

Reseña a Giovanni Parodi (ed.): *Working with Spanish Corpora. Research in Corpus and Discourse.* (Series Editors: Wolfgang Teubert and Michaela Mahlberg). Continuum, 2007. ISBN: 9780826494832, 280 páginas.

La obra editada por Giovanni Parodi se consagra a la variación lingüística medida a través de corpus. Este tema es abordado en diez trabajos en inglés, escritos por diversos autores procedentes de universidades americanas (Argentina, Chile, Venezuela, Estados Unidos) y de una universidad española. El libro, en palabras del propio editor (que es autor o coautor de tres de los diez capítulos) es “the first book published in English which has focused on Spanish variation among diversified corpora and registers” (p. 2), por lo que participa del auge vivido por la llamada *lingüística de corpus* en los últimos años, y orienta el crecimiento de la investigación en dicho campo hacia el terreno de la variación entre registros producidos en lengua española, ámbito que no había sido apenas estudiado por la lingüística de corpus, volcada sobre el inglés y sobre perspectivas contrastivas entre esta lengua y otras lenguas del mundo alejadas del foco occidental.

A continuación, se comentarán los aspectos de interés que pueden tenerse en consideración en cada uno de los capítulos que conforman la obra.

El segundo capítulo (tras la “Introducción” del editor), “Variation across registers in Spanish: Exploring the El Grial PUCV Corpus”, corresponde al propio Giovanni Parodi, quien emplea el citado corpus para identificar rasgos característicos de la lengua de varios registros técnico-científicos. Uno de los principios teóricos básicos (que incide, claro está, en el tipo de metodología empleada) es el de que la variación es multidimensional y también multiparamétrica: “there is a continuum of linguistic variation throughout the registers” (pág. 14). Así, estudia dimensiones como el tipo de modalización, los rasgos de la dimensión informativa, narrativa... mostrando, mediante cuantificaciones y gráficos varios, las grandes diferencias lingüísticas asociadas a los registros hablado y escrito, pero también las zonas funcionales comunes. En este primer artículo se presenta ya un listado de características computables que suscitan algunas interrogantes. Así, el rasgo (3) “Perfect past (indicative and subjunctive)” es descrito en virtud de la oposición peninsular entre aoristo / pasado hodiernal, aun reconociéndose que “vast areas of Spain and Latin America have preferred one form to the other due to prevalence of the perfective aspect in both” (pág. 41), o el subjuntivo (rasgo 10) es calificado como modo de la subjetividad. Es decir, observamos que en la caracterización se tiende a simplificar la amplia virtualidad de empleos que se dan en los textos en favor de un posible rasgo único común que ayude a distinguir lingüísticamente los textos, por ello, aunque posiblemente no llegue a deturpar lo conclusivo de los datos, quedan fuera hechos particulares como el que la ausencia de pasado perfecto se deba a un mero rasgo geolectal o que *Que el gobierno haya subido la gasolina va a a ayudarnos* no es simplemente un indicio de subjetividad.

En el tercer capítulo, “Dimensions of register variation in Spanish”, Douglas Biber y Nicole Tracy-Ventura presentan una aplicación al español sincrónico del análisis multidimensional que desde hace años ha sido puesto en marcha por el primero de los autores:

“This research approach is based on computational analysis of a large text corpus to identify the most important patterns of linguistic co-occurrence, known as the ‘dimensions’. Each dimension comprises a distinct set of co-occurring linguistic features, and each has distinct functional underpinnings. Registers can be compared in

this multi-dimensional space, enabling empirical analysis of both the extent and the ways in which any two registers are different” (pág. 55).

Las conclusiones a las que llegan confirman tendencias de otras lenguas romances y no romances en cuanto a modos de construcción e identificación de registros, pero también revelan datos menos tenidos en cuenta, como la postergación de las cláusulas irreales en registros escritos. Sin duda, es una aportación muy sugerente que podría seguir siendo trabajada en el futuro.

El capítulo IV, “Epistemic modality and academic orality: Pilot study for COTECA (Corpus Textual del español Científico de la Argentina)” de Guiomar Ciaspuscio presenta la exploración de un corpus de enorme interés, constituido por textos científicos en español derivados de producciones medialmente orales y no orales. Si el tipo de corpus constituido no es el convencional, tampoco lo es el tipo de estudio que se realiza, atingente a la implicación del hablante en el discurso mediante procesos de modalización, fundamentalmente marcadores discursivos de modalidad epistémica. El estudio revela cómo hay una intervención muy constante en esta clase de textos por parte del autor para que el lector interprete en un determinado sentido la factualidad de los asertos.

El capítulo V, “Multi-register analysis of prepositional schemes in communication verbs in Spanish”, debido a Omar Sabaj, estudia los objetos preposicionales del español, particularmente aquellos que dependen de verbos de acto verbal, separados en dos grupos a partir de la clasificación de Bosani —aunque en este sentido, trabajos como el de M.J. Rodríguez Espiñeira (2000: «Percepción directa e indirecta en español. Diferencias semánticas y formales», *Verba* 27: 33-85) podían haber ayudado a caracterizar mejor la diversidad interna de estos verbos—. Las conclusiones a las que se llega con este estudio inciden en la continuidad de algunos esquemas entre registros y su tendencia a la variación en otros.

Son muy conocidos en la comunidad investigadora los estudios sobre variación de Mercedes Sedano, autora que es aquí responsable del capítulo VI, “Future tense expressions in several Spanish corpora”; el análisis de la diferencia entre futuro perifrástico y morfológico, al que aquí se dedica la autora, es asunto de un gran interés, cuyo proceso histórico está aún en gran parte por determinar. Este trabajo es una excelente aportación a la materia, ya que se sistematiza qué tipo de aditamentos temporales activan un sentido de futuridad inmediata, cercana o remota que determinará la aparición de futuro perifrástico o de futuro morfológico. Se profundiza también en el peso de los marcadores epistémicos de modalidad a la hora de elegir un tipo de futuro u otro. El rastreo masivo en los corpus permite además a la autora extraer algunas conclusiones sobre el peso geolectal del futuro perifrástico frente al morfológico, siendo aquél más frecuente que éste en Hispanoamérica.

En “Technical-professional discourses: Specialized and dissemination text types” de Giovanni Parodi y Aída Gramajo se presenta un tipo de análisis textual mediante corpus diferente al que se encuentra en el resto de los trabajos del libro, ya que en este artículo se pretende, a partir de un análisis de elementos copresentes en un grupo de textos de tipo técnico, trazar una taxonomía interna de ellos; los resultados confirman la heterogeneidad interna de los constituyentes, aun concentrándose mayor grado de disparidad interna en algunas áreas de contenido que en otras.

El *Corpus 92* que utiliza como base de datos Carmen López Ferrero en “Academic writing: Exploring *Corpus 92*” reúne entre otros textos aquellos provenientes de los exámenes de ingreso en la universidad hechos por alumnos españoles. El análisis de López Ferrero, de foco sintáctico, contrapone el repertorio

(bastante reducido) de nexos paratácticos empleados por los estudiantes y el que emplean productores “expertos” también analizados en el estudio. Como la propia autora afirma, “[a]nalyzing this corpus enables one to determine certain needs and to gain an acquaintance with the actual command with which the Spanish education system aims to provide its students” (191), esto es, se muestra la efectividad práctica de este análisis aplicado, y vuelve a reabrir el debate de por qué la junción sintáctica es un síntoma de elaboración o logro de la distancia comunicativa, siguiendo los conocidos términos de Koch y Oesterreicher.

El trabajo de René Venegas “Using Latent Semantic Analysis in a Spanish research article corpus” se adentra también en un corpus de textos de discurso especializado. En primer lugar, el autor hace una descripción muy exhaustiva y condensada del tipo de discurso que representa el artículo de investigación en una revista científica, aclarando los presupuestos retóricos IMRD que para ellos ya señaló hace años Swales y advirtiéndolo del carácter no discreto de la separación entre los distintos tipos de discurso científico (del más especializado al más divulgativo). Metodológicamente, el trabajo se vale de la herramienta vectorial LSA (*Latent Semantic Analysis*), que mide la similitud semántica entre palabras, frases y párrafos, para verificar dos interesantes hipótesis propuestas para la organización discursivo-textual de los trabajos científicos de dos áreas de conocimiento diversas (la biología y las ciencias sociales):

“H1: When comparing lexical-semantic similarity indexes between text variables (Keywords-abstracts, keywords-contents and abstracts-contents) from scientific research articles of two areas of knowledge, the abstract-contents relation will show higher similarity values than the keywords-contents and keywords-abstracts relations.

H2: There will be significant differences between the two areas of knowledge researched when comparing lexical-semantic similarity indexes between text variables of the scientific research article in a specialized semantic space (keywords-abstracts, keywords-contents and abstract-contents)”.

En efecto, ambas hipótesis se comprueban mediante el análisis LSA, y esta conclusión resulta muy interesante para caracterizar los mecanismos de continuidad topical en la escritura científica, ya que, por una parte, se revela que los resúmenes o *abstracts* de los artículos suponen una condensación, una operación de *macrosemantización*, como la llama el autor, que permite aproximarse al artículo sabiendo que se va a profundizar en la satisfacción de unas hipótesis. En segundo lugar, el hecho de que las palabras-clave no sirvan para fundar también esa operación de macrosemantización confirman que su uso por parte de autores o comités editoriales de revistas es más argumentativo que informativo, ya que se usan como atractores generalistas para conseguir lectores, con independencia de que luego esas mismas palabras sean o no reutilizadas en el cuerpo del texto.

El último capítulo, “Lexical bundles in speech and writing”, de Nicole Tracy-Ventura, Viviana Cortes y Douglas Biber, repite el esquema común a los otros trabajos del libro de fundamentación teórica y análisis práctico en corpus. En este caso, la fundamentación teórica corresponde a la noción de *lexical bundle* empleada por Biber para aludir a secuencias múltiples de palabras que aparecen juntas en los textos y que, por tanto, pueden ser entendidas como “building blocks” a los que se recurre para construir el texto. La investigación de los autores pretende identificar y listar esos complejos léxicos en corpus de escritura académica, por una parte, y de entrevistas sociolingüísticas, por otra, adoptando la convención de que se seleccionen sólo compuestos de cuatro palabras. Los resultados son de gran interés, al presentarse un

paso más en la ya conocida técnica de análisis frecuencial. Así, puede verse todo lo que se refiere al repertorio de topicalizadores más comúnmente empleados, asunto sobre el que, curiosamente, nada se había hecho hasta el momento en español (cfr. el estado de la cuestión que ya describí en Pons Rodríguez 2005: “Los tematizadores en los manuales de ELE”, en *Las gramáticas y los diccionarios en la enseñanza del español como segunda lengua: deseo y realidad. Actas del xv Congreso Internacional de ASELE*. Sevilla: Universidad, 2005, pp.683-692). Además de conferir una nueva dimensión (la del tipo de discurso) al análisis frecuencial en español, el trabajo va un paso más allá y dirige también una mirada contrastiva a los usos semejantes en inglés, extrayéndose también algunas conclusiones de interés respecto al tipo de nominalizaciones que se practican en una y otra lengua.

La obra se cierra con una utilísima lista de corpus y aplicaciones en red para el análisis lingüístico.

En suma, esta obra muestra la fertilidad de la Lingüística de corpus y las muchas posibilidades abiertas para el estudio del español actual. Por otro lado, si el libro que aquí reseñamos atiende a hechos de variación sincrónica, para el futuro queda también abierta la posibilidad de ofrecer más análisis de este tipo aplicados a textos históricos de nuestro idioma. Se podría así conciliar dos áreas de trabajo, con sus debates y cuestionamientos internos, que parecen estar planteándose unos mismos interrogantes. Así, de la preocupación que están despertando estas cuestiones para los historiadores de la lengua da fe el incremento de reuniones científicas que se vertebran en torno a este asunto; así, el *Coloquio Internacional de Corpus Diacrónicos en Lenguas Iberorrománicas* (Palma de Mallorca, octubre de 2007), las *Jornadas sobre Edición de Textos e Historia de la Lengua* celebradas en la Universidad de Sevilla, o la sección al respecto en el *VII Congreso Internacional de Historia de la Lengua Española* (Santiago de Compostela, septiembre de 2009). Lejos de darse un enfrentamiento entre el análisis *filológico* de los datos y su entrada en grandes corpus donde el dato particular es olvidado en favor de la cuantificación numérica global, se puede reclamar la existencia de una nueva filología que atienda ambos frentes, lo cuantitativo y lo cualitativo.

LOLA PONS RODRÍGUEZ
(Universidad de Sevilla)