

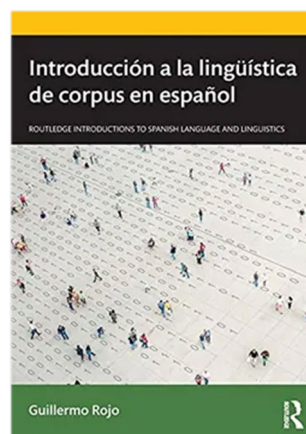
*Introducción a la lingüística de corpus en español* cumple lo que promete su título. Es una oportunidad de considerar la lingüística de corpus (LC) desde la perspectiva del español, tanto por los corpus del español que se discuten y se analizan como por el hecho de que el texto está escrito en español y provee innumerables ejemplos para realizar búsquedas en corpus.

Este monográfico está pensado desde una perspectiva didáctica para estudiantes de cursos de lingüística hispánica que estén finalizando un grado, un máster o un doctorado, y para aquellos que con mínimos conocimientos de lingüística quieran ampliar sus conocimientos de la lingüística de corpus. Con esta finalidad didáctica, cada capítulo incluye una sección con “cuestiones, problemas y temas de investigación” que permite revisar y practicar los temas principales del capítulo. Las actividades de práctica permiten explorar corpus de español disponibles *online*, tales como CREA, CORDE, CORPES XXI, Corpus del Español (Web/Dialectos) (*CdEdw*), Corpus del Español (Género/Histórico) (*CdEhist*) o ESLORA. Además, en los capítulos se dan ejemplos de una variedad de corpus textuales y otros recursos electrónicos que se compilan en un apéndice final. Solo por este apéndice ya merece la pena tener el libro a mano.

El primer capítulo da una visión general del concepto de corpus, las áreas de la lingüística para las que un corpus puede ser útil y los tipos de corpus que existen. En concreto, describe cómo los corpus sirven para análisis léxicos, tales como examinar la frecuencia de palabras o estudios de clases de palabras, por ejemplo, cómo investigar la forma *la* como artículo o pronombre personal. Explora cómo los corpus enriquecen investigaciones sobre aspectos semánticos (cómo la evolución en el significado de *enervar*) y cuestiones diacrónicas<sup>1</sup>. Además, examina cómo las investigaciones sobre aspectos sociolingüísticos y enseñanza de lenguas se beneficiarán del uso de corpus. Con estos objetivos en mente, revisa diferentes tipos de corpus: corpus técnicos, corpus de aprendices y corpus de entrenamiento. Esta introducción enfatiza las cuestiones y temas que se investigarán en profundidad en los capítulos siguientes.

En el segundo capítulo se resume el rol que cumple la LC, prestando especial atención al contexto histórico en el que ha aparecido. El capítulo empieza con una descripción de la lingüística en términos generales, mostrando su posición única como ciencia empírica y a la vez cultural. Se describe la tensión que se ha sentido en la disciplina durante el proceso de evolución desde los años sesenta: el surgimiento de la LC tiene que ver con este cambio de la lingüística y la ciencia en general. De esta manera, se enfatiza que la LC se puede considerar una revolución instrumental enormemente influyente en el campo, en cuanto a que no es necesariamente una nueva metodología en sí, pero es una herramienta que permite a los lingüistas seguir sus investigaciones empíricas de manera más eficaz y esclarecedora. Su desarrollo ha tenido un efecto profundo en la disciplina porque, aunque regresa al tipo de datos que se usaban en la lingüística

<sup>1</sup> En lingüística, las cuestiones diacrónicas examinan la evolución del idioma a través del tiempo. Rojo da el ejemplo de cómo un corpus compara las frecuencias normalizadas de *vegetal* y *vegetable* a través de los siglos (p. 18).



tradicional (o, más bien, la lingüística descriptiva antes del cambio en los años sesenta), emplea el modo de análisis que viene del método hipotético-deductivo.

En el tercer capítulo se define qué es un corpus. Aunque los corpus son conjuntos de objetos reunidos con el propósito de facilitar su uso y análisis, este capítulo deja claro que la obra se centrará en los corpus lingüísticos. Se describe en detalle el proceso de construcción, diseño, explotación y uso de los corpus lingüísticos. En cuanto al diseño, se presentan los diferentes elementos que uno debe considerar para crear corpus lingüísticos de una manera efectiva. También se compilan recomendaciones para la construcción de corpus lingüísticos. Por último, se resumen las diferentes maneras en las que los lectores y lingüistas pueden aprovechar al máximo los corpus lingüísticos gracias a su codificación y anotación.

En el cuarto capítulo se exploran las posibilidades de la LC para analizar el léxico del español. El capítulo se enfoca principalmente en dos secciones: en frecuencia de elementos y fenómenos léxicos y en variación del léxico. La primera parte empieza con formas ortográficas y la ley de Zipf<sup>2</sup>. La parte final de esta primera sección habla sobre las unidades multipalabras que son expresiones que consisten en más de una palabra con un significado común, por ejemplo, *por favor*. La segunda mitad del capítulo se enfoca en los cuatro ejes de la variación lingüística: el eje diatópico, el diacrónico, el diastrático y el diafásico. Todos estos ejes permiten aproximarnos a diferentes aspectos del léxico, por ejemplo, su formalidad o su uso, que la LC nos facilita explorar.

El quinto capítulo se organiza de manera paralela al cuarto capítulo, pero, en vez de tratar del léxico, analiza una amplia variedad de fenómenos relacionados con la gramática. Este capítulo se centra en cómo recuperar las frecuencias de uso de distintos elementos gramaticales (ej. verbos acabados en *-ar*, perífrasis verbales o adverbios con *-mente*) con la ayuda de corpus. También explora detalles sobre los cambios y diferencias gramaticales entre los países hispanohablantes y la variación diacrónica.

En el sexto capítulo se proporcionan antecedentes sobre la LC. Este capítulo enfatiza la importancia del empleo de las computadoras para su desarrollo, ya que ha facilitado el proceso de informatización de los corpus. Esta informatización hace que el proceso de extraer y analizar la información lingüística de los textos sea mucho más fácil, y se ha mejorado a medida que la capacidad y velocidad de procesamiento de las computadoras han avanzado con el tiempo. Además, se describen las ventajas y desventajas de los corpus textuales actuales antes de ofrecer varios datos estadísticos del CORPES XXI. El capítulo termina hablando de la evolución de la LC, específicamente sobre la necesidad de los corpus de referencia y la necesidad de una mejor anotación y codificación de los textos en los corpus. El tono del capítulo nos hace mostrarnos positivos con los futuros avances que aún le quedan por experimentar a la LC.

<sup>2</sup> “Tendencia según la cual las frecuencias de los elementos de un corpus mantienen una relación constante, de modo que la frecuencia del segundo equivale a la del primero partido por dos, la del tercero a la del primero partido por tres, etc. Como la ley de Pareto, refleja la tendencia a que unos pocos elementos concentren la mayor parte de los casos” (Rojo, 2021, p. 358).



En el séptimo y último capítulo se mencionan diferentes aplicaciones que los investigadores pueden utilizar para explorar corpus más allá de los corpus disponibles *online*, como CORPES XXI y Corpus del Español. Aparte de revisar expresiones regulares, se enumeran diferentes utilidades del mundo Unix que también se pueden usar en MacOS y en Windows. Se nos explica con detalle las diferentes codificaciones que deben incorporar en los ficheros para producir datos de frecuencia de los textos, por ejemplo, el número de palabras en un texto y la frecuencia de las mismas. Este capítulo ofrece diferentes ejemplos de comandos y sus resultados, que servirán a aquellos investigadores que quieran analizar sus propios corpus o corpus digitalizados que carezcan de un interfaz.

En resumen, este libro cumple los objetivos que nos presenta en la introducción. Al acabar el libro, el lector conoce una plétora de corpus disponibles para explorar el español a la vez que ha aprendido nociones básicas sobre cómo utilizar algunos de ellos y qué aspectos se pueden explorar. Sin duda nos encontramos en un buen momento para la LC en español, pues también contamos con otras publicaciones recientes significativas sobre el tema. Para investigadores y estudiantes avanzados, Parodi, Cantos-Gómez y Howe (2022) acaban de editar un libro sobre LC en español en esta misma editorial. Además, en la misma serie que se ha publicado este libro que reseñamos (Routledge Introductions to Spanish Language and Linguistics), podemos encontrar otros dos libros que complementan este volumen: *El léxico-gramática del español. Una aproximación mediante la lingüística del corpus*, de Brown, Paz y Brown (2021); y *Gramática española. Variación social*, de Potowski y Shin (2018). Si bien estos dos últimos libros están pensados como un recurso para que estudiantes que no tienen el español como lengua dominante analicen el español, los tres ofrecen actividades para analizar múltiples variedades del español con corpus. Claramente el libro aquí reseñado, Rojo (2021), es el que nos ofrece más información sobre qué es la LC y los corpus existentes en español. Aunque incluye también actividades de exploración de la gramática y el léxico del español, su función principal es explicar la LC. Estos tres libros se complementan en tanto que los anteriores nos dan más información sobre qué analizar y este último sobre cómo usar corpus para examinar la lengua.

Desde nuestra perspectiva como profesora y estudiantes de español en un programa de grado donde el español no es la lengua dominante, este libro supone ciertos desafíos para los estudiantes. Esta situación se debe tanto a que hemos leído el libro en el primer o segundo curso de lingüística de los estudiantes, como porque para entender la lingüística del corpus hay que tener conocimientos avanzados sobre la gramática y el léxico del español al igual que sobre el metalenguaje. Sin embargo, la repetición pedagógica que los capítulos incluyen facilita el entendimiento en su conjunto. Por ejemplo, el primer capítulo ofrece una panorámica del libro, mientras que los capítulos siguientes profundizan los temas mencionados. Siguiendo este enfoque pedagógico, el cuarto y el quinto capítulo se estructuran de manera similar para entender las diferencias entre búsquedas léxicas y gramaticales. Quizás pensando en estudiantes con un nivel muy básico, el número de actividades se podría aumentar, pero para eso ya contamos con los volúmenes de Brown, Paz y Brown (2021) y Potowski y Shin (2018) que nos ayudarán a buscar qué explorar.

Para lingüistas de otras áreas que quieran saber más sobre la LC, este libro es también un recurso excelente debido a su accesibilidad, sobre todo teniendo en cuenta que la LC ha permeado todas las áreas de la lingüística. Por otro lado,



para los profesores de español como segunda lengua (L2) que respondan a las múltiples llamadas para usar más corpus en sus clases de español L2, este recurso será fundamental (Ascensión-Delaney *et al.* 2015; Cruz Piñol 2012; Elvira García 2021; Jablonkai *et al.* 2020). No se debe olvidar que los profesores de L2 son uno de los colectivos que más se pueden beneficiar del aprendizaje del manejo de corpus (Callies 2019; Heather y Helt 2012; Zareva 2016) y este libro es sin duda asequible para ellos. Para investigadores más avanzados este no es probablemente su libro, pero siempre van a encontrar nuevos puntos sobre los que reflexionar, sobre todo relacionados con el desarrollo de la historia de la LC en español.

Por todos estos motivos, estamos de acuerdo con el autor cuando afirma que este será un libro fundamental para clases de LC en programas de grado e, incluso, en maestrías y doctorados.

**Zoya Gheisar**

Denison University

[gheisa\\_z1@denison.edu](mailto:gheisa_z1@denison.edu)

**Estrella Ortiz**

Denison University

[gheisa\\_z1@denison.edu](mailto:gheisa_z1@denison.edu)

**Jorge Ramírez**

Denison University

[ramire\\_j1@denison.edu](mailto:ramire_j1@denison.edu)

**José Sánchez**

Denison University

[sanche\\_j1@denison.edu](mailto:sanche_j1@denison.edu)

**Tonio Zacco**

Denison University

[zacco\\_a1@denison.edu](mailto:zacco_a1@denison.edu)

**Nausica Marcos Miguel**

Denison University

[marcosn@denison.edu](mailto:marcosn@denison.edu)



## Referencias bibliográficas

### Referencias bibliográficas

- Asención-Delaney, Yuly; Collentine, Joseph G.; Collentine, Karina; Colmenares, Jersus; Plonsky, L. (2015): "El potencial de la enseñanza del vocabulario basada en corpus: optimismo con precaución", *Journal of Spanish Language Teaching*, 2(2), pp. 140-151. <https://doi.org/10.1080/23247797.2015.1105516>
- Brown, Alan V.; Paz, Yanira. B.; Brown, Earl Kjar (2021): *El léxico-gramática del español: Una aproximación mediante la lingüística de corpus*, Oxon/New York: Routledge.
- Callies, Marcus (2019): "Integrating corpus literacy into language teacher education", Sandra Götz; Joybrato Mukherjee (eds.), *Learner Corpora and Language Teaching*, Amsterdam: John Benjamins, pp. 245-263. <https://doi.org/10.1075/scl.92.12cal>
- Cruz Piñol, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*, Madrid: Arco/Libros.
- Elvira-García, Wendy (2021): *Uso de corpus en clase de ELE. La lengua real como modelo*, Barcelona: Difusión.
- Heather, Julian; Helt, Marie (2012): "Evaluating corpus literacy training for pre-service language teachers: Six case studies", *Journal of Technology and Teacher Education*, 20(4), pp. 415-440.
- Jablonkai, Reka; Forti, Luciana; Abad Castelló, Magdalena; Salengros-Iguenane, Isabelle; Schaeffer-Lacroix, Eva; Vyatkina, Nina (2020): "Data-driven learning for languages other than English: the cases of French, German, Italian, and Spanish", Karen-Margrete Frederiksen; Sanne Larsen; Linda Bradley; Sylvie Thouësny (eds.), *CALL for widening participation: short papers from EUROCALL 2020*, Research-publishing.net, pp. 132-137. <https://doi.org/10.14705/rpnet.2020.48.1177>
- Parodi, Giovanni; Cantos-Gómez, Pascual; Howe, Chad (eds.) (2022): *Lingüística de corpus en español/The Routledge Handbook of Spanish Corpus Linguistics*, Oxon/New York: Routledge.
- Potowski, Kim; Shin, Naomi L. (2018): *Gramática española: variación social*. Oxon/New York: Routledge.
- Zareva, Alla (2016): "Incorporating corpus literacy skills into TESOL teacher training", *ELT Journal*, 71(1), pp. 69-79. <https://doi.org/10.1093/elt/ccw045>

### Corpus citados

- Davies, M. (2001): *Corpus del Español (Género/Histórico)*. <[www.corpusdelespanol.org/hist-gen/](http://www.corpusdelespanol.org/hist-gen/)>. [Consulta: 28/7/2022].
- Davies, M. (2016): *Corpus del Español (Web/Dialectos)*. <[www.corpusdelespanol.org/web-dial/](http://www.corpusdelespanol.org/web-dial/)>. [Consulta: 28/7/2022].
- ESLORA: *Corpus para el estudio del español oral* <<http://eslora.usc.es>>, versión 2.0 de septiembre de 2020. [Consulta: 28/7/2022].
- Real Academia Española: *Banco de datos (CORDE)* [en línea]. *Corpus diacrónico del español*. <<http://www.rae.es>> [Consulta: 28/7/2022].



Real Academia Española: *Banco de datos (CORPES XXI)* [en línea]. *Corpus del Español del Siglo XXI (CORPES)*.  
<<http://www.rae.es>> [Consulta: 28/7/2022].

Real Academia Española: *Banco de datos (CREA)* [en línea]. *Corpus de referencia del español actual*.  
<<http://www.rae.es>> [Consulta: 28/7/2022].