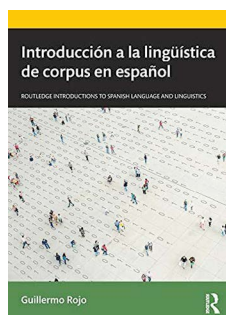


**Rojo, Guillermo (2021): *Introducción a la lingüística de corpus*.
Londres: Routledge. 380 páginas**

MARÍA EUGENIA CONDE NOGUEROL
GI *Gramática del español*. Universidade de Santiago de Compostela
meugenia.conde@usc.es
<https://orcid.org/0000-0002-5568-7423>



La lingüística de corpus (LC, en adelante) —entendida como enfoque metodológico o ámbito de la lingüística que basa sus investigaciones en datos obtenidos a partir de corpus (Parodi, 2010: 14-15)— ha ido adquiriendo en las últimas décadas una presencia considerable en el campo de los estudios lingüísticos hispánicos; hasta el punto de que ha ido modificando la concepción misma del objeto de estudio de la lingüística descriptiva tradicional. Son señales de su auge el creciente interés en la construcción y ampliación de numerosos y variados corpus en español, la cada vez más abundante bibliografía sobre la explotación de sus datos y la publicación de manuales y compendios. A ello han contribuido autores como

Guillermo Rojo, cuyas aportaciones al desarrollo de la LC en español han tenido enorme repercusión en la lingüística hispánica.

Una buena prueba de ello es la publicación de este libro que, aunque concebido originariamente como una introducción práctica con una marcada orientación didáctica —como así lo expresa el autor en el prólogo y así lo demuestra la propia concepción estructural del texto—, trasciende ese propósito inicial para convertirse en una obra que eleva el estatus epistemológico de la LC hispánica y contribuye a seguir dotando la investigación producida en español de una orientación científica de profunda solidez.

El autor se propone dos objetivos: enseñar LC a los estudiantes (de los últimos cursos de grado, de máster o doctorandos) que deseen profundizar o adquirir formación en LC, y publicar la primera introducción a esta corriente en español, pues no existen introducciones

Para citar esta reseña: Conde Noguero, M. E. (2022). Rojo, Guillermo (2021): *Introducción a la lingüística de corpus*. Londres: Routledge. 380 páginas. DOI: <https://doi.org/10.4324/9781003119760> ELUA, (38), 339-341. <https://doi.org/10.14198/ELUA.21356>

Recibido: 18/11/2021 Aceptado: 14/12/2021

© 2022 María Eugenia Conde Noguero



Este trabajo está sujeto a una licencia de Reconocimiento 4.0 Internacional de Creative Commons (CC BY 4.0)

escritas en español que muestren de forma práctica cómo se pueden resolver problemas de lingüística española con datos obtenidos de corpus de español.

Con este doble eje vertebrador —el de ayudar a adquirir conocimientos y el de dotar de una perspectiva global e integradora la investigación en LC en español— Guillermo Rojo ha creado una herramienta que responde a la necesidad actual de dotar de modelos y referentes a futuros investigadores en el manejo de datos lingüísticos en español. Para el desempeño de tal empresa asume la responsabilidad de atender tanto a los problemas sobre los que se puede trabajar, como a la forma en que hay que manejar las aplicaciones de consulta de diferentes corpus, aportando a la investigación en LC en español un soporte material similar al que ya posee la tradición anglosajona con los manuales *Routledge Corpus Linguistics Guides* (2015-2021), en los que la investigación en LC se aplica a una extensa variedad de áreas temáticas: pragmática, vocabulario, gramática, estudios contrastivos, enseñanza de la lengua inglesa, etc.

Es justo aquí donde reside otro de los intereses de esta obra: aportar un carácter globalizador e integrador a esta disciplina lingüística, que se desvela en el afán del autor por aprehender y actualizar con exhaustividad todo el saber relacionado con la LC, para lo que despliega todo un conjunto de conocimientos teórico-prácticos —fruto de una dilatada experiencia— acerca de los avances y desafíos que afronta la LC en el marco de la lingüística actual en español.

Desde sus primeras páginas se subraya el carácter empírico de la LC como ciencia cultural y en ella se destacan, no solo la complejidad y variabilidad como elementos consustanciales de esa concepción empírica y funcionalista del lenguaje, sino también, su estatus epistemológico, que se demuestra en la posibilidad de observación de datos objetivos a gran escala en favor de la *explicabilidad total* (*total accountability*).

El esfuerzo del autor por estructurar y ordenar la densidad de este campo de conocimiento se refleja en la vertebración de la obra en siete capítulos de complejidad progresiva, cuyos contenidos van de lo particular y teórico —qué es un corpus, para qué sirve, cuáles son sus tipos, qué es la LC—, a lo práctico y complejo —cómo se diseña y construye un corpus para su explotación y cómo se recupera e interpreta la información contenida en los corpus (en relación al léxico y los fenómenos gramaticales)—; para finalizar con un recorrido histórico por la LC y sus líneas futuras, y aportar, asimismo, una ampliación y resumen de los procedimientos de recuperación de datos.

Así, el capítulo 1, “La explotación básica de los corpus”, tiene la función introductoria de presentar cuáles son las características de un corpus textual, definir qué es la lingüística de corpus y describir sus aplicaciones a la investigación sobre elementos léxicos y categorías gramaticales, cuestiones diacrónicas, aspectos semánticos y sociolingüísticos y sobre enseñanza y aprendizaje de lenguas, entre otros.

El capítulo 2, “La lingüística de corpus y la metodología de la investigación lingüística”, se ocupa de cuestiones metodológicas y se centra en explicar el carácter empírico de la LC en el marco de la lingüística actual, frente a la lingüística racionalista y a la descriptiva tradicional. Sitúa el marco del debate teórico acerca de la concepción de la LC y desarrolla argumentalmente que se trata de una aproximación al estudio de los hechos lingüísticos de orientación empírica.

De la construcción de corpus se ocupa el capítulo 3, “Diseño, construcción y explotación de corpus”, que consiste en la descripción de las tareas y los aspectos más importantes del trabajo con estos recursos, y en el que se detallan los rasgos de diseño requeridos

para lograr una explotación adecuada de los datos, como son la representatividad, el formato, el volumen, la codificación y la anotación.

La aplicación de las técnicas ya explicadas a fenómenos léxicos y gramaticales son materia de estudio de los capítulos 4 y 5, respectivamente: “Recuperación de información contenida en corpus textuales: el léxico” y “Recuperación de información contenida en corpus textuales: fenómenos gramaticales”. En ellos se han seleccionado ejemplos prácticos reales que atienden a los diferentes ejes variacionales (diacrónico, sincrónico, diatópico y diastrático).

El capítulo 6, “Otras cuestiones centrales en lingüística de corpus”, trata de cuestiones generales relativas a la estructura estadística y al tamaño de los corpus, así como de las ventajas de su uso.

Por último, el capítulo 7, “Herramientas de recuperación de datos: resumen y ampliación”, es una introducción a la ciencia de los datos y se centra en cuestiones avanzadas para aplicaciones de consulta y herramientas informáticas de utilidad no tratadas en capítulos anteriores.

Cada capítulo concluye, asimismo, con un apartado dedicado a lecturas complementarias recomendadas y a cuestiones, problemas y temas de investigación que contienen actividades que retan y motivan a poner en práctica lo aprendido. El libro se cierra con un glosario que sirve de consulta rápida, un índice temático y las referencias bibliográficas a recursos y obras citadas en el texto.

El estudiante o investigador, tanto iniciado como experto, que se acerque a esta obra encontrará un recorrido teórico ordenado, pero también curiosidades, ejemplos de extracción de datos con resultados empíricos, ideas, líneas y temas de investigación que inspiran y estimulan a nuevas investigaciones. A su vez, en el contexto universitario, el docente tiene a su alcance un manual de uso pedagógico cuyos capítulos pueden reforzar y complementar la exposición en el aula.

En la obra se promueve, por tanto, una actitud activa de intercambio con un lector interesado por problemas lingüísticos, al que se dota del conocimiento y las habilidades necesarios para llevar a cabo su propia investigación basada en corpus ya existentes —a la vez que ofrece los pasos para diseñar el suyo propio—, y al que se le capacita para tomar el relevo del maestro/autor.

A lo largo de la lectura se desvela un entusiasmo contagioso por compartir con la comunidad científica, para inspirarla e impulsarla a utilizar la metodología, los recursos y las herramientas de que dispone la LC, gracias a los avances, entre otros factores, de la lingüística computacional. En definitiva, un libro que eleva el estatus de la investigación con datos de corpus y que da fe del compromiso de Guillermo Rojo con la comunidad científica a que va destinado.

REFERENCIAS BIBLIOGRÁFICAS

- Parodi, Giovanni (2010). *Lingüística de Corpus: de la teoría a la empiria*. Iberoamericana Vervuert.
McCarthy, Mike & Anne O’Keeffe (2015-2021). *Routledge Corpus Linguistics Guides*. Oxon, UK: Taylor & Francis.