

Camilo E. Díaz Romero  
Universidad Nacional de Colombia

RESEÑA DE BAKER, P. (2010). *SOCIOLINGUISTICS AND CORPUS LINGUISTICS*.  
EDIMBURGO, ESCOCIA: EDINBURGH UNIVERSITY PRESS

There is a growing realization that the basis of intersubjective knowledge in linguistics must be found in speech-language as it used in everyday life by members of the social order, that vehicle of communication in which they argue with their wives, joke with their friends, and deceive their enemies (Labov, 1972: xiii).

En este documento, el profesor Paul Baker (Universidad de Manchester) nos informa acerca del papel que están jugando los estudios sobre lingüística del *corpus* en la caracterización de la variación y el cambio de las formas y los significados de todo tipo de ítems lingüísticos dentro de un contexto social. Este libro, que es significativamente breve, de 200 páginas, se divide en tres partes: aspectos preliminares, los capítulos y sección final.

En la primera parte, de 11 páginas, tenemos la portada, la tabla de contenidos y las listas de figuras y cuadros. En la segunda parte, de 156 páginas, encontramos los siete capítulos en los cuales se van a tratar los contenidos principales del libro. El primero es la introducción, en la cual se nos precisa la conceptualización básica para entender lo que se explica en los capítulos que lo siguen, el segundo va a ilustrar el uso que se está haciendo de los diferentes *corpus* en la descripción de variaciones de pronunciación a nivel sociolectal o de uso de determinado vocabulario según si es hombre o mujer, si tiene menos de 35 años o más, o si perteneces a un estrato socioeconómico alto o bajo; el tercero nos enseña acerca del papel que tienen los registros escritos en la identificación de la variación lingüística que pudo haberse dado en siglos pasados dentro de una lengua, el cuarto expone los resultados que se han obtenido sobre la variación sincrónica de una lengua comparando grandes cantidades de datos que eran tomados sobre diferentes dialectos del inglés, el quinto resalta el papel de los *corpus* de archivos sonoros en la identificación de tendencias de variación en entonación y pausas en las conversaciones que se realizan entre personas que hablan diferentes dialectos, el sexto trata aquellos aspectos que, gracias a toda la información recolectada en los medios de comunicación y la literatura, ilustran variaciones semánticas y pragmáticas en la interpretación del uso de distintas formas lingüísticas por parte de hombres y mujeres del mundo angloparlante, y el séptimo resume todo lo tratado en los anteriores capítulos y expone las principales tareas que hacen falta por desarrollar en la interacción entre la lingüística del *corpus* y los estudios de sociolingüística.

Es de resaltar que en estos capítulos se nos informa que se está rescatando buena parte de los estudios de la Escuela de Londres en cuanto al surgimiento del significado de nuevas palabras con base en el contexto en donde aparecen (cf. Baker, 2010: 24-26, 113-116) y en cómo se emplean términos del dominio disciplinar de la fonología, como la prosodia, en la semántica o la pragmática (cf. *Ibid*:49, 50, 132-134).

En la tercera parte, de 33 páginas, se expone la bibliografía, un apéndice, que consiste en la lista de abreviaturas que se usaron en todo el libro, todas las notas con comentarios sobre información existente en los capítulos y referencias de textos disponibles en internet, y un índice de temas ordenado alfabéticamente.

Queda claro que los análisis expuestos en el libro son hechos principalmente sobre los cuerpos de información procedentes de registros orales y escritos del inglés. Aparte de investigaciones sobre datos del mundo angloparlante, tenemos una breve mención, en las páginas 46 y 48 del libro, de un estudio sobre variación del léxico del español, en el cual se cita el estudio de Biber *et.al.* (2006), el cual se basa en el *Corpus del Español*, diseñado por Mark Davies (2002).

Es interesante observar que la sociolingüística, en sus 50 años de desarrollo, ha conservado buena parte de sus objetivos de estudio y el propósito de los datos que se recolectan. Sobre esto último, esto era lo que Brigitte

Schlieben-Lange afirmaba en la década de 1970 que debían ser los objetivos de la información que se recolecte para estudios de variación lingüística condicionada socialmente:

1. Ha de aportar pruebas o *muestras de lengua actuales*, textos tal como de hecho se performan.
2. Ha de recoger *datos objetivos relativos a la persona y relativos a las modalidades del uso lingüístico*. 3. Ha de averiguar las *actitudes* respecto a las diversas formas de lenguaje (Schlieben-Lange, 1977: 158).

A continuación, se exponen los papeles que pueden tener los *corpus* en la sociolingüística: “A corpus is therefore a ‘body’ of language, or more specifically, a (usually) very large collection of naturally *occurring* language, stored as computer files” (Baker, 2010:6), “Corpora are usually carefully constructed so that they can be said to be representative of a particular language or language variety” (*Ibidem*) y “corpora themselves may reflect social attitudes or ‘discourses’ (ways of looking at the world)” (*Ibid*: 30).

También, se destaca aquí que se hace un llamado urgente a trabajar en la creación de cuerpos de archivos sonoros más grandes para poder identificar variaciones lingüísticas de la oralidad en tiempo real:

With the advent of technologies that make recording, transcription, segmentation, time alignment and phonetic and prosodic analysis more rapid and straightforward, it is hoped that larger spoken corpora will become available, although as many corpus builders will attest, it is still time consuming and potentially difficult to work with spoken data. However, hopefully the studies described in this section will indicate how corpora can be gainfully exploited as a resource for sociolinguists focusing on prosodic and phonetic variation (*Ibid*: 56).

La esperanza reside en los *corpus* de los archivos sonoros del inglés, de los cuales se informan sus principales desarrollos en buena parte del capítulo 2 del libro, en los estudios que la profesora Kate Beeching (Universidad del Occidente de Inglaterra en Bristol) ha estado haciendo con colecciones de grabaciones sobre diálogos del francés (Beeching 1997, 2006) y las investigaciones que los profesores de la Universidad de Gotemburgo J. Allwood, M. Björnberg, L. Grönqvist, E. Ahlsén, C. Ottesjö y J. Nivre están llevando a cabo en el sueco (Allwood *et.al*, 2000; Nivre y Grönqvist, 1999).

Sin embargo, esto aún es muy reciente, por lo que los mayores desarrollos, y eso lo deja muy claro el libro, provienen de los *corpus* escritos (Baker, 2010: 3, 6, 7, 12, 13, 15, 21), de los cuales existen muchísimos registros y han sido compilados a lo largo de varios siglos, primero, en monasterios, después, en bibliotecas, y ahora, en los computadores.

## REFERENCIAS

- Allwood J., Björnberg M., Grönqvist L., Ahlsén E. y Ottesjö C. (2000). “The Spoken Language Corpus at the Linguistics Department, Göteborg University”. *Forum Qualitative Social Research*, 1(3).
- Beeching, K. (1997). “French for specific purposes: the case for spoken corpora”. *Applied Linguistics*, 18(3), 374-394. doi:10.1093/applin/18.3.374
- Beeching, K. (2006). “Synchronic and diachronic variation: the how and why of sociolinguistic corpora”, en Wilson, A., Rayson, P y Archer, D (eds.), *Corpus Linguistics around the world*. Amsterdam, Holanda: Editions Rodopi S.V., 49-60.
- Biber, D., Davies, M., Jones, J. K. y Tracy-Venutra, N. (2006) “Spoken and written register variation in Spanish: A multidimensional analysis”. *Corpora*, 1(1), 1-38. doi:10.3366/cor.2006.1.1.1
- Davies, M. (2002). *Corpus del Español: 100 million words, 1200s-1900s*. <http://www.corpusdelespanol.org>. [fecha acceso: 10.4.2012].
- Labov, W. (1972). *Sociolinguistic Patterns*. Pennsylvania, Estados Unidos: University of Pennsylvania Press.
- Nivre, J. y Grönqvist, L. (2001). “Tagging a Corpus of Spoken Swedish”. *International Journal of Corpus Linguistics*, 6(1), 47-78. doi:10.1075/ijcl.6.1.03niv
- Schlieben-Lange, B. (1977) *Iniciación a la sociolingüística*. Madrid, España: Editorial Gredos.