



Serrano Zapata, M. (Ed.). (2024). *Corpus lingüísticos y enseñanza de lenguas*. Ediciones de la Universidad de Lleida, 151 págs. ISBN: 978-84-9144-469-5

Recibida el 22/10/2024

Aceptada el 04/05/2025

El volumen *Corpus lingüísticos y enseñanza de lenguas*, editado por Maribel Serrano y publicado por Ediciones de la Universidad de Lleida en 2024, constituye una valiosa aportación al campo de la lingüística aplicada. Esta obra colectiva reúne nueve estudios elaborados por especialistas de universidades nacionales e internacionales con el propósito común de profundizar en el potencial didáctico de los corpus lingüísticos en contextos educativos de ELE. Dirigida principalmente a investigadores, docentes de lenguas extranjeras, formadores de profesores y desarrolladores de materiales didácticos, la obra ofrece una panorámica amplia y actualizada de metodologías, recursos tecnológicos y propuestas pedagógicas relacionadas con el uso de corpus en el aula. La presente publicación se vincula a proyectos de investigación recientes y aparece en un momento de creciente interés por las tecnologías del lenguaje y la enseñanza basada en datos reales, lo que la convierte en una herramienta útil tanto para la investigación universitaria como para el desarrollo profesional docente.

En el primer capítulo, «El potencial inductivo de los corpus lingüísticos en la enseñanza de ELE», Marta Albelda Marco (Universidad de Valencia) presenta estos corpus como herramientas útiles en la enseñanza de lenguas debido a que permiten al alumnado familiarizarse con el idioma a través del enfoque inductivo. Distingue entre dos tipos de corpus cuya selección depende del aspecto lingüístico en que quiera centrarse el docente: el uso de corpus de concordancias para el aprendizaje de aspectos léxicos y gramaticales mediante el método *data-driven learning*, y el uso de corpus de acceso directo al texto aplicados a la enseñanza de cuestiones discursivas y pragmáticas. En ambos casos, se recurre a actividades inductivas que guían al alumnado en la deducción de las reglas pragmáticas y gramaticales de la lengua para elaborar hipótesis a partir de la observación. En el primer método, el objetivo es que el aprendiz identifique patrones sintácticos, morfológicos y léxicos presentes en los corpus de concordancias, y que comprenda el funcionamiento de las estructuras observadas. En cuanto al segundo, en los corpus discursivos, el alumno parte del contexto para identificar la conexión entre funciones pragmáticas e interpretar aspectos comunicativos, como valores, usos, intenciones, etc.

Las tres siguientes contribuciones se engloban en el proyecto *Elaboración y catalogación de un corpus de textos escritos en ELE producidos por estudiantes sinohablantes*,

del grupo Lenguas en Contacto y Enseñanza de Lenguas (LECEL) de la Universitat de Lleida. Como resultado de este proyecto, liderado por M^a Ángeles Calero, se creó el corpus lingüístico CINEAS, que recoge producciones textuales de aprendices sinohablantes de ELE.

En este capítulo, Calero aborda la relevancia de la informatización de los corpus de aprendices y de los sistemas de etiquetado, y «plantea si es posible determinar una taxonomía del error de la que pueda derivarse un sistema de etiquetado estandarizado semiautomático aplicable a corpus de aprendices digitalizados» (p. 31). Ante la amplia variedad de clasificaciones existentes, se concluye que todo sistema de distribución de errores en categorías es inexacto, ya que ha sido ideado para una investigación específica. Por lo tanto, ninguna tipología del error puede generalizarse de manera que se amolde a cualquier corpus. En cuanto al proceso de informatización, se establece que la creación de corpus de aprendices digitalizados y de programas informáticos para el etiquetado de errores permite sacar el máximo provecho de la explotación del corpus en las investigaciones de análisis de errores. Del mismo modo, facilita el estudio de la interlengua de los aprendices, la detección de las dificultades de los alumnos y la creación de métodos para sortearlas.

En «*Text Annot*, una herramienta web para la gestión y anotación de corpus lingüísticos», Roberto García González, Albert Berga Gatiús y Juan Manuel Gimeno Illa profundizan en la herramienta web de etiquetado colaborativo *Text Annot*, que diseñaron con el objetivo de facilitar la gestión y anotación del Corpus CINEAS. En este capítulo, los autores abordan la metodología seguida en su creación y explican el proceso de implementación de esta, centrándose en la gestión de los perfiles de usuario, la definición de los campos de anotación de muestras, la carga de ficheros de muestras con anotaciones, la configuración del etiquetado y la utilización del mismo. Además, describen las posibilidades de explotación del corpus que este sistema ofrece al usuario final; entre ellas, la visualización de las estadísticas globales o la de las especializadas por campos de anotación, junto a la búsqueda por palabras, por anotaciones o por etiquetas. Al concluir, los investigadores hacen hincapié en que el diseño se encuentra «enfocado en la flexibilidad de la herramienta resultante, que la hace adaptable a diferentes tipos de muestras, así como a múltiples tareas de análisis y anotación de textos» (p. 61). Dicha flexibilidad y su disponibilidad gratuita en línea incentivan su uso en otros proyectos de corpus lingüísticos.

Francisco Javier Terrado Pablo, Zhang Peng y Zhang Bohua presentan la última contribución relacionada con el proyecto del grupo LECEL. En este capítulo, los autores estudian la ausencia de metataxis en la producción escrita de los estudiantes sinohablantes de ELE, que se produce por las amplias diferencias existentes entre los patrones lingüísticos del chino, que es una lengua paratáctica, y los del español, que es hipotáctica. Para salvar esta distancia, al traducir, el aprendiz sinohablante de ELE debe recurrir a la metataxis, renunciando a la estructura de la frase en la lengua de origen para verter la idea que se quiere expresar en los moldes adecuados en la lengua de destino. Por lo tanto, resulta indispensable llevar a cabo una transformación estructural que tenga en cuenta las diferencias lingüísticas y culturales entre los dos idiomas, alejándose así de la

clonación directa de la estructura de partida. De hecho, recomiendan que, para evitar los errores relacionados con la ausencia de metataxis lineal y estructural, el docente preste especial atención a la tendencia de los aprendices a establecer una correspondencia entre la forma externa de las palabras de ambas lenguas, y ponga más énfasis en las diferencias lingüísticas y culturales entre el chino y el español, además de explicar en el aula los errores que se cometen con mayor frecuencia.

Por otro lado, «*¿Rabito o rápido? Análisis de errores en la producción escrita de sinohablantes de ELE*», de Demelsa Ortiz Cruz, se centra en el estudio de los errores cometidos por aprendices de nivel B1 en un contexto de inmersión lingüística. El objetivo es identificar los errores y analizarlos cualitativa y cuantitativamente partiendo de una tarea controlada, que los alumnos completan sin ayuda externa y con limitación temporal bajo la supervisión del docente, y de otra libre, bajo las condiciones opuestas. Al finalizar dicho análisis, Ortiz concluye que son más numerosos los errores en los ejercicios controlados que en los libres. Además, constata que los errores cometidos en las actividades vigiladas aparecen en todos los niveles de análisis (fonético-ortográfico, morfosintáctico, léxico-semántico), aunque no con la misma recurrencia: casi dos tercios son morfosintácticos, un tercio es de tipo fonético-fonológico y el resto de equivocaciones son léxico-semánticas. Por último, en la misma línea, deja constancia de que los errores en ejercicios libres también afectan a todos los niveles de análisis lingüístico.

A continuación, la obra sigue con dos artículos relacionados con el proyecto *Diccionario de español para aprendices sinohablantes (DEAS)* de la Universitat Rovira i Virgili. El primero de ellos, escrito por Maria Bargalló Escrivà, plantea los pasos a seguir para confeccionar un diccionario para aprendices sinohablantes de ELE de los niveles B1 y B2 del MCER. En primer lugar, es necesario estudiar el perfil del usuario mediante cuestionarios para así analizar las necesidades de información de estos estudiantes en el proceso de codificación y decodificación de la lengua. Seguidamente, debe seleccionarse un leuario que responda a estas necesidades recurriendo al análisis de manuales y al estudio de la disponibilidad léxica de los aprendices sinohablantes, prestando especial atención a la selección de unidades pluriverbales y de datos culturales. En el último paso, deberán elegirse los ítems por incluir en el diseño de la base de datos lexicográfica español-chino. En el *DEAS*, estos quedan distribuidos en un total de 39 categorías, que incluyen definiciones, información lingüística (gramatical, semántica, combinatoria, diafásica y pragmática), ejemplos, frecuencias, ilustraciones, notas explicativas, traducciones al chino o aclaraciones sobre posibles errores, entre otras.

El séptimo capítulo, titulado «Los índices de disponibilidad léxica como criterio complementario para la selección del leuario de un diccionario de ELE», ahonda en los criterios de selección de los lemas en la creación de un diccionario de ELE dirigido a estudiantes sinohablantes. Para ello, Moreno Villanueva y Madrona Cao analizan la disponibilidad léxica de tres centros de interés (*Educación, Política y Justicia*) comparando los resultados de un grupo de alumnos nativos con otro formado por estudiantes chinos. En todos los casos, se observa que la media de unidades léxicas de los estudiantes del grupo nativo es superior a la de los estudiantes chinos y que esa media disminuye si el centro de interés presenta mayor especificidad o se aleja del ámbito personal. En cuanto al

número de voces distintas empleadas por los alumnos sinohablantes, se observa que el centro *Educación* supera con creces a los otros dos, siendo *Justicia* el que evidencia un claro déficit respecto al grupo de nativos. Por otra parte, los resultados obtenidos se correlacionan con varios inventarios e índices de frecuencias, entre los que se encuentran el de Nociones específicas del *Plan Curricular del Instituto Cervantes* (2006), el Davies y Davies (2018), y el Corpes XXI de la RAE. De esta forma, se pretende delimitar el vocabulario fundamental a través de la intersección entre el léxico disponible y el frecuente.

En la penúltima aportación de la obra, Gómez Devis aborda la construcción de un corpus de léxico disponible en alumnos valencianos de educación infantil de 6 años de edad. Para ello, se propone descubrir cuáles son los términos de los que disponen en el centro de interés *Colores*, con el objetivo de ampliar dicho léxico y establecer criterios en su selección para cada nivel curricular. Al presentar los resultados a través de DispoGrafo, se observa la representación de las redes semánticas mentales de los sujetos de estudio, en las que las palabras con mayor disponibilidad léxica se encuentran en el centro. En cambio, las principales asociaciones semánticas, que vinculan cada nueva palabra con un término previo, se encuentran en la periferia. Mediante el análisis de los resultados representados, el estudio de la vertiente cuantitativa permite saber qué vocabulario han incorporado en etapas tempranas, mientras que el de la vertiente cualitativa posibilita estudiar qué mecanismos de la memoria infantil se usan para afianzar el léxico y recuperar el ya adquirido previamente.

El volumen finaliza con el artículo de Mira Conejero sobre la creación y el análisis de corpus para el estudio del español (L2) como lengua de instrucción. Partiendo de la necesidad de facilitar la integración del colectivo de inmigrantes en edad de escolarización cuya lengua materna no es el español, la autora alerta sobre la inexistencia de medidas educativas unánimes al respecto que ofrezcan herramientas a los discentes para afrontar estas dificultades. En consecuencia, el objetivo del estudio es crear un corpus lingüístico de libros de texto para identificar los rasgos más característicos del español académico usado en estos manuales y así poder diseñar recursos que respondan a las necesidades de los estudiantes inmigrantes.

En definitiva, *Corpus lingüísticos y enseñanza de lenguas* constituye una publicación colectiva con una sólida coherencia interna que articula de forma equilibrada perspectivas teóricas y aplicadas sobre el uso de corpus en la enseñanza del español como lengua extranjera. A través de los nueve capítulos que lo componen, el volumen ofrece una visión multidisciplinar en la que convergen investigaciones sobre construcción de corpus, análisis de errores, herramientas digitales para su explotación, propuestas lexicográficas y estudios de disponibilidad léxica. Esta pluralidad se mantiene vertebrada por un eje común: la voluntad de demostrar la vigencia de los corpus lingüísticos como herramientas de análisis y de intervención didáctica.

Sobre este trasfondo, uno de los principales logros de la obra es su capacidad para enlazar las necesidades del aula de ELE con desarrollos técnicos y metodológicos provenientes de la lingüística de corpus, ofreciendo ejemplos concretos de proyectos que responden directamente a los desafíos actuales en el ámbito de la enseñanza de lenguas. La combinación de estudios empíricos con propuestas de implementación práctica confie-

re solidez y aplicabilidad a las reflexiones presentadas, reforzando así la utilidad del volumen tanto en el plano académico como en el pedagógico.

Así pues, por su rigor académico, su aplicabilidad pedagógica y su atención a contextos de enseñanza plurales, este volumen merece una lectura atenta por parte de quienes investigan, enseñan o diseñan materiales para el aprendizaje de lenguas en el siglo XXI.

Referencias bibliográficas

CORPES XXI-ENCLAVE RAE. *Corpus avanzado*. Real Academia Española. <https://enclave.rae.es/recursos/corpus>

DAVIES, Mark; 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. http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/

JÚLIA MACHETTI VALLVERDÚ

Universidad de Salamanca, España

machetti.julia@usal.es