

# Introducción

MAR CAPILLA MARTÍN  
*Centro de Estudios de la Real Academia Española*  
mcapilla@rae.es



Es para mí un gran honor presentar este monográfico dedicado a los corpus orales, un proyecto que me ha hecho especial ilusión coordinar debido a mi vínculo directo con el tema que aborda. Mi experiencia en el Corpus del Español del Siglo XXI (CORPES XXI), centrada en la selección y codificación de textos escritos y orales, me permite trabajar de cerca con los corpus lingüísticos. Por ello, dedicar este número de *Biblioteca de Babel*, revista del departamento de Filología Española de la Universidad Autónoma de Madrid, a los corpus, y especialmente a los corpus orales, supone un auténtico privilegio.

En el ámbito del español, los corpus lingüísticos constituyen una herramienta fundamental para el análisis de la oralidad, ya que facilitan un estudio sistemático y empírico del uso de la lengua en contextos reales de comunicación. A través de los datos recopilados en los corpus, es posible identificar y examinar fenómenos lingüísticos característicos del discurso oral que, mediante métodos tradicionales, resultarían difíciles de observar y analizar. Considerando la importancia de los corpus en la investigación lingüística y su notable desarrollo en las últimas décadas, resulta razonable que el tema central del monográfico haya sido precisamente el empleo de corpus orales para la investigación del habla coloquial en español. De esta manera, este monográfico reúne un conjunto de investigaciones que abordan, desde perspectivas teóricas, metodológicas y aplicadas, hasta retos y avances en el estudio de corpus orales. Estas contribuciones subrayan la riqueza de los datos empíricos para la investigación lingüística, así como el impacto de las nuevas tecnologías y metodologías en el análisis del discurso oral.

A lo largo de este monográfico, especialistas de diferentes áreas de la lingüística se encargarán de resaltar la relevancia de los corpus orales en las investigaciones lingüísticas centradas en el habla coloquial. Tenemos el honor de incluir una variedad de artículos que abordan enfoques diversos. El volumen comienza con un marco conceptual general y una exposición de las bases metodológicas necesarias para trabajar con corpus. A continuación, se presentan estudios centrados

en el análisis pragmático-discursivo, que reúnen investigaciones sobre el uso de marcadores pragmáticos desde perspectivas variadas, conectando estudios específicos con cuestiones más amplias. Finalmente, el monográfico concluye con una serie de artículos dedicados a corpus orales específicos, que ofrecen una perspectiva transversal e interesante desde distintos enfoques. La sección de artículos de este volumen se inaugura con el trabajo de Renata Enghels, profesora en la Universidad de Gante y codirectora del grupo de investigación CROS (Crossing the Border between English and Spanish), Marlies Jansegers, profesora en la Universidad de Gante y Nele Van Den Driessche, estudiante de doctorado en la misma universidad. Su contribución, titulada «Reflexiones metodológicas y teóricas sobre el análisis de marcadores pragmáticos», se centra en los desafíos teóricos y metodológicos asociados al estudio de estos marcadores. En particular, analiza el caso de *es que* como marcador pragmático polifuncional en el habla madrileña, utilizando una muestra representativa del CORMA (Corpus Oral de Madrid).

A continuación, encontramos el artículo titulado «La transcripción de los elementos prosódicos en corpus de habla coloquial espontánea» presentado por Antonio Hidalgo Navarro, catedrático de Lengua Española en la Universidad de Valencia, y Carlos Castelló Vercher, investigador de la Universidad de Valencia, y ambos miembros del grupo Val.Es.Co. Este estudio aborda la relevancia de la prosodia en las interacciones coloquiales, proponiendo un sistema de transcripción basado en el modelo Val.Es.Co., el cual se apoya en herramientas tecnológicas como ELAN y Praat para identificar fenómenos prosódicos específicos. Otro trabajo basado en el corpus Val.Es.Co. es el propuesto por Sara Badia Climent, profesora de la Universidad de Valencia, y Salvador Pons Bordería, catedrático de la misma institución, ambos integrantes del grupo Val.Es.Co. Con el título «De la transcripción al análisis: desarrollos técnicos del corpus Val.Es.Co. 3.0.», esta investigación revisa los principales corpus orales del español, en especial del corpus Val.Es.Co. 3.0., y profundiza en los avances técnicos para la construcción de corpus orales accesibles digitalmente, poniendo especial énfasis en los retos asociados a la segmentación y el etiquetado automatizado en el procesamiento automatizado de los datos. Por último, Marta Garrote Salazar, profesora de la Universidad Autónoma de Madrid, bajo el título «El etiquetado gramatical automático en el procesamiento del habla coloquial», presenta las dificultades que plantea el etiquetado gramatical, o *part-of-speech tagging*, en los textos de habla coloquial y revisa el estado actual de la investigación en este campo, identificando posibles soluciones o mejorad del etiquetado gramatical en contextos coloquiales.

En relación con los marcadores discursivos, este monográfico reúne diversos estudios que abordan su análisis desde perspectivas

variadas. Dos de las contribuciones se centran específicamente en el estudio de *un poco*. En el artículo titulado «Los corpus orales en la investigación pragmática: el caso de la locución *un poco*», Beatriz Méndez Guerrero, profesora de la Universidad Autónoma de Madrid, examina los usos y las variantes semánticas y pragmáticas de este marcador a partir de datos extraídos de los corpus orales PRESEEA-Palma, Val. Es.Co. y COJEM. Por otro lado, en su trabajo «*Un poco* como recurso pragmático-discursivo en corpus orales de nativos y de aprendientes de español: un estudio comparado», Marta Blanco Domínguez, profesora de la Universidad de Santiago de Compostela, y María Eugenia Conde Noguerol, profesora de la Universidad de La Coruña, realizan un análisis contrastivo del uso de este marcador pragmático entre hablantes nativos y aprendientes de español, basándose en los corpus ESLORA y el Spanish Learner Language Oral Corpora (SPLLOC).

Además, el volumen incluye otras investigaciones relacionadas con los marcadores discursivos. Entre ellas, destaca el estudio de Florencio del Barrio de la Rosa, profesor titular en la Universidad de Venecia Ca' Foscari, titulado «*Ni tan mal*: un operador argumentativo en el español del siglo XXI». Este trabajo caracteriza gramatical y discursivamente dicho operador, apoyándose en corpus electrónicos sincrónicos de lengua hablada, coloquial y subestándar, como el Corpus del español, CORPES XXI y EsTenTen18, para analizar su función y su evolución hacia un marcador conversacional. Asimismo, se presenta el estudio de Mar Capilla Martín, filóloga en el CORPES XXI en el Centro de Estudios de la Real Academia Española, titulado «Entre lo oral y lo escrito: un estudio de corpus sobre los reformuladores *es decir*, *o sea* y *en plan* en español». En este artículo, la autora examina las diferencias contextuales en el uso de dichos reformuladores, destacando nuevos enfoques en el análisis de marcadores discursivos gracias a la integración de datos provenientes de corpus y subcorpus orales.

Seguidamente, se exponen tres artículos que se centran en el análisis de corpus lingüísticos específicos como herramientas para la investigación en el ámbito de la lingüística. El primero, titulado «Un acercamiento al estudio paralingüístico de las emociones en el corpus PRESEEA Valencia», ha sido elaborado por Adrián Cabedo Nebot, catedrático de la Universidad de Valencia, y Noelia Ruano, doctoranda en la misma universidad. Este trabajo examina la influencia de las emociones en el discurso y la prosodia a través del corpus PRESEEA-Valencia, que reúne datos sociolingüísticos del español hablado en Valencia (España).

Por su parte, Carles Navarro-Carrascosa, profesor en el University College Dublin, presenta en su artículo «Descripción y análisis de un corpus de español oral de la comunidad de habla LGTBI» un análisis

del Corpus Oral de la Comunidad de Habla LGTBI. Este corpus recoge muestras de habla propias de dicha comunidad, lo que facilita el estudio de sus códigos lingüísticos y prácticas comunicativas. Además, el autor propone diversas mejoras destinadas a optimizar el corpus para investigaciones más detalladas sobre variación lingüística, influencias socioculturales y la evolución del lenguaje en contextos LGTBI.

La tercera contribución es obra de Carlota de Benito Moreno, profesora de la Universidad Autónoma de Madrid, Antonio Corredor Avelado, investigador posdoctoral en la Universidad de Neuchâtel, y Elena Padrón Castilla, profesora en la Universidad de Neuchâtel, quienes forman parte del proyecto RurICan en la Universidad de Zúrich (UFSP Sprache und Raum). En su artículo «Heterogeneidad e innovación en la isla de La Palma: aproximación sociolingüística y dialectal», los autores exploran diversos fenómenos gramaticales utilizando los datos del corpus RurICan. Este estudio busca describir el grado de heterogeneidad e innovación en las hablas de la isla de La Palma, atendiendo a parámetros tanto sociolingüísticos como geográficos.

Finalmente, Adriana Raya Palmer, profesora en el University College Dublin, presenta su artículo «El lenguaje juvenil en la ficción traducida: una comparación basada en corpus de una selección de marcadores pragmáticos en catalán y español». La autora pone de relieve las particularidades del habla juvenil en las traducciones literarias, explorando las diferencias en la representación de este registro. Además, el estudio investiga la frecuencia y distribución de dichos marcadores en catalán y español, utilizando como base un corpus paralelo de diálogos traducidos de novelas contemporáneas y dos corpus orales, lo que permite una aproximación tanto lingüística como sociocultural.

En definitiva, este monográfico ofrece una mirada amplia y profunda sobre la utilización de los corpus orales para la investigación del habla coloquial, abordado desde diversas perspectivas gracias a la colaboración de expertos en la materia. A lo largo de los capítulos, los lectores encontrarán un análisis riguroso y enriquecedor que no solo desglosa los aspectos clave del tema, sino que también aporta nuevas reflexiones y enfoques. Con la participación de los distintos autores, cuyas contribuciones han sido fundamentales, este trabajo aspira a convertirse en una referencia imprescindible para quienes deseen profundizar en el tema. Confiamos en que este recorrido cumpla las expectativas y sirva de punto de partida para futuros debates y estudios.