

**RESEÑA A KABATEK, JOHANNES (ED.) Y CARLOTA DE BENITO MORENO (COL.)  
(2016): LINGÜÍSTICA DE CORPUS Y LINGÜÍSTICA HISTÓRICA  
IBERORROMÁNICA, (BEIHEFTE ZUR ZEITSCHRIFT FÜR ROMANISCHE  
PHILOLOGIE, BAND 405), BERLIN / BOSTON: DE GRUYTER  
[ISBN: 9783110460223]**

ANA PAZ AFONSO  
*Universitat Autònoma de Barcelona*  
ana.paz@uab.cat

La obra editada por Johannes Kabatek en colaboración con Carlota de Benito se corresponde con el volumen 405 de la colección «Beihefte zur Zeitschrift für romanische Philologie», a cargo de Claudia Polzin-Haumann y Wolfgang Schweickard, y publicada por la editorial De Gruyter. Se trata de un volumen que recoge 18 investigaciones cuyo origen se encuentra, por un lado, en las ponencias y comunicaciones presentadas en el marco del Tercer Coloquio Internacional sobre Corpus Diacrónicos en lenguas iberorrománicas (CODILI III), celebrado en Zúrich, del 23 al 25 de junio de 2014; y, por otro lado, en las reflexiones surgidas a lo largo del curso de invierno Abriendo Líneas en el Pasado del Español (ALPES), celebrado en Kandersteg (Berna) del 18 al 22 de enero de 2016. Todas ellas se articulan en torno a un tema común que actúa como eje vertebrador de todo el volumen: el tratamiento de la lingüística histórica basada en corpus desde una perspectiva crítica. Tal y como señala el propio Kabatek en su «introducción a la introducción» (p. 1), el volumen pretende cuestionar especialmente los datos primarios, en particular, su relación con la historia de la lengua, así como el diseño de los corpus y las posibilidades que estos ofrecen.

Los estudios que componen el volumen se organizan en cuatro grandes apartados en función de sus principales objetivos, precedidos por una introducción realizada por Johannes Kabatek y titulada «Un nuevo capítulo en la lingüística histórica iberorrománica: el trabajo crítico con los corpus. Introducción a este volumen» (pp. 1-17). En ella se plantea el tema central de la obra y los puntos principales de debate, mencionados en el párrafo anterior. Justifica, asimismo, el enfoque iberorrománico de los estudios reunidos (pp. 2-3) y realiza una interesante reflexión acerca del concepto actual de corpus y su representatividad (pp. 3-5) para, finalmente, centrarse en la defensa de los estándares de la red CHARTA (pp. 6-8), la descripción de las tendencias principales en la configuración de corpus históricos, así como una consideración final sobre los desafíos que presenta el incremento del volumen de datos y los factores a tener en cuenta para su interpretación (pp. 9-10). Por último, finaliza con una breve revisión de los trabajos que componen el volumen (pp.10-14) y una lista de corpus y bases de datos (p. 15-16).

En primer lugar, en la primera sección, titulada «I. Contribuciones a la lingüística de corpus desde las lenguas iberorrománicas» (pp. 19-112) se encuentran cuatro trabajos que tratan sobre los problemas generales de la lingüística de corpus. El primero de ellos, de Andrés Enrique-Arias, se titula «Sobre la noción de perspectiva en lingüística de corpus: algunas ventajas de los corpus paralelos» (pp. 21-39) y constituye una defensa de este tipo de corpus a partir del ejemplo de *Biblia Medieval*. Aboga por las ventajas que estos presentan respecto a los demás corpus «convencionales» (*CORDE*, *CDH* y *CdE*, principalmente), como la posibilidad de disponer de equivalentes de traducción (p. 26) y de realizar búsquedas «de las unidades de expresión posibles para la estructura que se está investigando» (p. 28).

Por su parte, en «Traducción y tradición en los corpus: nuevas perspectivas para la lingüística histórica» (pp. 40-56), el propósito fundamental de Santiago del Rey Quesada es la defensa de los estudios de traducción basados en corpus debido a la importancia de las traducciones en la historia de la lengua y el desarrollo de las tradiciones discursivas; de ahí que afirme que es necesario trabajar en la elaboración de más corpus paralelos con el fin de calibrar la influencia de las traducciones respecto a los textos originales en diferentes momentos históricos.

Álvaro S. Octavio de Toledo y Huerta, en «Aprovechamiento del *CORDE* para el estudio sintáctico del primer español moderno (ca. 1675-1825)» (pp. 57-89), destaca la importancia de esta etapa para los estudios históricos y diacrónicos y respalda la idea de que el *CORDE* es «a día de hoy, la única herramienta mínimamente fiable desde el punto de vista filológico» (p. 61), que sustenta a partir de un análisis del volumen de textos incluidos en dicho corpus. A partir de búsquedas en el *CORDE* de las construcciones *antes sí/sí solo*, *nada SV/no SV nada*, *en el ínterin/en el ínterin que/en ínterin que/ínterin que/ínterin* plantea las limitaciones del uso de las frecuencias correlativas y, en su lugar, aboga por el empleo del «índice de difusión textual» (p. 71) y la consideración de la «dinámica variacional» (p. 76).

El último capítulo de esta sección es «Tres propuestas en el ámbito de la lingüística de corpus» (pp. 90-112), de Joan Torruella, que aporta sugerencias para mejorar, en este caso, el *Corpus Informatizat del Català Antic (CICA)*: en primer lugar, dividir los periodos cronológicos en cortes de 50 años con el fin de evitar sesgos de tipo histórico o sociocultural; en segundo lugar, mantener el equilibrio textual mediante un sistema de correspondencia «equivalente» para los ejes diacrónico y tipológico (p. 105) –si bien opta por el sistema proporcional para el eje dialectal–; y, en tercer lugar, cuantificar la riqueza léxica mediante un índice NOMC % (pp. 105-110).

La segunda sección lleva por nombre «II. Corpus iberorrománicos» (pp. 113-212) y reúne cinco investigaciones que presentan nuevos corpus. El primero de ellos, «Iluminar los *Séculos Escuros: Gondomar*, un corpus para el estudio del gallego de la Edad Moderna» (pp. 115-136), de Rosario Álvarez y Ernesto González Seoane, muestra las características del *Corpus dixital de textos galegos da Idade Moderna (Gondomar)*, que recoge documentación gallega desde los *Séculos Oscuros* hasta el siglo XVIII, y se detiene en aspectos relacionados con la caracterización general de los textos y su tipología, así como en los problemas para el establecimiento del corpus, como la «dispersión física y la invisibilidad de sus piezas» (p.

124), la transmisión lingüística a través de personas que no dominaban la lengua (p. 125) o la ausencia tanto de referentes contemporáneos como de cánones (p. 125).

A continuación, el texto de Maria Francisca Xavier, «O CIPM – Corpus Informatizado do Português Medieval, fonte de um Dicionário exhaustivo» (pp. 137-156), expone los rasgos principales de este corpus, que contiene muestras de «textos portugueses e galegos» (p. 137) desde el siglo XII al XVI, así como las decisiones tomadas en cuanto a la anotación y las herramientas informáticas adaptadas para su explotación. Se hace hincapié, además, en que parte de sus materiales se han empleado para la elaboración del *Dicionário da Língua Portuguesa Medieval*, de cuya confección se describe la metodología y el tratamiento de las variantes gráficas, fonéticas y morfológicas (pp. 146-150).

El siguiente capítulo, «La documentación medieval de Miranda de Ebro: presentación del corpus y rasgos lingüísticos» (pp. 157-174), de Vicente J. Marcet Rodríguez y M.<sup>a</sup> Nieves Sánchez González de Herrero, trata sobre un corpus medieval de ámbito geográfico restringido a la zona de Miranda de Ebro formado por 203 documentos desde el siglo XIII al XV. En el texto se describe el proceso de edición del corpus (p. 161-162), a partir de los criterios de la red CHARTA, y el análisis lingüístico realizado (p. 162), centrado en cuestiones gráfico-fonológicas, diatópicas, gramaticales, morfológicas y de cambio lingüístico, en general.

«A idade dos *desvios*: diacronía, variação social e linguística de corpus» (pp. 175-196), de Catarina Carvalheiro, Ana Luísa Costa, Rita Marquilhas, Clara Pinto, Fernanda Pratas y Gael Vaamonde, presenta el corpus *Post Scriptum*, que incluye cartas privadas tanto en español como en portugués europeos y americanos desde el siglo XVI hasta principios del XIX. Los autores ponen de manifiesto su utilidad a través del análisis del marcador discursivo *pois* y el relativo *cujo* en portugués y los fenómenos de laísmo, loísmo y leísmo en español.

El último capítulo de esta sección es «CITIVS, MAIVS, MELIVS: del CREA al CORPES XXI» (pp. 197-212), en el que Guillermo Rojo presenta las características de este último corpus y realiza una revisión de los corpus de la RAE desde el CORDE y CREA de 1998 a la nueva versión del CORPES XXI. Aborda, además, la cuestión de la representatividad y se plantean los nuevos objetivos del CORPES XXI para el futuro.

Como su nombre indica, la tercera sección, «III. Corpus y análisis cuantitativos» (pp. 213-381), se centra en los aspectos estadísticos y es la que contiene más contribuciones, un total de seis. El primero de ellos es «Notas sobre la aportación del análisis estadístico a la lingüística de corpus» (pp. 215-237), de Dorien Nieuwenhuijsen, en el cual el autor defiende la utilidad de los datos estadísticos mediante el análisis de un corpus —extraído del CdE— de oraciones subordinadas interrogativas indirectas negadas dependientes del verbo *saber* desde el siglo XIII al XX. Nieuwenhuijsen se centra en el cálculo de las frecuencias generales a partir del programa SPSS, el estudio de la influencia de la etapa cronológica a través de un test de regresión logística binaria (p. 222), la relación entre los sintagmas interrogativos y el modo verbal empleado, la probabilidad de que cada sintagma interrogativo aparezca con una forma verbal de subjuntivo en la oración subordinada (p. 226) y la distribución geográfica del modo verbal en las interrogativas indirectas negadas con *saber* + *si* en textos americanos y peninsulares (pp. 228-233).

A continuación, Kim Schulte y José Luis Blas Arroyo exponen en «Entrenchment and frequency effects in the diffusion and replacement of modal periphrases in Spanish: a diachronic variationist analysis» (pp. 238-269) la existencia de correlaciones importantes desde el punto de vista estadístico a partir de un análisis de frecuencias de cinco perífrasis modales —*haber de* + infinitivo, *deber* + infinitivo, *deber de* + infinitivo, *tener de* + infinitivo y *tener que* + infinitivo— desde el siglo xvi. El corpus ha sido elaborado por los propios autores a partir de textos de la «inmediatez comunicativa» (p. 247) y se han analizado variables extralingüísticas, fonológicas y fonotácticas, morfológicas y morfosintácticas, sintácticas, semánticas y léxicas.

Seguidamente, en «La posposición pronominal con futuros y condicionales en el código escurialense i.i.6: un examen de varias hipótesis morfosintácticas» (pp. 270-298), Miriam Bouzouita emplea el análisis estadístico para abordar el estudio de la presencia de pronombres átonos en posición posverbal en tiempos verbales de futuro y condicionales sintéticos en el siglo xiii y el porqué de esta posición. El corpus está compuesto por documentaciones procedentes de la biblia E6 y pretende servir de base para examinar tres hipótesis de trabajo: la influencia del entorno sintáctico en la presencia posverbal del pronombre, la correlación entre la forma morfológica del verbo y la posposición pronominal y la posibilidad de que este fenómeno sea una consecuencia de la exposición a la sintaxis pronominal de la *Vulgata*, de modo que esta influya en el orden de la traducción castellana.

La creación de nuevas herramientas para el análisis cualitativo y el estudio paleográfico se combinan en «El nacimiento de la letra jota como grafía consonántica» (pp. 299-321), de María Jesús Torrens Álvarez y Hiroto Ueda, capítulo en el cual los autores emplean el programa estadístico LETRAS para determinar, a partir de 250 documentos del corpus *CORHEN* fechados entre el siglo x y xiii, el momento en que la grafía <j> empieza a emplearse con valor consonántico.

En «El castellano en los orígenes del cambio gramatical: el pretérito imperfecto de la 2.<sup>a</sup> y 3.<sup>a</sup> conjugación (-ié/-ía)» (pp. 322-357), M.<sup>a</sup> del Carmen Moral del Hoyo también parte del *CORHEN* para crear un corpus de 278 documentos burgaleses. A través de un análisis de frecuencias, plantea la interacción de los factores geográficos y estructurales en la variación y evolución de las terminaciones -ié/-ía de imperfecto de indicativo en el siglo xiii.

Por último, en el «Análisis de la sufijación en el corpus *DITECA*» (pp. 358-381), de Inés Carrasco Cantos y Livia Cristina García Aguiar, las autoras realizan un análisis de las frecuencias de los sufijos que dan lugar a derivados nominales, adjetivales y verbales a partir de ordenanzas concejiles con el objetivo de determinar las funciones que estos van adoptando con el paso del tiempo desde el siglo xiii hasta el xviii y los fenómenos de competencia y predominancia de unos sobre otros: -ción/-sión frente a -zón; la alternancia -icia/-ez/-eza y la competencia entre -azgo y -aje.

La cuarta y última sección, denominada «iv. Cuestiones lingüísticas diacrónicas iberorrománicas y lingüística de corpus» (pp. 383-448), es la más breve y recoge estudios que realizan análisis diferentes a los cuantitativos a partir del trabajo con corpus, como los cualitativos, variacionales o idiolectales. El capítulo «Argumentos dialectológicos y sociolingüísticos que ayudan a la caracterización del español en la Nueva España en el siglo xvi» (pp. 385-400), de Beatriz Arias Álvarez y Juan Antonio Hernández Mendoza, se centra en

el estudio de las estructuras de [artículo indefinido + posesivo + sustantivo] y [artículo definido + *cual* + (*dicho*) + sustantivo], así como de la forma gramatical *dende* y las voces *comadre* y *compadre*, a partir del *COREECOM* en la variante del español de la Nueva España en el siglo *xvi* teniendo en cuenta la tradición discursiva y el registro.

Asimismo, en «Manifestaciones de la variación del español colonial en un corpus epistolar multidimensional» (pp. 401-423), Marta Fernández Alcaide toma como corpus de análisis una carta de mediados del siglo *xvi* de 3.274 palabras y estudia desde el punto de vista de la variación el grado de presencia del hablante, la sintaxis segmentada, los problemas en el uso de oraciones relativas, la acumulación de coordinadas, las expresiones indirectas (Searle 1969) y el empleo de construcciones con formas no personales.

«"Se le quedó mirando": la atracción de clíticos en un corpus de idiolectos (ss. *xix-xxi*)» (pp. 424-448), de Olivier Iglesias, constituye el último trabajo de la sección y del volumen y tiene la particularidad de realizar un análisis idiolectal, siguiendo las propuestas de Barra-Jover (2015), con el propósito de examinar el fenómeno conocido como «subida de clíticos» (p. 424) en las construcciones *poder* + infinitivo, *querer* + infinitivo y *deber* + infinitivo y su evolución en los siglos *xix* y *xx*. El corpus lo conforman seis producciones textuales de escritores y blogueras y se analizan las muestras desde el punto de vista cuantitativo para demostrar la validez de la hipótesis idiolectal.

Para concluir, el volumen aquí reseñado constituye un compendio de las últimas propuestas y tendencias en lingüística de corpus de corte diacrónico y ámbito iberorrománico. Por un lado, se exponen los problemas a los que se enfrenta actualmente esta área de estudio en su afán por reconstruir la historia de las lenguas románicas peninsulares, y, por otro, plantea soluciones innovadoras para seguir avanzando en la investigación, así como nuevas reflexiones sobre conceptos y temas tratados previamente en la bibliografía sobre el tema.