

HUMANIDADES DIGITALES, CORPUS Y TECNOLOGÍAS DEL LENGUAJE. UNA MIRADA DESDE DIVERSOS CASOS DE ESTUDIO¹

DIGITAL HUMANITIES, CORPUS AND LANGUAGE TECHNOLOGIES:
A LOOK FROM DIVERSE CASE STUDIES

**Andrés GRAJALES RAMÍREZ, Jorge MOLINA MEJÍA
y Pablo VALDIVIA MARTÍN (eds.)**

Groningen: University of Groningen Press / Universidad de Antioquia, 2023, 344 pp.
ISBN: 9789403430249

Esta monografía es tanto en su hechura como en los diferentes capítulos de los que consta un trabajo compilatorio, transdisciplinar y de carácter práctico, que pretende inspirar investigaciones futuras en los ámbitos de las humanidades digitales, la lingüística de corpus y las tecnologías del lenguaje. Se trata de un trabajo coordinado por tres expertos: Andrés Grajales Ramírez, de la Universidad de Antioquia, especialista en lingüística computacional y de corpus, *proofreading* y la enseñanza de español como lengua extranjera; Jorge Molina Mejía, asociado a la misma institución colombiana, donde comparte campos de investigación con su colega y ejerce como profesor de lingüística computacional y de español como lengua extranjera; y Pablo Valdivia Martín, catedrático de la neerlandesa University of Groningen, acreditado también como tal por la ANECA española, y *research associate* en la estadounidense Harvard University, que es además consejero, director académico y miembro de varias instituciones de investigación internacionales. Los tres académicos tienen en común, aparte de la edición de este libro, su participación en el grupo de investigación de la Universidad de Antioquia *Corpus Ex Machina*, coordinado por Jorge Molina Mejía, dedicado al estudio del procesamiento del lenguaje natural y a la lingüística de corpus.

El título de la monografía es muy efectivo en cuanto que se refiere directamente a las disciplinas que abarca y al elemento casuístico de los capítulos —algo no tan habitual en un área como las humanidades digitales, dada a publicaciones de tónica más teórica— y a su vez presenta una claridad que puede llamar la atención del público no especializado.

¹ Libro digital de libre acceso: <https://doi.org/10.21827/646242d096beb> [03/06/2024].

Asimismo, hace visible la pulsión internacional de un libro para cuya publicación unieron fuerzas las universidades de Antioquía y Groningen, y en el que conviven textos en castellano e inglés, firmados por autores y autoras de origen diverso: en total hay contribuciones de especialistas de más de quince centros de educación superior de ocho países de Europa y América.

La monografía principia con un prefacio (pp. 11-13) de Pablo Valdivia Martín y una introducción (pp. 15-30) a cargo de los tres editores. En las primeras páginas queda definido el objetivo del libro respecto a las tres áreas que comprende: “to provide a shared territory where it will be easier to move, get inspired, and move forward together” (p. 11). Además, Valdivia Martín destaca la calidad de la revisión por pares superada por cada una de las participaciones y establece una serie de necesarias definiciones de trabajo para los conceptos centrales de la monografía: “digital humanities”, “corpus studies” y “language technologies” (p. 11). Por último, presenta brevemente las tres partes en las que se agrupan los diferentes artículos del libro, tituladas descriptivamente “digital humanities”, “corpus construction” y “corpus analysis and natural language processing”.

En la introducción se avanza el contenido de cada uno de los capítulos de cada parte. Así, los cuatro primeros textos “versan sobre temas acerca de las humanidades digitales tales como: las artes visuales, las bibliotecas en línea, las bases de datos relacionales para el estudio del griego y el latín clásicos, y el empleo de Python en el análisis epistolar” (p. 25). La segunda parte está integrada por seis capítulos vinculados a la lingüística de corpus que tratan el desarrollo de un corpus para un atlas lingüístico, una propuesta metodológica para el tratamiento de datos multimodales, las primeras fases del proceso de creación del primer corpus paralelo amuzgo-español, la propuesta de una metodología para la construcción de un corpus con el que evaluar la credibilidad a partir de indicadores lingüísticos, el diseño de un corpus específico de alemán a partir de los diálogos de la serie televisiva *Türkisch für Anfänger*, y la presentación del Corpus Colombiano de Aprendices de Inglés, CLEC. La tercera parte incluye aportaciones a la lingüística de corpus “pero desde una perspectiva más relacionada con el análisis y sus métodos, en el que a menudo se valen de la lingüística computacional y el procesamiento del lenguaje natural —PLN—, como también del análisis estadístico” (p. 28), y cuenta con un estudio de las consonantes checas de pronunciación más difícil para hispanohablantes, la propuesta de un modelo estadístico basado en redes bayesianas que tienen por resultado la creación de corpus, una investigación con estadística de corpus sobre la correlación entre la metáfora orientacional *bueno es arriba / malo es abajo* y la polaridad positiva / negativa en verbos del castellano, y la presentación de UnderRL Tagger, un programa patentado que ofrece recursos para el etiquetado de lenguas poco habladas. Esta última colaboración, de la que Jorge Molina Mejía es autor secundario, es la única en la que participa alguno de los tres editores.

En relación con la disposición del libro, se puede apuntar que, precisamente debido a esa interdisciplinariedad que caracteriza la monografía, hay capítulos que encajarían en varias de las partes al mismo tiempo: por ejemplo, el capítulo II, de Jaime Peña Arce y

María Ángeles García Aranda (pp. 55-72), en el que se presenta la Biblioteca Virtual de la Filología Española y se estudia su componente latinoamericano, es constituyente de la sección dedicada a las humanidades digitales, pero también explica el proceso de construcción del corpus base de la BVFE, por lo que debería pertenecer a la parte II, y analiza el archivo latino de la biblioteca, lo que haría que se localizase en la última parte. Por casos como este se considera la separación de los contenidos en tres espacios innecesaria, ya que, debido al enfoque transdisciplinar y fluido del volumen y de las disciplinas de las que se ocupa, la mayoría de las contribuciones se pueden asociar tanto a las humanidades digitales como al estudio o a la creación de corpus mediante tecnologías del lenguaje.

Con todo, se trata de un volumen que en definitiva trabaja en la dirección correcta para cumplir ese ambicioso objetivo de crear un terreno común entre las tres disciplinas que engloba. Gracias a una interesante exposición de metodologías y propuestas de trabajo variadas, esta monografía se convierte además en una valiosa plataforma compilatoria que reúne no solo textos académicos, sino también *corpora*, un programa de *software* libre y otras herramientas de enorme utilidad cuyo conocimiento resulta vital para incrementar la productividad científica de la ya efectiva relación entre las ciencias computacionales y la lingüística y las humanidades.

Roque Fernández Alcañiz
Universidade de Santiago de Compostela



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND).