

CAROLINA VIERA. *Reseña de Alonso-Ramos, Margarita. 2016. Spanish Learner Corpus Research. Current trends and future perspectives.* Amsterdam / Philadelphia: John Benjamins.

1. Introducción

El volumen colectivo que aquí se reseña ha sido coordinado por Margarita Alonso-Ramos (Universidad de Coruña) y está constituido por once capítulos, que discuten diferentes aspectos de un tema del que no abunda información: la creación y utilidad de los corpus de aprendices de español (Alonso Ramos 2016). La Lingüística de Corpus ha despertado el interés de muchos investigadores recientemente, ya que las nuevas tecnologías relacionadas con la informática han permitido el desarrollo de aplicaciones que analizan de manera automática o semiautomática grandes cantidades de textos y, por tanto, permiten la extracción de información y ejemplos que estadísticamente son representativos del género textual o fenómeno lingüístico que se desea estudiar (McEnery y Wilson 2001; Flowerdew 2017). Un corpus consiste en una colección de textos escritos, orales o archivos multimodales de los cuales se puede extraer información estadística sobre las distribuciones de ciertos fenómenos (Gries 2009). Estos archivos/textos cumplen con una serie de requisitos: a) son representativos, b) son auténticos, c) están codificados de manera estandarizada, d) pueden ser procesados por herramientas tecnológicas diseñadas con ese fin y e) suelen estar anotados (McEnery y Hardie 2011; Tono 2016). Este enfoque de investigación basada en análisis de corpus proporciona a los investigadores una rigurosa metodología y las herramientas técnicas necesarias para seleccionar y analizar una colección de textos, tanto escritos como orales, representativos de la población o género textual que se investiga y de los cuales se puede extraer los datos para investigaciones lingüísticas de corte empírico (Parodi 2008). Al igual que otras subdisciplinas de la Lingüística, el campo de Adquisición de Segundas Lenguas se ha beneficiado con este aporte metodológico, que permite al investigador un análisis tanto cuantitativo como cualitativo y, dependiendo del corpus, en diferentes niveles del lenguaje (Granger 2012).

La progresiva especialización de la Lingüística de Corpus ha permitido el surgimiento y desarrollo de la Lingüística de los Corpus de Aprendices (LCA), que se ha ido transformado progresivamente en un campo de estudios en sí mismo (Granger 2012). En LCA, la gran mayoría de los estudios de aprendices de lengua se centran en el idioma inglés (Alonso-Ramos 2016: 4). En este sentido, este volumen significa un gran aporte para quienes se interesan tanto en la Lingüística de Corpus como en el estudio del español como lengua extranjera o segunda lengua. Los autores que integran esta colección son referentes en su campo, lo cual otorga un valor agregado a los trabajos de investigación presentados en este libro.

El libro se estructura en tres secciones. La sección 1 ofrece un panorama del estado actual de la investigación de corpus de aprendices, haciendo hincapié en los problemas y desafíos del campo. La sección 2 está integrada por cinco manuscritos, que describen aspectos técnicos relacionados con el diseño y creación de corpus de aprendices, así

como su anotación y explotación. Por último, la sección 3 está constituida por cuatro capítulos, que se centran en el análisis de datos obtenidos a partir de corpora. Esta estructuración permite al lector una aproximación gradual, que combina el proceso que conlleva la creación y el uso de un corpus, a la vez que facilita el acceso a la información específica.

2. Descripción de capítulos

El capítulo 1, “Spanish learner corpus research: Achievements and challenges”, escrito por Margarita Alonso-Ramos, la compiladora del libro, ofrece una revisión histórica del incipiente campo de los estudios de corpus de aprendices, en especial en lo que atañe a idiomas que no sean el inglés. De hecho, la autora señala que no ha sido hasta 2014, cuando una institución oficial se abocó a la tarea de compilar un corpus de aprendices de español (Alonso-Ramos 2016: 3). La autora presenta información relevante tanto de la situación del español como lengua de estudio, como de los desafíos a la hora de diseñar y analizar corpus de aprendices de español. Estos problemas se agudizan si se considera la creación de corpus de español hablado y corpus multimodales. De igual interés, en este capítulo se plantea un tema de controversia, que concierne a los estudios de corpus de aprendices. En general, este tipo de corpus se crean para analizar el nivel de desarrollo de la lengua estudiada. El problema principal radica en que para poder medir ese desarrollo se debe establecer una norma comparativa que obre como referente, esto es, una meta de ese desarrollo. Alonso-Ramos resume la discusión que existe entre los profesionales dedicados a las segundas lenguas, en relación con la idealización del hablante nativo y la pertinencia del concepto del “error” y cómo estos conceptos teóricos han afectado la anotación y etiquetado de errores en los corpus de aprendices. Concluye el capítulo con una revisión de los diferentes corpus de español existentes, lo que representa una importante contribución para quién esté interesado en investigar aspectos de la lengua en los que este enfoque metodológico sea clave.

El segundo capítulo que integra la primera sección de este libro, escrito por Yukio Tono y titulado “What is missing in learner corpus design?”, complementa el capítulo anterior, centrándose en las brechas de conocimiento que existen en este campo de estudios. La revisión bibliográfica de Tono es un excelente material de consulta para quienes se inicien en la lingüística de corpus de aprendices, puesto que la autora cubre los conceptos básicos que deben ser tenidos en cuenta a la hora de la creación de un corpus, como los son la representatividad, autenticidad y la recolección de datos. Uno de los principales fallos de los estudios de corpus es justamente su diseño y es precisamente por este motivo por lo que este capítulo cumple una función fundamental para quienes se interesen en la construcción de un corpus con propósitos de investigación. El capítulo concluye con recomendaciones para futuros estudios en lo que respecta al diseño y a un análisis que tenga en cuenta la variación sociolingüística y contextual, que afecte a los textos que se compilan en un corpus.

El capítulo 3, escrito por Guillermo Rojo e Ignacio M. Palacios, se titula “Learner Spanish on computer: the CAES ‘Corpus de Aprendices de español’ project”. Comienza la sección 2 del libro, llamada “Compilación, anotación y explotación de los datos de un corpus de aprendices” (en inglés en el original). La elección del capítulo de Rojo y Palacios es acertada, ya que se presenta la construcción del Corpus de Aprendices de Español (CAES), al que Alonso-Ramos (2016) considera como la primera empresa oficial relacionada con corpus de aprendices. Rojo ha estado involucrado con la

creación de corpus de referencia del español moderno e histórico y, por tanto, conoce a fondo los principios básicos del diseño de un corpus y los posibles problemas que pueden surgir en su creación. El capítulo 3 describe el proceso de creación del corpus de textos escritos (CAES), posibles usos del mismo y postula la importancia que los estudios basados en corpus pueden tener en el entendimiento de procesos lingüísticos y en la enseñanza de lenguas extranjeras. Asimismo, se discuten las limitaciones de los estudios de corpus y se advierte de posibles problemas asociados con el análisis no riguroso de los datos que se obtienen. Rojo y Palacios ofrecen una sólida bibliografía de referencia al realizar su revisión de la contribución de la Lingüística de Corpus, revisión que incluye autores no citados en los capítulos anteriores y que han sido de gran importancia para este campo de estudio. Otra contribución es la diferenciación que Rojo y Palacios detallan entre los corpóra de aprendices y los de aprendices de español. De igual manera, se reseñan otros tres proyectos similares al CAES. Entre ellos, un corpus oral de aprendices, el SPLLOC. En suma, este es un capítulo de gran interés por cuanto se detalla (2) el proceso de creación de un corpus de aprendices de español de gran magnitud como es el CAES, con 570.000 palabras, (2) LA ejemplificación de su aplicación a la investigación y a la enseñanza y (3) la discusión de las diferentes decisiones metodológicas relacionadas con la anotación, etiquetado y las herramientas de análisis usadas.

El Capítulo 4, escrito por Campillos Llanos se titula “PoS-tagging a Spanish oral learner corpus: Criteria, procedure, and a sample analysis”. Como su título indica, este capítulo se enfoca en el proceso de etiquetado y análisis de un corpus oral de aprendices. Una de las principales dificultades del análisis morfológico y sintáctico de corpus orales de aprendices consiste en lo demandante del proceso de su etiquetación. Existen herramientas automáticas de segmentación (parsing) y etiquetado (tagging), pero estas no han resultado eficaces para el procesamiento de corpus de aprendices. Campillos Llanos informa del éxito en el etiquetado del corpus oral CORELE (Corpus Oral de Español como Lengua Extranjera) con el etiquetador GRAMPAL (Moreno y Guirao 2006) y demuestra su utilización presentando un estudio que analiza estadísticamente el uso de los artículos definidos e indefinidos por parte de los aprendices que están representados en corpus. El estudio tiene como marco teórico el análisis contrastivo de interlenguaje (Granger 1996) y, Campillos Llanos considera que este enfoque teórico permite un acercamiento cuantitativo al estudio de las capacidades de los aprendices de español. Asimismo, el artículo de Campillos Llanos ofrece una interesante revisión bibliográfica sobre este la idoneidad del análisis contrastivo de interlenguaje en los estudios de corpus de aprendices, así como del uso de los artículos por parte de los aprendices. En suma, este capítulo ofrece una detallada descripción de los criterios y métodos de etiquetado y análisis estadístico del corpus oral de aprendices CORELE en el marco teórico del análisis contrastivo de interlenguaje.

El capítulo 5, “The LANGSNAP longitudinal learner corpus”, escrito por Tracy-Ventura, Mitchel y McManus, describe el diseño y uso de un corpus que es novedoso por dos razones importantes: representa un estudio longitudinal en un programa de estudios en el extranjero. Existe muy poca información en lo que refiere a estas dos áreas de estudio, especialmente, en lo que compete a la Lingüística de Corpus y, en ese sentido, el capítulo de Tracy-Ventura et al. es de sumo interés. En especial, porque los anteriores estudios en el campo de la Adquisición de Segunda Lengua en estudios en el extranjero han dejado claro que es necesario profundizar la investigación en estos

contextos (Ventura et al. 2016). Los autores de este capítulo presentan el diseño de un estudio que se desarrolló en el curso de 20 meses con 27 participantes, que cumplieron tres tareas diferentes: narración, descripción y escritura de ensayo argumentativo. Otro aspecto interesante de este estudio es que se recolectó tanto datos escritos como orales. La mayoría de los estudios de corpus compilan textos escritos y, en este sentido, este estudio es relevante para aquellos investigadores interesados en el desarrollo de la conocimiento y capacidad de expresión oral de los aprendices. Los investigadores buscaban analizar el desarrollo de la variedad léxica en los textos y en la realización de las diferentes actividades. De interés es la discusión de la unidad de análisis usada para medir la variedad léxica: el “D score” (Tracy-Ventura et al. 2016: 131) o índice parámetro D (Capsada Blanch y Torruella Casañas 2017: 361). Si bien los resultados del estudio que se presenta son de carácter preliminar, este capítulo ofrece un sólido aporte, que combina el campo de SLA con la Lingüística de Corpus, en lo que respecta al diseño de un corpus longitudinal, con muestras orales, de aprendices que se encuentran realizando estudios en el extranjero.

El capítulo 6, “The Aprescrilov corpus, or broadening the horizon of Spanish language learning in Flanders”, escrito por Kris Buyse, Lydia Fernández Pereda y Katrien Verveckken, describe otro corpus de aprendices, en este caso, de más de 2700 textos escritos por estudiantes universitarios de español de nivel básico hasta avanzado en Bélgica y Holanda. El corpus está organizado por diferentes tareas (exámenes y trabajos de curso). Los autores afirman que el estudio que presentan permite ver las ventajas del uso de corpus, y de las herramientas de la Lingüística de Corpus en la investigación de adquisición de segunda lengua o una lengua extranjera. También resulta de interés, el hecho de que se presente un estudio longitudinal dentro del marco teórico de Análisis de Error (Corder 1981) y el Interlenguaje (Granger 1996). Los estudios que se discuten en este capítulo analizan el aprendizaje del uso de los verbos de cambio de estado en español y de la noción de género y concordancia por parte de aprendices que recibieron una instrucción explícita sobre este tema. Asimismo, se realizó un análisis cuantitativo de los tipos de errores más frecuentes en el corpus Aprescrilov. Los autores concluyen que los datos extraídos en esta investigación cuestionan el énfasis que se da en la enseñanza del español a aspectos del lenguaje que no aparecen como los más frecuentes en sus resultados, al menos en lo que concierne a textos escritos representados en su corpus. Por ende, los autores consideran que la enseñanza de lenguas extranjeras o segundas lenguas debe tener en cuenta los estudios de corpus. De igual manera, plantean las limitaciones que tiene este tipo de estudios y la necesidad de tener en cuenta dichas limitaciones. Este capítulo, por tanto, es de interés para quienes busquen aplicaciones de los estudios de corpus al campo de la enseñanza del español y un mejor entendimiento del interlenguaje de los aprendices, basado en estudios empíricos de una colección amplia de textos escritos.

El capítulo 7, escrito por Dale Koike y Jennifer Witte y titulado “Spanish Corpus Proficiency Level Training Website and Corpus: An open-source, online resource for corpus linguistics studies”, describe un corpus compuesto por 327 segmentos videograbados de 38 aprendices de los cuales 17 son hablantes de español como lengua de herencia. Como en el caso de otros artículos publicados en este volumen, las autoras ofrecen una completa revisión bibliográfica del estado de la cuestión en la Lingüística de Corpus de aprendices, centrándose en los conceptos de validez y representatividad. La discusión presentada por las autoras es pertinente, puesto que el corpus que discuten

es atípico en varios aspectos, concretamente, por el hecho de que no ha seguido métodos estandarizados de transcripción, anotación y etiquetado, ni incluye herramientas de búsqueda automática como programas de concordancia. Esto se debe a que el propósito original del sitio web en donde se encuentran alojadas las muestras de habla era el de entrenamiento de docentes y no el de investigación en el marco de la LC. A pesar de esto, las autoras defienden la viabilidad de este corpus para investigaciones sobre la adquisición de una segunda lengua o sobre las posibles diferencias dialectales y, para ello, presentan un estudio de expresión del sujeto pronominal en el marco teórico de la Lingüística Variacionista, realizado a partir de las muestras de habla que comprenden su corpus. Las autoras concluyen que los datos resultantes de este estudio son válidos y, por lo tanto, este corpus permite su investigación. Asimismo, como este corpus está integrado por videograbaciones, las autoras destacan una contribución potencial de este corpus es su aplicación a investigación de aspectos paralingüísticos poco estudiados en SLA.

La tercera sección del libro, que se centra en el análisis de los datos obtenidos a partir de un corpus, comienza con el capítulo 8, escrito por Ana Blanco Canales y titulado “Factors that can have an impact on the processes of perceiving Spanish/L2”. La propuesta de Blanco Canales es muy interesante, puesto que detalla las ventajas de un enfoque basado en la Lingüística de Corpus, con el fin de determinar la relación entre percepción y producción de elementos fónicos y suprasegmentales (competencia fónica) en el proceso de adquisición de una lengua no materna. Para ello, se presenta un estudio cuantitativo y cualitativo de percepción, aplicado a un total de 300 estudiantes – reducidos posteriormente a 209– procedentes de 15 países diferentes. La autora también describe el corpus de referencia *Fono.ele*, que usó en su estudio compuesto de 96 individuos de 6 nacionalidades diferentes. El estudio de Blanco Canales permite ver la manera en que un corpus puede ser diseñado y analizado cuantitativa y estadísticamente para arrojar datos fidedignos, que explican el proceso de adquisición de los diferentes elementos del lenguaje por parte de diferentes poblaciones de aprendices de español. En particular, se centra en la competencia fónica de estos aprendices y en las variables sociolingüísticas que afectan la percepción de los sonidos de la lengua extranjera y es, por tanto, una contribución a un área de estudios en donde tradicionalmente la obtención de datos se limitaba a un menor número de participantes, lo que llevaba a cuestionar la representatividad de los resultados. Otro de los aportes de este trabajo está dado por las herramientas de análisis diseñadas con el fin de obtener los datos, en particular, la discusión del diseño e implementación del test de percepción usado en este estudio y el detallado análisis de resultados, que serán de interés por la escasez de instrumentos de investigación en el área de la competencia fónica.

El capítulo 9, escrito por Cristóbal Lozano y titulado “Pragmatic principles in anaphora resolution at the syntax-discourse interface: Advanced English learners of Spanish in the CEDEL2 corpus”, discute el nivel sintáctico y discursivo de la lengua en aprendices de español, en concreto, en lo que se refiere a la resolución de la anáfora. Además de información pertinente a otro corpus de aprendices de español (CEDEL2), Lozano informa sobre una herramienta de anotación de corpus, diseñada con el propósito de anotar los factores sintácticos, gramaticales y discursivos que inciden en dicha resolución. El aporte de Lozano radica, por un lado, en estudiar un tema que ha despertado gran interés, como es el de la resolución de anáforas, usando las herramientas y metodología de la Lingüística de Corpus, y, por otro lado, en demostrar

que se puede contribuir con datos nuevos, que, en parte, complementan así como desafían las teorías existentes. CEDEL2, el corpus en el que se basa la investigación de Lozano consiste en textos escritos, que se recolectan en línea y constituyen exámenes de nivelación, biografía y composiciones de más de 1500 participantes. El estudio de Lozano comprende 20 textos y compara la producción de hablantes nativos con la de aprendices. Lozano se enfoca en la tercera persona (humana) por motivos que explicita en su artículo. El etiquetado se realizó con UAMCorpus Tool (O'Donnell 2009) y las etiquetas se basaron en investigaciones previas de resolución de anáfora. Lozano realiza un análisis cuantitativo en base a porcentajes y otro de corte cualitativo. Estos análisis combinados le llevan a proponer una nueva teoría, que explica los factores que intervienen en la resolución de anáfora y que este autor llama "Pragmatics Principle Violation Hypothesis".

El capítulo 10, escrito por Nancy Vázquez Veiga, se titula "Discourse markers in CEDEL2 and SPLLOC corpora of learner Spanish: Analysis of some lexical-pragmatic failures". Es un estudio cuantitativo y cualitativo del uso de los marcadores de discurso (MDs) por parte de los aprendices de español. En particular, se centra en el uso de los siguientes MDs: *oh*, *actualmente*, *eventualmente*, *usualmente* y *entonces* y los MDs que contienen la forma "fin", tales como *por fin* o *al fin*. La autora trabaja con datos extraídos de dos corpora discutidos en capítulos anteriores, el corpus de textos escritos CEDEL2 (Lozano 2009) y el corpus de textos orales SPLLOC (Mitchell et al. 2008), a los cuales escoge por estar integrados por aprendices de español de nivel intermedio y avanzado, cuya primera lengua es el inglés y porque cuentan con un corpus paralelo de hablantes nativos. El uso de estos dos corpora combinados le permite a la autora presentar una serie de datos y análisis cualitativos y concluir que hay fallas a nivel pragmático y léxico en el uso de MDs por parte de los aprendices, y que esto representa un problema de comunicación importante, especialmente en aquellos MDs que contienen la forma "fin".

El último capítulo del volumen tiene como autores a Orsolya Vincze, Marcos García-Salido, Ana Orol and Margarita Alonso-Ramos y se titula "A corpus study of Spanish as a Foreign Language learners' collocation production". El estudio compara la producción de colocaciones de aprendices y hablantes nativos, y para esto usan los textos escritos compilados en el corpus CEDEL2. En este estudio, los autores investigan un área poco investigada en el estudio de colocaciones: la descripción detallada y la clasificación del tipo de errores (taxonomía) que los estudiantes realizan en el proceso de combinar elementos léxicos en patrones determinados por las características de estos elementos. Cien textos de nivel intermedio y avanzado y ciento tres textos de hablantes nativos fueron anotados manualmente. El capítulo ofrece amplia información de las decisiones metodológicas tomadas durante la recolección de datos y de la anotación del corpus, así como de los resultados del estudio que se presentan siguiendo un análisis cualitativo y cuantitativo para ejemplificar su uso. Sin embargo, las autoras concluyen que las dimensiones del corpus estudiado fueron un factor limitante a la hora de obtener resultados más concluyentes. En suma, este capítulo obra como cierre a una colección de estudios que ilustran los diferentes temas del lenguaje, que pueden ser estudiados con el enfoque de la Lingüística de Corpus. Al igual que otros capítulos, existe una interesante descripción del proceso de investigación en un área nueva de estudios, por lo que estos capítulos ofrecen un camino a seguir por parte de otros investigadores.

3. Conclusiones

Los diferentes capítulos que integran este volumen presentan investigaciones sólidas, que avalan la idea de que el análisis de corpus y, en especial, de corpóra especializados en aprendices de lengua, permite realizar análisis cuantitativos y cualitativos a una colección de textos representantes del género y población que sea de interés. Es un libro fundamental para quien se interese por las investigaciones de diferentes aspectos que atañen a la adquisición de la lengua en diferentes niveles del lenguaje, porque los capítulos que se incluyen discuten temas variados a nivel discursivo, léxico, gramatical y fónico. Otra contribución es que se estudian diferentes poblaciones de estudiantes de lengua y en diferentes contextos, aunque priman las investigaciones de hablantes universitarios. Asimismo, ofrece una revisión muy completa de la situación de la Lingüística de corpus de aprendices y su importancia en la investigación empírica de fenómenos lingüísticos relacionados con los procesos de aprendizaje de una lengua. Los autores que participan en este volumen incluyen descripciones de decisiones metodológicas que pueden ser replicadas por otros investigadores, lo cual representa un claro aporte a un campo de estudio incipiente.

El volumen se encuentra bien balanceado en lo que refiere al tipo de corpus y temas lingüísticos que generan mayor interés. Una observación es que, si bien el volumen está integrado por investigadores que trabajan en Europa y Estados Unidos, no se incluye información ni referencia a la Lingüística de corpus que se realiza en Latinoamérica ni a corpus académicos como el Corpus Académico y Profesional del Español PUCV-2006 (Parodi 2009). Esta es probablemente la única deficiencia de este volumen editado.

Finalmente, los investigadores que integran este volumen hacen un llamado de la necesidad de una mayor sistematización de los pasos metodológicos necesarios para el diseño de corpus representativos y válidos. De igual manera, la mayoría de los investigadores discuten los problemas existentes y las limitaciones en la investigación de corpus. En este sentido, el volumen también realiza una contribución al plantear líneas de investigación futura, así como de un mejor desarrollo profesional de esta área de estudios.

En suma, el presente volumen representa una lectura obligada para aquellos interesados en la investigación empírica del lenguaje natural producido por aprendices de español y en la Lingüística de Corpus en general.

Referencias bibliográficas

- Alonso-Ramos, Margarita. 2016. Spanish learner corpus research: Achievements and challenges. En M. Alonso Ramos, ed. *Spanish learner corpus research: Current trends and future perspectives*. Amsterdam / Philadelphia: John Benjamins
- Capsada Blanch, Ramón; Joan Torruella Casañas. 2017. Métodos para medir la riqueza léxica de los textos: Revisión y propuesta. *Verba* 44: 347-408
- Corder, Stephen. 1981. *Error Analysis and Interlanguage*. Oxford: OUP
- Flowerdew, John. 2017. Corpus-based approaches to language description for specialized academic writing. *Teach* 50.1: 90–106
- Granger, Sylvie. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. En K. Aijmer, B. Altenberg y M.

- Johansson, eds. *Language in Contrast*. Lund: Lund University Press, pp. 37-51
- Granger, Sylvie. 2012. How to use foreign and second language learner corpora. En A. Mackey & S. Gass, eds. *Research Methods in Second Language Acquisition: A Practical Guide*. Malden: Blackwell, pp. 7-29
- Gries, Stefan Th. 2009. What is Corpus Linguistics? *Language and Linguistics Compass* 3: 1-17
- Lozano, Cristobal. 2016. Pragmatic principles in anaphora resolution at the syntax-discourse interface: Advanced English learners of Spanish in the CEDEL2 corpus". En M. Alonso Ramos, ed. *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam / Philadelphia: John Benjamins, pp. 235-266
- McEnery, T., & Wilson, A. 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press
- McEnery, T., & Hardie, A. 2011. What is corpus linguistics? En T. McEnery y A. Hardie, eds. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, pp. 1-24
- Mitchell, R., Domínguez, L., Arche, M., Myles, F., & Marsden, E. 2008. SPLLOC: A new database for Spanish second language acquisition research. *EuroSLA Yearbook*, 8.1: 287-304
- Moreno, Antonio y Guirao, José María. 2006. Morpho-syntactic tagging of the Spanish C-ORAL-ROM Corpus. En Y. Kawaguchi, S. Zaima y T. Tagakaki, eds. *Spoken Language Corpus and Linguistic Informatics (Usage Based Linguistic Informatics 5)*, Amsterdam / Philadelphia: John Benjamins, pp. 199-218
- Parodi, Giovanni. 2008. Lingüística de corpus: una introducción al ámbito. *Revista de Lingüística Teórica y Aplicada* 46.1: 93-119
- Parodi, Giovanni. 2009. El Corpus Académico y Profesional del Español PUCV-2006: semejanzas y diferencias entre los géneros académicos y profesionales. *Estudios Filológicos* 44: 123-147
- Tono, Yukio. 2016. What is missing in learner corpus design? En M. Alonso Ramos, ed. *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam / Philadelphia: John Benjamins, pp. 33-54
- Tracy-Ventura, Nicole; Mitchel, Rosamond; McManus, Kevin. 2016. Spanish learner corpus research: Achievements and challenges. En M. Alonso Ramos, ed. *Spanish Learner Corpus Research: Current trends and future perspectives*. Amsterdam / Philadelphia: John Benjamins