Traducción de Esther Diamante].