



RESEÑA

Guillermo Rojo. 2021. *Introducción a la lingüística de corpus en español*, Oxon/New York: Routledge, 404 págs.

María Angélica Verdú

Universidad Nacional del Comahue

Email:

maria.verdu@fadel.uncoma.edu.ar

ORCID

<https://orcid.org/0009-0003-9118-7260>

En la corriente británica y estadounidense de estudios de diversas variedades del inglés, la lingüística de corpus (LC) ha acumulado numerosos volúmenes tales como Sinclair (1991)¹, O’Keefe & McCarthy (2010)² o McEnery & Hardie (2012)³, por mencionar sólo algunos, que les han permitido a los estudiosos del campo de la lingüística, investigar, de la mano de la informatización y el desarrollo de las ciencias de la computación, diferentes aspectos del lenguaje en uso en clave empírica. *Introducción a la lingüística de corpus en español* del reconocido lingüista Guillermo Rojo (GR de aquí en más) viene a cubrir la necesidad de contar con un número mayor de manuales introductorios que discutan temas transversales de la LC en la tradición hispanoparlante. Sin embargo, aquello que resulta novedoso desde un punto de vista de vacancia en la literatura de LC en español podría resultar limitado para expertos en la materia especializados en otras lenguas, debido a que esta obra se centra en el español.

Sin duda, GR nos ofrece una acabada y accesible introducción a cuestiones de relevancia en torno a la exploración científica de esta lengua, al emplear tanto un análisis cualitativo como cuantitativo de los datos compilados en corpus. Partiendo de la premisa de considerar a la LC como “uno de los grandes elementos renovadores de los estudios lingüísticos” (p. xix), el autor se dedica a la investigación de la estructura estadística de textos, entre otros temas afines a la LC actual, utilizando algunas de las herramientas y técnicas de análisis y extracción de datos que nos proporciona esta disciplina, mostrando, en específico, corpus en español.

En cuanto a su organización interna, este manual introductorio, de enfoque e impacto didáctico, cuenta con siete capítulos graduados en la progresión de su complejidad, adecuadamente definidos y de lectura amena, sin jerga técnica innecesaria. Cada una de estas siete secciones bien ordenadas se inicia con un breve resumen de los contenidos a discutirse, y desarrolla aproximaciones teóricas pertinentes, vertebrando las discusiones de manera tal que se facilita el seguimiento del contenido, a la vez que se propicia la lectura selectiva de capítulos individuales en específico para quienes se acerquen al volumen con una línea investigativa particular y no

¹Sinclair, John (1991). *Corpus, concordance, collocation*. Oxford University Press.

²O’Keefe, Anne & Michael McCarthy (Eds.). (2010). *The Routledge Handbook of Corpus Linguistics*. Routledge.

³McEnery, Tonny & Andrew Hardie (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

busquen leerlo en su totalidad. A modo de cierre, cada uno de estos siete títulos incluye un apartado en el que GR despliega una exhaustiva lista de lecturas complementarias recomendadas, agrupadas por área de interés a explorar en mayor detalle. Sin perder de vista el tono práctico y didáctico que caracteriza a esta obra, la sección final plantea interesantes aplicaciones a través de problemas y temas de investigación de casos específicos que permiten acceder al uso del corpus para examinar la lengua, entre otras cuestiones.

Asimismo, se detalla un glosario bilingüe (inglés-español) relativamente útil para quienes están dando sus primeros pasos en este apasionante mundo, al que el autor suma el listado completo de obras citadas y una minuciosa recopilación de los corpus textuales y otros recursos electrónicos mencionados en el texto. Se agradece la variedad de los ejemplos y numerosos elementos ilustrativos de análisis y extracción de datos incorporados a este volumen, que podrán resultar de utilidad y brindar claridad tanto para estudiantes de diversos espacios de formación de grado o posgrado, como para especialistas en la materia. A modo de cierre, el texto incorpora un práctico índice de materias de consulta rápida. A continuación, se hará mención a algunos de los contenidos más significativos a los que GR da tratamiento en cada una de las siete secciones principales que integran la presente publicación.

A través del primer capítulo, titulado “La explotación básica de los corpus”, GR abre una ventana para examinar de manera global y a modo de sección introductoria algunas cuestiones relacionadas con qué es un corpus y sus posibles usos. Abundan las referencias a posibles investigaciones de tono más empírico en torno al estudio de la lengua: elementos léxicos, clases de palabras y otras categorías gramaticales, aspectos semánticos, cuestiones diacrónicas, aspectos sociolingüísticos, combinaciones de palabras, fenómenos fónicos, enseñanza y aprendizaje de lenguas, entre otros. Además, se listan posibles tipos de corpus y se enfatiza el rol central de la representatividad y el cuidado en la selección de los textos constitutivos del corpus en desarrollo. En palabras del autor, este capítulo sirve para “mostrar, en diferentes áreas, que los corpus constituyen el modo más adecuado de llevar a cabo estudios de carácter empírico⁴ (con datos objetivos, replicables y pertinentes con relación al fenómeno estudiado)” (p. 1). Muchos de estos puntos serán retomados en mayor profundidad en secciones siguientes.

En “La lingüística de corpus y la metodología de la investigación lingüística”, el segundo capítulo de este volumen, Rojo articula la discusión, atendiendo al enfoque histórico, con cuestiones metodológicas generales relacionadas con la organización del conocimiento científico y el método hipotético-deductivo en la metodología de la investigación. El tercer capítulo denominado “Diseño, construcción y explotación de corpus” plantea, con un poco más profundidad de la esbozada en el primer capítulo, la caracterización general de los corpus lingüísticos y sus diversos tipos. La propuesta recupera aclaraciones generales estrechamente ligadas al trabajo con corpus como son el diseño, el formato, el volumen, la codificación y la anotación lingüística. Tal como expresa GR, el objetivo que se persigue aquí está guiado por la idea de que “la familiarización con todas las fases del trabajo permitirá una mejor comprensión de los factores implicados y contribuirá a una explotación más adecuada de los datos obtenidos” (p. 62). En esta tercera sección, se discuten, también, cuestiones legales y éticas relevantes de cara a la explotación de corpus que ya habían sido planteadas por algunos de los volúmenes ya mencionados, pero que resultan más que bienvenidas.

Hasta aquí, el autor ha abordado cómo se pueden utilizar las técnicas de análisis de corpus para estudiar la variación en diferentes niveles lingüísticos, desde la morfología hasta la sintaxis y la pragmática. A partir de los capítulos cuatro y cinco, el manual incorpora temas centrales a la investigación-acción basada en corpus en función de las técnicas ya puntualizadas en aportaciones vertidas en secciones anteriores. El cuarto,

⁴Véase Reppen, Randi & Rita Simpson-Vlach (2020). Corpus linguistics. En Norbert Schmitt & Michael Rodgers (Eds.), *An Introduction to Applied Linguistics* (3^{ra} edición. pp. 89 - 108). Routledge. “Empírico”, en tanto posibilita el “análisis de patrones reales de uso en textos naturales” (p.91, traducción propia), sin que esto implique que se trate de estudios de carácter empírico con recolección de datos.

“Recuperación de información contenida en corpus textuales: el léxico”, es, posiblemente, uno de los más interesantes, ya que, como se explica en el resumen de inicio, en él se explora la explotación de corpus anotados y lematizados para la recuperación e interpretación de información sobre elementos léxicos y se “muestran las formas en las que se puede extraer y analizar el contenido de los corpus textuales mediante ejemplos reales para investigar los diferentes aspectos relacionados con el componente léxico del español” (p. 127). GR aprovecha para discutir cuestiones como la frecuencia de elementos y fenómenos léxicos de ciertas formas ortográficas, lemas, unidades multipalabra y expresiones complejas. Asimismo, plantea la importancia de la variación en el léxico desde los puntos de vista diatópico (geográfico) y diacrónico (temporal), diastrático (de estratificación socio-cultural) y diafásico (de acuerdo a diferentes situaciones comunicativas). Ahora bien, la profundidad dedicada a la dimensión semántica en el análisis de corpus es quizás un punto perfectible, si se persigue el fin de realizar una contribución más acabada en este sentido.

El capítulo quinto, “Recuperación de información contenida en corpus textuales: fenómenos gramaticales”, en vez de tratar el léxico tal como lo hace el cuarto, se centra en la recuperación selectiva de información en corpus que no ofrecen un análisis sintáctico, sino que incorporan únicamente la anotación morfosintáctica en torno a investigaciones con foco en aspectos gramaticales, la variación gramatical y el cambio gramatical. Aquí, GR ilustra y analiza una amplia variedad de fenómenos relacionados con la frecuencia de las clases de palabras y de categorías y subcategorías gramaticales (conjugaciones, modos y tiempos verbales, perífrasis verbales, adverbios en *-mente*) y concordancia y otros fenómenos afines. No escapan a este muestreo algunos fenómenos gramaticales en perspectiva diacrónica, como así también algunas aplicaciones en enseñanza y aprendizaje de segundas lenguas (L2) muy necesarias para profesionales de L2 o lenguas extranjeras, quienes pueden llegar a encontrar inspiración en estas páginas para dar vida a la explotación de corpus en el aula de español como L2 al diseñar sus propios ejercicios.

De los antecedentes históricos y evolución de la LC se ocupa el sexto capítulo, “Otras cuestiones centrales en lingüística de corpus”, en el marco de un presente atravesado por los vertiginosos avances y el alto nivel de sofisticación de la computación y la informática. En este sentido y para alcanzar un mayor entendimiento de la trama que sostiene los contenidos revisados hasta aquí con relación a las aplicaciones de la LC, GR señala algunas de las ventajas e inconvenientes del uso de corpus textuales hoy, al hacer referencia a la necesidad de una anotación y codificación cada vez más avanzada (en los textos de los corpus existentes). Aparte de esto, el capítulo también discute la dimensión que tiene que ver con una visión más cuantitativa al referirse a la estructura estadística y el tamaño de los corpus, el tema de la representatividad y el equilibrio. Esta sección concluye con la discusión del futuro de la LC y su diálogo con otras ramas y disciplinas lingüísticas.

GR dedica el séptimo y último capítulo, titulado “Herramientas de recuperación de datos”, a la revisión de procedimientos de recuperación por medio de herramientas informáticas de utilidades de carácter general no introducidas en secciones anteriores. Como expone el autor, se trata de “herramientas propias de algunos sistemas operativos que pueden facilitar considerablemente el trabajo con los textos, las listas de ejemplos, listas de frecuencias y, en general, tanto con los textos como con los resultados de su análisis” (p. 303). Quienes buscan una guía más detallada pueden necesitar recursos adicionales, pero este capítulo es un buen punto de partida.

En resumen, este libro se constituye en un manual adecuadamente balanceado y bien logrado en el campo de los estudios lingüísticos hispanicos y de la lingüística de corpus actual que, como su título lo indica, no presume ser nada más y nada menos que lo que, en esencia, es: una muy efectiva y amigable puerta de acceso a la exploración de los temas que analiza. Se trata de una obra que es una fuente de información complementaria sumamente provechosa para adentrarse en la profundización de aquellos conceptos que resulten de interés al

lector que la aborda. Además, puede estimular a quienes persiguen metas investigativas personales, ya que se verán beneficiadas por los aportes de la LC en sus ramas específicas de estudio.

En un momento histórico que nos invita a usar las avanzadas herramientas de desarrollo de software a nuestra disposición, la lingüística de corpus pone a nuestro alcance los medios que permiten iluminar el proceso de estudio de las lenguas como nunca antes. En nosotros radica la responsabilidad de apropiarnos de ellas y utilizarlas a nuestro favor. Tal como destacan McEnery & Xiao (2011)⁵, la formación en análisis de corpus es un área que no ha recibido el protagonismo que merece si ponemos en la balanza las prestaciones y utilidades que ofrece a los profesionales de las lenguas e investigadores en ejercicio. Indudablemente, desde su aporte a explicar la LC y el manejo de datos lingüísticos en español, la obra de GR en su totalidad debiera ser entendida como una contribución innegable a la concientización en este sentido. *Introducción a la lingüística de corpus en español* es, ciertamente, un manual de utilidad, necesario en cualquier biblioteca de lingüística y de la lengua española, y una valiosa adición al campo de la LC en general, como una disciplina fascinante y, en palabras de Parodi (2008: 93)⁶, “en (constante) ebullición”.

⁵McEnery, Tony & Richard Xiao (2011). What corpora can offer in language teaching and learning. En Eli Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (Vol. 2, pp. 364-380). Routledge.

⁶Parodi, Giovanni (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de Lingüística Teórica y Aplicada*, 46 (1), 93-119. <http://dx.doi.org/10.4067/S0718-48832008000100006>