



Corbella Diaz, Dolores; Fajardo Aguirre, Alejandro; Langenbacher-Lieb Gott, Jutta (eds.), *Historia del léxico español y Humanidades digitales*, Berlín, Peter Lang, 2018, pp. 502.

La entrada del siglo XXI ha supuesto una revolución en el ámbito de los estudios lingüísticos. La digitalización de muchos recursos y la aparición en la web de tantos otros especialmente preparados para este medio pone a disposición de los usuarios una gran cantidad de datos y le proporciona, al mismo tiempo, una mejora de su capacidad de procesamiento que ha transformado la manera en la que los investigadores se aproximan tanto a estudios sincrónicos como diacrónicos. Esta monografía da a conocer una serie de recursos y proyectos disponibles en el ámbito hispánico para los estudios de tipo histórico, y lo hace desde un punto de vista crítico: es decir, no solo recoge aquellas herramientas con un alto grado de fiabilidad filológica, sino que además presenta sus ventajas e inconvenientes para el estudio, en este caso, del léxico desde una perspectiva diacrónica. Así, este libro constituye un buen resumen de los proyectos de ámbito panhispánico que han ido surgiendo en los últimos años y que se inscriben dentro del marco de las *Humanidades digitales*.

La obra se divide en tres secciones. La primera presenta una serie de proyectos internacionales de gran relevancia actualmente que se dedican a la creación de corpus y otros recursos digitales para el estudio de la lengua desde una perspectiva histórica. La segunda recoge dos capítulos dedicados al análisis crítico de algunos de estos recursos mencionados en la primera parte. La tercera, por otro lado, ejemplifica las posibilidades de las nuevas herramientas con estudios léxicos concretos. Cada capítulo viene acompañado, además, de una extensa bibliografía que recoge no solo los recursos mencionados, sino también diversos trabajos relacionados con la historia del léxico y las Humanidades digitales.

En la primera sección, titulada *Corpus y recursos actuales*, Mar Campos Souto, Mark Davies y Virginia Bertolotti y Concepción Company presentan tres de las herramientas más destacadas que se emplean hoy en día en los estudios diacrónicos: el *Corpus del Nuevo diccionario histórico del español (CDH)* de la Real Academia Española, el *Corpus del español (CE)* y el *Corpus Diacrónico y Diatópico del Español de América (CORDIAM)*, respectivamente. En cada capítulo se presentan los subcorpus de cada uno de ellos, sus fuentes, sus características (lematización, presentación de los documentos, etc.), sus posibilidades de búsqueda y sus posibles aplicaciones para los estudios del léxico. Algo similar hace Esther Hernández en el capítulo dedicado al *Tesoro léxico de los americanismos contenidos en los vocabularios hispano-amerindios coloniales (1550-1800) (TELEAM)*, un proyecto en elaboración que tiene como objetivo recoger los americanismos que se encuentran en los vocabularios bilingües de español y lenguas indígenas elaborados por misioneros durante la época colonial en América. También de tesoros lexicográficos, tanto en castellano (españoles y americanos) como en otras lenguas de la Península, habla

Dolores Corbella, quien hace un breve recorrido por los más destacados del siglo XX y comienzos del XXI, tanto en papel como digitales, y muestra cómo la informática ha modificado la planta de estas obras, permitiendo nuevas opciones de búsqueda y presentación de los datos. Rafael Arnold, Stefan Serafin, Anna-Sussan Franke y Jutta Langenbacher-Liebgott, por su parte, presentan el proyecto de digitalización de los testimonios recogidos para la elaboración del *Diccionario del español medieval* de Bodo Müller, y proponen qué características y posibilidades de búsqueda presentará el resultado de esta digitalización, el futuro *Diccionario del español medieval electrónico (DEMel)*. Otro proyecto de informatización es el que cuenta Pilar García Mouton en un capítulo dedicado a la creación de un atlas virtual interactivo a partir de los datos del *Atlas lingüístico de la Península Ibérica (ALPI)*, algunos de cuyos mapas, como el de *jabalí*, ya se encuentran disponibles en red. Rolf Eberenz, por su parte, analiza las fuentes, macroestructura y microestructura del *Diccionario de la alimentación y la culinaria medievales y renacentistas*. Para cerrar esta sección, José Calvo Tello, Ulrike Henny-Krahmer y Christof Schöch dan a conocer *Textbox*, una colección de corpus literarios históricos etiquetados mediante XML-TEI y accesibles *online*, y muestran, mediante ejemplos concretos, las diversas posibilidades de análisis, fundamentalmente cuantitativos, que estos presentan.

La segunda sección, titulada *Crítica de los recursos en línea: el desorden digital*, recoge dos capítulos dedicados al análisis crítico de los recursos citados en los apartados anteriores. Alejandro Fajardo analiza las ventajas e inconvenientes de algunos corpus, tanto generales (como el *CORDE* o el *CREA*) como regionales o específicos (por ejemplo, el *CORDIAM*, el *CORDEREGRA* o *PostScriptum*), y propone algunas alternativas (como los ficheros digitales, las hemerotecas, las bibliotecas virtuales o los tesoros lexicográficos informatizados) para paliar los problemas que estos presentan para ser usados con fines específicamente lexicográficos. En su misma línea, Francisco Javier Herrero Ruiz de Loizaga, tras un breve repaso de la transición del corpus analógico al digital, expone las dificultades que presentan el *CORDE*, el *CREA* y el *CE* para el estudio de los procesos de gramaticalización en español, dificultades que proceden sobre todo de la mala datación y tratamiento de los textos que los componen y que el autor ilustra con numerosos ejemplos. Estos capítulos aportan, con una perspectiva crítica, una síntesis de las características y la utilidad de los recursos ya presentados previamente en el libro.

La última sección, *Del corpus a los estudios léxicos*, recopila un conjunto de trabajos sobre el léxico desde un punto de vista histórico que ilustran la aplicación de algunos recursos a estudios concretos. Miguel Calderón Campos, en primer lugar, se vale del *Corpus diacrónico del español del reino de Granada (CORDEREGRA)* para estudiar el significado y la distribución geográfica de algunos andalucismos de los siglos XVII y XVIII. Pedro Sánchez-Prieto Borja, Delfina Vázquez Balonga y María Jesús Torrens Álvarez dedican dos capítulos a las fuentes, características y posibilidades para los estudios léxicos del futuro *Atlas Lingüístico Diacrónico e Interactivo de la Comunidad de Madrid (ALDICAM-CM)*. En el cuarto capítulo de esta sección José R. Morala y María Cristina Egido muestran las posibilidades de análisis del léxico cotidiano de los Siglos de Oro en todo el ámbito hispánico que presenta el *Corpus Léxico de Inventarios (CorLexIn)* mediante el estudio de varias voces cuya variación diatópica tiene cierto interés. También en el léxico de la vida cotidiana se centra Miguel Ángel Puche Lorenzo, quien muestra la utilidad de la documentación

de archivo murciana, que se está digitalizando gracias al proyecto *Carmesí*, para estudiar esta parcela del vocabulario y para comprobar cómo esta región ha influido en la configuración del español general a lo largo de la historia. El trabajo de M<sup>a</sup> Ángeles Blanco, Gloria Clavería y Enrique Jiménez es el único de esta sección dedicado a la historia de la lexicografía: los autores presentan una comparación de la macroestructura y microestructura de las distintas ediciones del *Diccionario de la lengua castellana* de la Academia publicadas entre 1817 y 1852 con el objetivo de poner de manifiesto qué cambios se dan, en base a qué fuentes se realizan y a qué circunstancias sociales y políticas responden. Cierra este volumen el trabajo de José Ignacio Pérez Pascual, que analiza cómo la prensa periódica digitalizada, en este caso la que corresponde a la Edad de Plata (1885-1936), puede ser una herramienta muy valiosa para la documentación temprana de algunos términos que aparecen en esta época como neologismos, se instalan en el español general y llegan, en algunos casos, hasta nuestros días.

Así, esta monografía da cobertura a las tres vertientes de la lingüística de corpus que menciona Kabatek (2016: 3) en su *Lingüística de corpus y lingüística histórica iberorrománica*: la que expone de forma crítica la creación y características de los corpus; la que trabaja el tratamiento informático de los datos que estos contienen; y la que estudia fenómenos lingüísticos concretos a partir de las herramientas disponibles. A esto suma la mención de toda una serie de materiales disponibles en red (hemerotecas y bibliotecas digitales, tesoros lexicográficos informatizados, etc.), todo lo cual resulta imprescindible para los estudios de tipo diacrónico. El rápido avance de la informática hace que cualquier publicación sobre este tema, como ocurre con el caso de este libro, resulte inevitablemente una aportación a este campo. El grado de actualización de este trabajo puede observarse en el hecho de que en él se mencionan muchos recursos que no se encuentran disponibles en red todavía, como el *ALDICAM-CM*, el *DEMel* o el *Diccionario de la alimentación y la culinaria medievales y renacentistas*, entre otros. La virtud que tiene, por tanto, no es solo lo novedoso de su contenido, sino el hecho de que presenta al investigador toda una serie de herramientas cuyos avances y modificaciones podrán seguirse, una vez se conozcan, en las propias páginas web de los proyectos que las crearon. Pese a que existen muchos artículos dispersos que describen estos recursos, son pocos los casos en los que estos se encuentran reunidos en un solo volumen como el que nos ocupa; cuando sí lo están, sin embargo, o bien se encuentran en trabajos de tipo más general, no dedicados a la historia del léxico, o bien se recogen en obras algo antiguas y, por tanto, desactualizadas. Unido a esto se encuentra el hecho de que son muy pocos los trabajos que, como este, recogen capítulos dedicados a recursos americanos (como el *CORDIAM*, el *TELEAM*, el *Tesoro lexicográfico del español de Puerto Rico* o el *CE* de Davies) o mencionan herramientas en otras lenguas románicas, aunque luego no exista ningún trabajo concreto en la tercera sección de este libro dedicado al léxico de estas zonas. Esta tercera sección, si bien compuesta por estudios demasiado específicos sobre algunas palabras muy concretas, tiene la virtud, frente a otros trabajos sobre historia del léxico, de presentar nuevos modos de acceso a la información que pueden ayudar a rellenar lagunas existentes en los estudios diacrónicos del vocabulario. En suma, este libro combina la presentación crítica de un conjunto de recursos recientes, con especial atención a sus posibilidades para el estudio del léxico en el ámbito diacrónico, con una serie de aportaciones concretas a la historia del léxico español; y resulta, por

tanto, una prueba más de que, tal y como señalan sus autores en la introducción (p. 11), «la conjunción filología/informática parece ya indisoluble».

### **Referencias bibliográficas**

Kabatek, Johannes (ed.) (2016). *Lingüística de corpus y lingüística histórica iberorrománica*. Berlín: De Gruyter.

Irene Bello Hernández  
Universidad de La Laguna  
irenebellohernandez@gmail.com