
Dialectología digital: innovaciones técnicas y metodológicas

Digital Dialectology: Technical and Methodological Innovations

VERENA WEILAND
Universität Bonn
vweiland@uni-bonn.de

En los últimos tres decenios, la digitalización ha cambiado en gran medida tanto las bases de datos como la forma en que se llevan a cabo los estudios dialectológicos del español. Trasladar la información del espacio impreso al espacio virtual es un logro fundamental, si bien no agota en absoluto las posibilidades de la dialectología digital. Numerosas áreas y propuestas de investigación se derivan actualmente no solo del tratamiento de los datos, sino también de la concepción de las cuestiones planteadas y sus metodologías de análisis.

Los pilares de la dialectología que permanecen en el siglo XXI son los atlas y corpus lingüísticos, así como las monografías individuales. Aproximadamente desde mediados del siglo XX, la cartografía ha desempeñado un papel central en la investigación dialectológica en los campos léxico, fonético y fonológico, morfológico y sintáctico (por ejemplo, el *Atlas lingüístico Diatópico y Diastrático del Uruguay ADDU*, Thun 2000). Más tarde, la digitalización ha facilitado el acceso a los atlas lingüísticos, tal y como queda reflejado, por ejemplo, en el *Atlas Lingüístico-Etnográfico de Colombia (ALEC)* o el *Atlas Lingüístico y Etnográfico de Chile por regiones (ALECh)*. Asimismo, los análisis dialectológicos más centrados en cuestiones locales y menos en la cartografía fina han sido posibles gracias a numerosos corpus del español. Esta característica, sin embargo, no implica su limitación a un solo lugar como se demuestra en corpus de gran tamaño como Val.Es.Co (*Valencia, Español Coloquial*), COSER (*Corpus Oral y Sonoro del Español Rural*), PRESEEA (*Proyecto para el Estudio Sociolingüístico del Español de España y de América*), el *Atlas interactivo de la entonación del español*, FEC (*Fonología del Español Contemporáneo*), FRONTESPO (*Frontera hispano-portuguesa*) o CEMC (*Corpus del Español*

Mexicano Contemporáneo) que cubren áreas geográficas de gran extensión (Moreno Fernández 2005, Pustka *et al.* 2018).

La realidad es que la mayoría de los atlas lingüísticos digitales todavía no ha explotado todo el potencial que tendría el poner a disposición de los usuarios archivos de audio a través de repositorios (de acceso abierto). Esto no solo supondría un enorme beneficio para la fonética y la fonología, sino también para otros campos como muestra, para el italiano, el *Atlante sintattico dell'italiano meridionale (AsiCa)*. Ahora bien, detrás de este tipo de proyectos surgen procesos de extremada complejidad, puesto que no se trata simplemente de escanear los atlas lingüísticos existentes y ponerlos a disposición de forma electrónica, sino también de combinar la cartografía con la posibilidad de reproducir los archivos de audio. En el futuro es deseable, por tanto, abandonar la demarcación fija entre los atlas lingüísticos y los corpus. Esto incluye, por un lado, la explotación y el uso de atlas lingüísticos como corpus y, por otro, la visualización de la variación regional que resulta de estos últimos. Por otro lado, aparece una problemática de exclusión, ya que la cartografía lingüística hasta ahora ha apartado en gran medida al menos dos áreas lingüísticas centrales, la semántica y la pragmática, que son ciertamente de interés desde una perspectiva dialectológica (Colin Salazar 2015).

En cuanto a los actuales corpus del español hablado, todavía se observa una carencia de corpus digitales en línea, que puedan ser creados, ampliados y, si es necesario, corregidos de acuerdo con los principios de la gestión ágil de proyectos (*agile project management*). Asimismo, cabría la posibilidad de desarrollar este tipo de corpus como proyectos de *Citizen Science* con el apoyo de no expertos interesados (véase el proyecto *Dialectos del español*, respectivamente Bouzouita/Castillo Lluch/Pato, en este número, para el español o el *Atlante linguistico digitale dell'Italia e della Svizzera meridionale AdIS* para el italiano). A este respecto, una nueva forma de recopilación y tratamiento de datos (Krefeld/Lücke/Mages 2016) también debería conllevar cambios en los parámetros de estudio y el modo de las anotaciones. Esto se traduciría en la inclusión de factores sociolingüísticos como la movilidad de los hablantes, la migración, el contacto lingüístico y, como aspecto lingüístico-geográfico adicional, la diferenciación entre espacios rurales y urbanos (Britain 2013, Almeida 2014, Klee/Caravedo 2019).

Por consiguiente, el volumen temático *Dialectología digital: innovaciones técnicas y metodológicas* del presente número de la *Revista Internacional de Lingüística Iberoamericana* tiene por objetivo examinar las consecuencias y los próximos trámites de las humanidades digitales y, en particular, las posibilidades de la Web 2.0 en el ámbito de la dialectología. En este sentido, resulta preciso centrarse, en primer lugar, en las nuevas formas de atlas lingüísticos, pilar esencial para la dialectología, y en la cuestión de cómo romper las fronteras entre la cartografía (*mapping*) y los corpus en el espacio digital, en segundo. De este modo, tras considerar los retos generales de la dialectología digital, se utilizarán ejemplos para presentar las formas en las que los datos existentes pueden integrarse en el espacio virtual y los retos que surgen para los nuevos tipos de cartografía lingüística. Por otra parte, en lo que respecta a la recopilación de datos, no nos preocupa tanto la explotación de material producido en redes sociales como

Twitter, puesto que ya es objeto de la sección temática del número 28 de *RILI* (28.2, volumen XIV 2016, *Variación lingüística e internet: cayendo en la red*)¹, sino más bien el planteamiento de cómo debe concebirse la investigación lingüística en el espacio virtual. Con este fin, se plantean diferentes alternativas para la recogida de datos de la lengua hablada, como la aplicación *Dialectos del español*. Además, quedarán expuestas las diferentes posibilidades digitales de anotación de los datos, así como los retos que se plantean en este sentido, especialmente a nivel morfosintáctico (ej.: *COSER-AP*) y fonológico (ej.: *FEC*).

LAS CONTRIBUCIONES DE LA SECCIÓN TEMÁTICA SON LAS SIGUIENTES:

Carolina Julià Luna se dedica a la geografía lingüística del español. Su artículo discute la necesidad de complementar los datos procedentes de la geolingüística tradicional impresa en papel con bases de datos relacionales georreferenciadas que permitan un mejor manejo, explotación y preservación de la información que contienen los mapas lingüísticos. En este marco, se presenta un estudio sobre el cambio léxico-semántico en español europeo a partir de algunos resultados obtenidos del análisis de los datos que contiene la herramienta *CORPAT* (*Corpus de los Atlas Lingüísticos*) para el concepto ‘cantero de pan’ en contraste con resultados obtenidos en una encuesta lanzada a través de las redes sociales. Las conclusiones muestran las ventajas de poder consultar los datos de los atlas regionales del español europeo en una herramienta común para el conocimiento y el estudio de la historia de la lengua.

Verena Weiland, Linda Bäumlér y Carmen Quijada Van den Berghe revelan otra faceta de la dialectología digital: a través de Internet, es posible realizar un test de percepción lingüística en línea, con la participación de hablantes de diferentes lugares del mundo. Las autoras, que trabajan en el campo de la fonética y la fonología del español, eligen el caso del yeísmo (desfonologización de /k/ en favor de /j/) para su estudio. A partir de grabaciones de habla natural, procedentes del corpus *FEC*, examinan si jueces nativos de español pueden distinguir diferentes pronunciaciones de palabras que contienen <ll> e <y> y en qué medida. Los resultados son relevantes para investigar la influencia del origen de los hablantes en el reconocimiento de los alófonos, arrojar luz sobre las razones del cambio lingüístico, y, especialmente importante para el objetivo último de la investigación, para desarrollar sistemas de codificación en corpus fonológicos.

El tercer artículo, escrito por Miriam Bouzouita, Mónica Castillo Lluch y Enrique Pato, presenta la aplicación *Dialectos del español*, centrada principalmente en la obtención de datos lingüísticos que permitan estudiar ciertos fenómenos de variación morfosintáctica del español en el mundo hispanohablante. Los objetivos son tanto

¹ Índice: https://www.iberamericana-vervuert.es/indices/Indexe_R122132.pdf.

la actualización de datos y análisis de los fenómenos morfosintácticos ya estudiados, como la inclusión de estructuras no examinadas hasta ahora. Un aspecto particular de la aplicación es su éxito: en octubre de 2022, más de 735.000 personas habían respondido a las preguntas que se les planteaban sobre el uso de ciertas estructuras morfosintácticas. Además de las respuestas a las cuestiones lingüísticas (p. ej., el complemento del adverbial locativo *enfrente* o la pregunta sobre cómo contestar al teléfono), los participantes indican también su origen con el propósito de crear mapas digitales a partir de los datos. Así el artículo muestra las posibles formas de recogida de datos masivos que propicia la digitalización, y también un aspecto tan esencial como la realización del análisis lingüístico en este contexto. Sin duda, esto marca pautas esenciales para trabajos futuros.

El último artículo destaca los retos técnicos de la creación de corpus digitales anotados y parseados, para la cual se escogió una metodología interdisciplinaria entre la lingüística computacional, la interacción persona-computadora (*Human-Computer Interaction*) y la dialectología. Johnatan E. Bonilla, Miriam Bouzouita y Rosa Lilia Segundo Díaz exponen los principios teóricos y metodológicos de la construcción del *COSE-AP*, cuyo objetivo consiste en anotar morfosintácticamente y parsear los materiales del corpus *COSE* mediante las más recientes tecnologías. Trata en detalle el proceso de anotación de las categorías gramaticales (*PoS tagging*), así como la polifuncionalidad y las fronteras borrosas entre algunas categorías encontradas en los datos dialectales del *COSE*. Esto sirve para facilitar a los lingüistas la exploración de la microvariación diatópica oral de la morfosintaxis del español europeo.

Así, el volumen temático aborda cuestiones centrales de la dialectología digital y plantea, al mismo tiempo, otras nuevas. No pretende resolver todos los retos en este campo, pero sí contribuir al debate actual, desarrollar ideas novedosas e indicar qué aspectos deberían explorarse en el futuro.

REFERENCIAS BIBLIOGRÁFICAS

- ALMEIDA, Manuel (2014): “El concepto de ‘hablas de tránsito’ y el español canario”. En: *Revista de Filología Románica* 31, 1, pp. 37-47.
- BOUZOUITA, Miriam/CASTILLO LLUCH, Mónica/PATO, Enrique (2018): “Dialectos del español. Una nueva aplicación para conocer la variación actual y el cambio en las variedades del español”. En: *Dialectología* 20, pp. 61-83.
- BRITAIN, David (2013): “The Role of Mundane Mobility and Contact in Dialect Death and Dialect Birth”. En: Schreier, Daniel/Hundt, Marianne (eds.): *English as a Contact Language*. Cambridge: Cambridge University Press, pp. 165-181.
- COLIN SALAZAR, Leticia del Carmen (2015): *Usos de ‘hasta’ en el español de México*. Tesis de maestría. Universidad Autónoma de Querétaro.
- FERNÁNDEZ-ORDÓÑEZ, Inés (2016): “Dialectos del español peninsular”. En: Gutiérrez Rexach, Javier (ed.): *Enciclopedia lingüística hispánica*. London/New York: Routledge, vol. 2, pp. 387-404.

- KLEE, Carol A./CARAVEDO, Rocío (2019): “Migration and Orders of Indexicality in Lima”. En: Lynch, Andrew (ed.): *The Routledge Handbook of Spanish in the Global City*. Milton: Routledge, pp. 176-199.
- KREFELD, Thomas/LÜCKE, Stephan/MAGES, Emma (eds.) (2016): “Korpus im Text (2). Zwischen traditioneller Dialektologie und digitaler Geolinguistik: Der Audioatlas siebenbürgisch-sächsischer Dialekte (ASD)”. München: Universitätsbibliothek der Ludwig-Maximilians-Universität, <epub.ub.uni-muenchen.de/25627/1/Korpus-im-Text_Band_2.pdf> (07/06/2021).
- MORENO FERNÁNDEZ, Francisco (2005): “Project for the Sociolinguistic Study of Spanish from Spain and America (PRESEEA) – A Corpus with a Grammar and Discourse Bias”. En: Takagaki, Toshihiro *et al.* (eds.): *Corpus-Based Approaches to Sentence Structures*. Amsterdam: John Benjamins, pp. 265-288.
- PUSTKA, Elissa/GABRIEL, Christoph/MEISENBURG, Trudel/BURKARD, Monja/DZIALLAS, Kristina (2018): “(Inter-)Fonología del Español Contemporáneo/(I)FEC: metodología de un programa de investigación para la fonología de corpus”. En: *Loquens* 5.1. <<https://doi.org/10.3989/loquens.2018.046>> (07/06/2021).
- THUN, Harald (ed.) (2000): *Atlas Lingüístico Diatópico y Diastrático del Uruguay – Norte (AD-DU-Norte)*. Kiel: Westensee.