

## RESENHA/REVIEW

MCENERY, Tony & Andrew HARDIE. *Corpus Linguistics* (Cambridge Textbooks in Linguistics). Cambridge, UK: Cambridge University Press (CUP), 2012. 312 p. ISBN 978-0-521-54736-9

Resenhado por/by: Camilo Enrique Díaz ROMERO  
(*Universidad Nacional de Colombia*)

“As corpus linguistics is a collection of methods, researchers need to determine which ones are most applicable in addressing their research questions, along with deciding which software will be used (often the affordances of the latter will heavily impact on the former)”. (Baker, 2010a: 19)

“Corpus linguistics is an increasingly popular field of linguistics which involves the analysis of (usually) very large collections of electronically stored texts, aided by computer software. The word *corpus* is Latin for *body* – a corpus is therefore a ‘body’ of texts” (Baker, 2010b: 93)

Los profesores Tony McEnery y Andrew Hardie, de la Universidad de Lancaster, nos presentan un libro que expone los avances y desarrollos que ha venido teniendo los estudios de la lingüística del *corpus* en los últimos 30 años.

Si se comparan las citas sobre lo que el profesor Paul Baker, de la Universidad de Manchester, afirmó sobre lo que conforman este grupo especial de modernos estudios del lenguaje humano con lo que aparece a continuación, nos daremos cuenta que se ha venido consolidando una definición sobre lo que es la lingüística del *corpus*:

“What is corpus linguistics? It is certainly quite distinct from most other topics you might study in linguistics, as it is not directly about the study of any particular aspect of language. Rather, it is an area which focuses upon a set of procedures, or methods, for studying”. (p.1)

El libro, como tal, se divide en 3 partes: aspectos preliminares, los capítulos y la sección posterior.

La primera parte, de 18 páginas, se compone de: una portada, una contraportada, una lista de referencias a otros libros de lingüística que CUP ha publicado desde la década de 1970 hasta nuestros días, un índice de figuras, un índice de tablas, una nota de agradecimientos a los lingüistas que les colaboraron con la redacción y edición del libro, y un prólogo. En esto último, dejan claro que este texto se ha diseñado con una visión crítica de los conceptos que se vienen manejando dentro de esta rama de los estudios del lenguaje, con miras a que logren ser precisados aún más y lleguen a ser más confiables para futuros estudios (p. xiii-xv).

La segunda parte, de 237 páginas, se compone de 9 capítulos: “1. What is corpus linguistics” (p.1), “2. Accesing and analyzing corpus data” (p.25), “3. The web, laws and ethics” (p.57), “4. English Corpus Linguistics” (p.71), “5. Corpus-based studies of synchronic and diachronic variation” (p. 94), “6. Neo-Firthian corpus linguistics” (p. 122), “7. Corpus methods and functionalist linguistics” (p. 167), “8. The convergence of corpus linguistics, psycholinguistics and functionalist linguistics” (p. 192) y “9. Conclusion” (p. 225).

El primer capítulo ilustra la conceptualización básica para entender lo que se entiende actualmente por *lingüística del corpus*, las clases de datos que pueden constituir estas grandes masas de información, así como las dos clases de enfoques sobre esta rama de las ciencias del lenguaje: estudios que se basan en los cuerpos de datos para corroborar o refutar alguna tesis (*corpus-based linguistics*) frente a los que consideran que estas grandes cantidades de datos son las fuentes sobre las cuales se generan las teorías lingüísticas actuales (*corpus-driven linguistics*). Es de resaltar que aquí ya no solo se menciona la existencia de *corpus* constituidos de registros escritos u orales, sino también de gestos de la cara y movimientos de la mano en las diferentes lenguas de señas existentes en Norteamérica (p.3). Además, si se compara lo que se presenta en este capítulo con el pensamiento que había sobre la cantidad de datos a utilizar en cada análisis basado en esta naciente rama de la lingüística por parte del profesor de la Universidad de Birmingham John Sinclair (2004), observamos que los cuerpos de datos ya no tienen que ser constituidos por millones de entradas léxicas de diccionario o miles de registros sonoros para poder mostrar significativamente

la funcionalidad de una lengua en tiempo real y que, incluso, a veces es preferible hacer selecciones de información para hacer estudios sobre variación de formas lingüísticas (cf.: p.8-13).

El segundo capítulo presenta un resumen histórico de las técnicas y herramientas que se han estado empleando en la lingüística del *corpus*. Llama la atención que estos los autores del libro reconocen aquí que existe un debate acerca de si la posesión de grandes cantidades de datos organizadamente recolectados pueden responder todas las preguntas actualmente existentes en las ciencias del lenguaje. Para ello, ponen un ejemplo procedente de los resultados sobre los estudios que se llevan a cabo sobre los verbos del inglés:

“without a corpus, we could certainly examine, and seek to describe, the use of non-finite verbs in English –many scholars have, for instance O’Dwyer (2006: 58–9). However, the number of examples of non-finite verbs that we could base our investigation on would remain relatively small, being limited by the hand-and-eye techniques that we would need to find them.” (p.27)

El tercer capítulo expone uno de los aspectos de la lingüística de *corpus* más polémicos: la legislación y los aspectos éticos que se deben tener en cuenta en el momento de usar grandes cantidades de información en investigaciones, sobre todo, cuando proceden de internet.

El cuarto capítulo presenta el desarrollo de estudios de *corpus* del inglés que han realizado 5 universidades del mundo angloparlante y la Universidad Católica de Lovaina, Bélgica. Queda claro que buena parte de las investigaciones que se están realizando son sobre lexicología y morfología, por lo que aún falta mucho por hacer para que se constituyan más *corpus* de registros sonoros para estudios de fonética y fonología.

El quinto capítulo exhibe las ventajas que han producido los estudios de lingüística de *corpus* en la caracterización de la variación de sincrónica y diacrónica del vocabulario y la gramática del inglés.

El sexto capítulo expone los desarrollos que se están llevando a cabo en lingüística del *corpus* teniendo como base teórica los postulados de la Escuela de Londres, cuyo principal exponente es el profesor John Rupert Firth, los cuales fueron actualizados con los estudios lexicológicos del profesor de la Universidad de Birmingham John Sinclair.

El séptimo capítulo presenta los avances que han tenido los estudios funcionalistas del inglés al hacer uso de *corpus* con registros escritos del inglés, especialmente, en sintaxis y semántica.

El octavo capítulo expone los avances que han tenido los estudios psicolingüísticos de adquisición y desarrollo de una lengua en un niño, así como en la modelación computacional del lenguaje al hacer uso de *corpus* con registros escritos y orales.

El último capítulo resume el panorama en que se encuentran los desarrollos de los estudios de la lingüística del *corpus*, así como las posibilidades que tienen de seguir expandiéndose. La pregunta que queda es por qué no se han contemplado aún desarrollar estudios de esta naturaleza para corroborar o refutar modelos y tipologías que se han venido postulando en la fonología en los últimos años (Sedlak, 1969; Maddieson, 1984; Hyman, 2007, 2009). No obstante, los autores del libro son conscientes de que esta rama de los estudios del lenguaje, todavía emergente, ya encuentra unas fronteras teóricas que cada vez más se sabe que no podrían traspasar: “There are few areas of linguistics where there is no possible role for corpus methods (the most obvious example being Chomskyan theory)” (p.236).

En general, es buen libro para conocer las principales tendencias en estudios de la lingüística que se están realizando con datos recogidos en su cotidianidad.

Recebido em maio de 2012  
Aprovado em agosto de 2012  
E-mail: cediazr@unal.edu.co

## REFERENCIAS

- BAKER, Paul. *Sociolinguistics and Corpus Linguistics*. Edinburgh, Scotland: Edinburgh University Press Ltd, 2010a.
- \_\_\_\_\_. *Corpus Methods in Linguistics*. In: LITOSSELITI, Lia. Ed. *Research Methods in Linguistics*. London, UK: Continuum International Publishing Group, 2010b. p. 93-113.
- HYMAN, Larry M. Where's phonology in typology. *Linguistic Typology* 11, 2007. p.265-271

- \_\_\_\_\_. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31, 2009. p. 213-238.
- MADDIESON, Ian. *Patterns of sounds*. Cambridge, MA: Cambridge University Press, 1984.
- O'DWYER, B. *Modern English Structures: Form, Function and Position*. New York, USA: Broadview Press, 2006.
- SEDLAK, Philip. Typological considerations of vowel quality systems. *Working Papers on Language Universals* 1, 1969. p. 1-40.
- SINCLAIR, John. *Trust the Text: Language, corpus and discourse*. London, UK: Routledge (Taylor & Francis Group), 2004.