

Gorana Zečević Krneta
Universidad de Kragujevac
Serbia

Una aproximación al estudio del corpus de aprendices serbios de E/LE

Recibido 5 de marzo de 2013 / Aceptado 22 de mayo de 2013

Resumen: El estudio del corpus de aprendices se sustenta en fundamentos de la lingüística del corpus, las teorías de la adquisición de idiomas y la enseñanza de las lenguas extranjeras, y así ofrece numerosas posibilidades de la labor didáctico-pedagógica. Sin embargo, lo más exigente del estudio de corpus es la creación del mismo. Con el presente trabajo abrimos la cuestión de la elaboración de un corpus de aprendices y presentamos uno de ellos, el corpus de aprendices serbios de E/LE. Asimismo, desarrollamos una visión crítica que capacite al personal docente para saber cuándo un determinado corpus de este tipo puede resultar útil, por qué y para qué. Este trabajo lo cerramos con una agenda de prioridades dado que nuestra tarea de elaboración de un corpus de aprendices lo suficientemente representativo y asequible todavía no ha conseguido la fase pretendida.

Palabras clave: adquisición de lenguas extranjeras, E/LE, estudios del corpus de aprendices, implicaciones pedagógicas, lingüística computacional.

Abstract: Due to the fact that learner corpus research relies on findings from corpus linguistics,

linguistic theory, second language acquisition, and foreign language teaching, it is commonly claimed that language corpora have many uses in terms of relevance for language teaching and learning. However, corpus elaboration is a very difficult and demanding part of this methodological approach. In this paper, special attention is given to learner corpora and, more specifically, to its elaboration. The learner corpus in the case of Spanish as a Foreign Language has been presented, along with the implications of its use. However, there is a need for more systematic studies on the effectiveness of learner corpora in language teaching. The article closes with a list of priorities for further work on an electronic database of learner corpus in the case of Spanish as a Foreign Language.

Key words: computational linguistics, learner corpus research, pedagogical implications, second language acquisition, Spanish as a Foreign Language.

1. Introducción

Se ha destacado y demostrado varias veces que la lingüística del corpus ofrece una gran cantidad de posibilidades de investigación, dado que el estudio lingüístico de corpus permite el manejo de una cantidad enorme de datos, contextualiza ejemplos y los interpreta de acuerdo con una serie de variables de naturaleza lingüística y social (Biber 2012; Buckingham 2009; Granger 2009; Adolphs 2008; Leistyna, Meier 2003). El corpus es una

colección de textos, orales o escritos, en formato electrónico o no, pero, al mismo tiempo, una fuente de ejemplos tanto realistas y típicos como excepcionales para diferentes objetivos lingüísticos y pedagógicos (Sánchez 2000: 7).

Si se toman en consideración los trabajos realizados hasta ahora dentro del ámbito del estudio del corpus, se pueden citar tres focos de interés: (a) el empleo del corpus en actividades relacionadas con los estudios de traducción, la estilística, los estudios gramaticales y la construcción de gramáticas y diccionarios; (b) un desarrollo de la asociación del corpus y docencia, que consiste en la compilación y empleo de corpus pedagógicos con objeto de descubrir aspectos importantes relacionados con el proceso de enseñanza-aprendizaje de una lengua extranjera; (c) el empleo de concordancias y otros datos extraídos de un corpus con el fin de utilizarlos en diversas tareas docentes (Fuertes Olivera 2007: 214). La utilidad de los análisis de los tres focos se argumentará luego en el artículo.

Sobre la aplicación de los estudios del corpus en el ámbito pedagógico ya se ha escrito en muchas ocasiones (Biber 2012; Zecevic Krneta 2010; Belz, Vyatkina 2008; O'Keeffe, McCarthy, Carter 2007; Sinclair 2004; Granger, Hung, Petch-Tyson 2002; Hunston 2002; Osborne 1999). Por un lado, los corpus facilitan a los aprendices el contacto con el lenguaje auténtico, les ayudan a formular sus propias conclusiones sobre aspectos del uso y significado de palabras, colocaciones y otras unidades fraseológicas; por otro lado, permiten conocer los mecanismos de la adquisición de idiomas y hacen concienciar a los interesados acerca del proceso de enseñanza-aprendizaje de una LE. Barnbrook (1996: 24) añade que un corpus representa una muestra de colección de datos lingüísticos que permite que se deduzcan conclusiones sobre la muestra más que sobre la lengua misma, por lo que cualquier estudio de textos de corpus debería cuestionar la muestra lingüística, y ajustarla a los objetivos que se intenten lograr.

Pues bien, claro está que el estudio del corpus es una tarea fructífera y beneficiosa tanto para el personal docente como para los aprendices; sin embargo, lo difícil y exigente del estudio del corpus es la creación del mismo. Justamente, en esta ocasión queríamos abrir esa cuestión de elaboración de un corpus de aprendices y presentar uno de ellos, el que nosotros hemos logrado crear hasta ahora. Asimismo, queríamos desarrollar una visión crítica que capacite al personal docente para saber cuándo un determinado corpus de este tipo puede resultar útil, por qué y para qué.

En nuestro caso se trata de un corpus de aprendices del español como lengua extranjera (E/LE) que incluye textos descriptivos escritos por los estudiantes de español de los cuatro cursos reconocidos oficialmente, y el corpus se está elaborando en el Departamento de Hispanística de la Universidad de Kragujevac, con la ayuda de la Facultad de Ciencias Naturales y Matemáticas de la Universidad de Belgrado. El impulso hacia la creación de un corpus de aprendices serbios de E/LE, como suele ser, ha partido de la necesidad de elaboración de una base de datos para la tesis doctoral, pero la idea es que su rentabilidad no termine con eso, ya que son muchas las oportunidades de su uso y provecho didáctico.

Finalmente, nuestra intención ha sido ofrecer a todos los interesados este corpus codificado e informatizado, pero, lamentablemente, como esta tarea costosa en tiempo y esfuerzo no depende en totalidad de la autora, esperamos en un futuro próximo poder dar toda la información necesaria para manejar el corpus. Por el momento, el corpus está en el formato de *Word*, con la posibilidad de manejarlo con herramientas de *Word*¹.

¹ La idea es que el corpus se encuentre a disposición de todos los interesados en una web página, la que se podría consultar bajo la autorización de la autora.

2. Sobre el corpus de aprendices de idiomas

Antes de describir el corpus de aprendices serbios de E/LE, vamos a definir qué entendemos por corpus de aprendices y cómo se destaca este entre otros tipos de corpus.

De acuerdo con los trabajos realizados hasta ahora dentro del campo de la lingüística del corpus, Johansson (2009: 36) nombra tres tipos de corpus: corpus textuales de un nativo (ing. *native textbook corpora*), corpus de aprendices de una LE (ing. *learner corpora*), y corpus paralelo multilingüe (ing. *multilingual corpora*). Los tres tipos pueden subclasificarse en corpus escritos u orales. Asimismo, Granger (2002: 10) añade que la mayoría de los corpus suele ser de carácter sincrónico, es decir, ofrecen muestras lingüísticas recopiladas en un momento dado, a diferencia de los corpus longitudinales, que engloban varios cortes de material lingüístico, mejor dicho, evidencian una evolución lingüística o casos de fosilización.

El corpus de aprendices supone un corpus de materiales producidos por estudiantes de una LE, de forma que se trabaje sobre datos reales recogidos de forma sistemática y organizada (Cestero et al. 2001: 527). El estudio de un corpus de este tipo no solo facilita la formulación de generalizaciones sobre diferentes aspectos del proceso de adquisición de idiomas extranjeros, sino también enriquece nuestro entendimiento de la relativa frecuencia de expresiones o errores determinados. Cestero et al. añaden:

lo más importante es que un corpus sea asequible y que permita llevar a cabo un amplio número de investigaciones diferentes dentro del ámbito de la lingüística aplicada a la enseñanza y adquisición de E/LE, tales como: análisis de errores lingüísticos, problemas de adquisición relacionados con la LM, con el nivel de conocimientos, con el sexo o con la edad de los estudiantes; luego, tipos, frecuencias, condicionamientos e implicaciones de las transferencias lingüísticas; y fundamentalmente, elaboración de materiales didácticos de refuerzo y de consolidación. Estos corpus, por lo tanto, ofrecen un acceso directo a la competencia transitoria, a la interlengua de los estudiantes de idiomas y permite entender una serie de peculiaridades de la misma en distintas etapas de su aprendizaje. (Cestero et al. 2001: 527)

A veces se pone en cuestión la autenticidad de un material producido dentro del aula, pero Granger (2002: 8) argumenta que la autenticidad de las redacciones escritas de aprendices, como uno de los elementos clave de cualquier corpus, proviene de una auténtica actividad realizada en clase, limitada solo por lo que concierne al tema y al tamaño.

La elaboración de un amplio corpus de estas características es una tarea exigente, por lo que los investigadores suelen elaborar corpus reducidos que les permitan realizar trabajos puntuales atendiendo a los objetivos de su propia investigación, lo mismo que se ha hecho aquí. La creación y preparación de un corpus para el estudio del proceso de la enseñanza-aprendizaje de una LE se entiende como la primera fase, mientras que la informatización, el almacenamiento y el etiquetado del corpus recogido se ve como la fase posterior. Por eso es muy importante que se establezcan unas relaciones firmes entre los lingüistas, los docentes y los informáticos, porque cada uno contribuye con sus conocimientos y esfuerzos.

Actualmente, en el caso del español como lengua extranjera encontramos varios corpus de aprendices de E/LE, como, por ejemplo, el Corpus de aprendices ingleses del

español (*CEDEL2*)², creado por la Universidad Autónoma de Madrid en colaboración con la Universidad de Granada, luego el corpus de aprendices de E/LE *CORANE*³ elaborado por el personal docente de la Universidad de Alcalá, también *SPLLOC*⁴ (*Spanish Learner Language Oral Corpus*) realizado por la Universidad de Southampton, Reino Unido, y otros que se han elaborado en otras universidades de Europa y también en otros continentes (*SAELE*⁵ de la Universidad de Jönköping, Suecia, y el corpus multilingüe *CoMAprend*⁶ de la Universidad de São Paulo, Brasil).

3. Sobre el corpus de materiales escritos de aprendices serbios de E/LE

A continuación describimos la situación de la recogida de datos y presentamos nuestro corpus numéricamente.

En cuanto a la recogida de datos, se ha realizado en dos ocasiones y en dos universidades de Serbia, las dos públicas: en la Universidad de Belgrado, el Departamento de Estudios Ibéricos de la Facultad de Filología, durante mayo y junio de 2007, es decir, a finales del año escolar 2006/07; y, en la Universidad de Kragujevac, el Departamento de Hispanística de la Facultad de Filología y Artes, durante mayo y junio de 2011, o sea, a finales del año escolar 2010/11.

Con ello hemos pretendido obtener el mayor número posible de redacciones de estudiantes de los cuatro niveles de aprendizaje que se reconocen en los estudios académicos de dos universidades serbias. A fin de conocer las tendencias normales de adquisición y aprendizaje manifestadas en las redacciones era necesario que no existieran grandes diferencias entre los estudiantes. Así, para garantizar la homogeneidad de la muestra hemos elegido estos dos centros de estudios puesto que tanto el nivel de conocimiento de español como el de la destreza escrita concuerdan con el número de horas de instrucción recibidas (aproximadamente un total entre 8 y 10 horas semanales de español), con una leve ventaja en el caso de la Universidad de Belgrado, la que se debe a un mayor número de clases con lectores nativos de español y también a un mayor número de ocasiones para acercarse a la lengua y cultura españolas (eventos culturales organizados por el Instituto Cervantes de Belgrado, por la Embajada española y por las embajadas hispanoamericanas). En las dos universidades se imparten tanto clases lectivas de la lingüística española como clases prácticas de español y de la conversación.

En la recogida de datos, además de pruebas específicas de gramática, pedimos a los estudiantes que redactasen composiciones escritas, aproximadamente 20 líneas de extensión, sobre el tema «El viaje de mis sueños», un tema que parece ser motivador y cercano a los jóvenes, y aún más a los estudiantes de filología, quienes a la pregunta :
–¿por qué se han matriculado en los estudios de filología?– contestan que les gusta mucho viajar. Además, de todos los procedimientos que existen para recoger datos, la redacción libre es tal vez el menos controlado, ya que el investigador influye solo con la decisión del

² <http://www.uam.es/proyectosinv/woslac/cedel2.htm>

³ http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/12/12_0527.pdf

⁴ <http://www.splloc.soton.ac.uk/>

⁵ <http://urn.kb.se/resolve?urn=urn:nbn:se:hj:diva-20244>

⁶ <http://www.fflch.usp.br/dlm/comet/comaprend.html>

tema y del tamaño. Se procuró que la redacción de los textos resultase lo más abierta posible, evitando imposiciones y dando a los informantes un alto grado de libertad. Los estudiantes no disponían de material de consulta, de manera que emplearon sus propias estrategias para la superación de las dificultades comunicativas. El propósito de la recopilación de los datos no se explicó a los alumnos, solo recibieron instrucciones sobre la tarea misma minutos antes de realizarla. El tiempo asignado para redactar la composición fue de treinta minutos en el propio entorno académico de los estudiantes.

Asimismo, se les pidió a los estudiantes que rellenasen un cuestionario escrito respecto a diversas variables relacionadas con el aprendizaje del idioma tales como: edad, sexo, educación, conocimiento de otros idiomas, estudios realizados de español, lugar y duración de su realización, luego ocasiones fuera del aula en las que usan español para comunicarse. Toda esa información⁷ se recoge en una ficha complementaria a la redacción correspondiente.

En lo que concierne a los informantes que constituyen nuestro corpus, la muestra de estudiantes está formada por la mayoría de los estudiantes de primero, segundo, tercero y cuarto curso de estudios hispánicos en las dos universidades mencionadas. Cabe decir que los aprendices se encontraban en un entorno académico de asimilación del material lectivo y práctico, con lo que sus objetivos eran estrictamente académicos y orientados a la realización del examen final.

La muestra de informantes de Belgrado es de 120, y de ese número son 22 pertenecientes al primer curso de español, 36 al segundo, 28 al tercero y 34 al cuarto. En el momento de la recopilación del corpus sus edades comprendían entre diecinueve y veintiocho años, con un total de 108 mujeres frente a 12 hombres, lo que suponía una inmensa mayoría de informantes femeninos. El recuento de palabras utilizadas da un total de 17.531 palabras.

Por otro lado, en la universidad de Kragujevac participaron 102 estudiantes, de los cuales 23 pertenecían al primer curso de español, 20 al segundo, 28 al tercero y 31 al cuarto. Cabe destacar que, aunque el número de los participantes de Kragujevac es menor que el de Belgrado, el recuento de palabras utilizadas llega a ser 19.028 palabras. De nuevo nos encontramos con una inmensa mayoría de informantes femeninos, 87 mujeres frente a 15 hombres. Sus edades van desde los veinte hasta los veintinueve años.

4. La rentabilidad del corpus de aprendices

Llegamos al punto donde cabría justificar la existencia de un corpus de aprendices serbios de E/LE, por qué y para qué es importante tener uno a mano. La rentabilidad es una de las características principales de cualquier corpus, o sea, es de suma importancia que un corpus permita realizar un amplio número de investigaciones diferentes.

Lo esencial para el desarrollo de una enseñanza efectiva de idiomas es el conocimiento de los errores más frecuentes y más característicos de los alumnos que aprenden español, para poder conocer las principales áreas de dificultad a las que se

⁷ Los estudiantes de español a los que se les pidieron las composiciones que forman el corpus han dado su autorización para utilizarlas. En la definitiva base de datos, los apellidos y el nombre del informante son sustituidos por un código numérico para guardar la confidencialidad de los datos personales.

enfrentan los alumnos y para preparar materiales didácticos adecuados a los alumnos en función de su lengua materna (LM). Al examinar la interlengua se puede comprobar si existen etapas inestables en el proceso de adquisición de español por un grupo homogéneo de aprendices, cuáles son los problemas principales que encuentran los estudiantes en cada uno de estos niveles, cuáles desaparecen con la siguiente etapa del aprendizaje y cuáles presentan mayor riesgo de fosilización. También se deduce qué estrategias se activan para resolver esos problemas. Es recomendable que este tipo de investigaciones tengan un carácter fundamental práctico, o sea, junto a una descripción cualitativa del proceso de enseñanza-aprendizaje de E/LE se debe pensar en una futura aplicación didáctica que permita la superación de errores.

Por todo lo que acabamos de exponer, encontramos oportuno que el mismo corpus pueda servir de estímulo investigador en los estudios de posgrado, tanto por parte de los docentes como por los estudiantes.

A continuación, diferenciando los estudios basados en un corpus y derivados de ese (*cf.* Biber 2012: 162), presentamos algunas sugerencias para la posible explotación del material recopilado⁸. Lo que nosotros proponemos son las líneas de investigación basadas en el estudio del corpus, mientras que lo que emergiera espontáneamente de los análisis se atribuiría al estudio derivado del trabajo con el corpus. Así, se sugiere:

- estudiar las características (morfosintácticas, léxicas, discursivas, pragmáticas, culturales) de la producción escrita de aprendices serbios de E/LE;
- comprobar si existen diferentes etapas perceptibles en el proceso de adquisición del español y cuáles son las características de esas etapas;
- diferenciar los textos de acuerdo con el nivel de conocimiento de español, con el conocimiento de otros idiomas extranjeros⁹ y otras variables;
- comparar el progreso de estudiantes en distintas universidades, es decir, averiguar el impacto de la enseñanza sobre el aprendizaje en dos centros diferentes, claro, si hay algún contraste;
- mejorar los materiales de enseñanza de lenguas destacando los puntos que resultan problemáticos para los serbios aprendices de E/LE.

Como se puede observar, son muchas las oportunidades para trabajar con un corpus de aprendices de una LE, solo nos falta tiempo, un trabajo coordinado y buena fe para trazar todas esas líneas de investigación.

5. Procedimientos de análisis computacional del corpus

Como ya se ha mencionado en el trabajo, después de recoger las composiciones, procedimos a hacer la transcripción de las composiciones al ordenador en formato de texto,

⁸ En esta ocasión no pretendemos analizar la interlengua de aprendices serbios de E/LE, más bien invitamos a los interesados a que hicieran diferentes análisis, para poder obtener unos resultados lo más posible englobadores. Nosotros ya hemos analizado el uso del artículo determinado por parte de los hablantes serbios de E/LE, pero los resultados de ese análisis junto con la propuesta del material didáctico de refuerzo se expondrán en la tesis.

⁹ Se ha argumentado en varias ocasiones que los aprendices de idiomas extranjeros tienden a desarrollar estrategias de aprendizaje que les facilitan el proceso de aprendizaje/adquisición de una L3 o L4 (*cf.* Jovanovic, Zecevic Krneta 2012; Ringbom 2007).

se ha empleado el procesador de textos Microsoft® en la digitalización de las composiciones, e hicimos el etiquetado de los códigos identificadores a cada composición: número del estudiante, sexo, curso, universidad. La ficha con los datos personales de los informantes se encuentra en las tablas de Excel y bajo un código unificador se encuentra una redacción y los datos personales correspondientes.

Ahora bien, cuando ya hemos montado un corpus lo suficiente grande para obtener informaciones variadas e interesantes, resulta que es demasiado grande para ser manejado sin ayudas computacionales. Gracias a los colegas informáticos, a lo largo de las dos últimas décadas se han ido desarrollando herramientas que pueden, por un lado, almacenar cantidades de informaciones y, por otro, trabajar las informaciones, es decir, contar palabras, ocurrencias; luego indicar colocaciones, grupos sintácticos etc.

Uno de los programas que se utiliza para el estudio de corpus hoy en día es el programa *Wordsmith*¹⁰ de Mike Smith para el inglés, luego *VocabProfile*¹¹ de Paul Nation para el inglés y el francés, y en el ámbito académico serbio los informáticos suelen usar el programa *Unitex*¹², elaborado por los miembros de *The RELEX Network* (La red internacional de los laboratorios especializados en la lingüística del corpus), fundada por Maurice Gross y su equipo LADL. Los miembros de la red mencionada se dedican a la elaboración de exhaustivos diccionarios electrónicos codificados (*coded*) y etiquetados (*tagged*), también a la elaboración de las tablas léxico-gramaticales que permiten consultar toda la información gramatical de los vocablos de la lista¹³. Actualmente, Vitas, Nakamura, Voyatzi (2009) van elaborando un programa para el manejo cuantitativo y cualitativo del diccionario electrónico paralelo español-francés de palabras compuestas, que sin duda será otra herramienta muy útil para la labor docente.

Pues, como ya hemos dicho, nuestro proyecto a largo plazo es la codificación y la etiquetación del corpus de aprendices serbios de E/LE a base del programa *Unitex*, con el fin de que, en un futuro cercano, el corpus sea de fácil acceso y uso para todo aquel interesado en el estudio de algún aspecto relacionado con la enseñanza o el aprendizaje de E/LE. Nuestra tarea más inmediata es revisar y ampliar el diccionario electrónico del español del que dispone el programa computacional *Unitex*, creado por Xavier Blanco de la Universidad autónoma de Barcelona (*cf.* Blanco 2001a; 2001b), para poner una base firme a posibles análisis computacionales tanto cualitativos (los cambios que se producen en las clases de palabras, palabras gramaticales versus palabras de contenido, la aparición de los usos idiomáticos, o la utilización de frases lexicalizadas), como cuantitativos (hallar las frecuencias absolutas y relativas de los tipos (*type*)¹⁴ y las muestras (*token*)¹⁵ de palabras, así como los patrones de colocación de las palabras dentro de contextos lingüísticos etc.).

Además, cabrá establecer una serie de marcas para identificar los distintos tipos de errores que se manifiestan en la palabra entendida como signo, teniendo en cuenta que las

¹⁰ <http://www.lexically.net/wordsmith/>

¹¹ <http://www.lextutor.ca/>

¹² <http://www-igm.univ-mlv.fr/~unitex/index.php?page=0>

¹³ A toda esa información relativa al mundo de Unitex y las posibilidades de su funcionamiento nos dio acceso el profesor Dusko Vitas de la Facultad de Ciencias Naturales y Matemáticas, por lo que se lo agradecemos mucho.

¹⁴ El tipo (*type*) se refiere al análisis cualitativo y se relaciona con el análisis de ciertas categorías gramaticales (morfemas, lexemas etc.) (Dobrić 2009: 51).

¹⁵ La muestra (*token*) se refiere a la aparición de una palabra cuya frecuencia de uso, o sea, ocurrencia, forma parte del análisis cuantitativo (Dobrić 2009: 51).

palabras se distribuyen en clases y se combinan para formar sintagmas, oraciones y textos. Las marcas que se establezcan serán incluidas en la edición informatizada de textos, de modo que los distintos tipos de errores se puedan fácilmente localizar en la base de datos. Una vez almacenado, etiquetado y marcado, el corpus permitirá realizar un sinfín de estudios, lo cual esperamos cumplir en un futuro no muy lejano.

6. Conclusión

A modo de conclusión, creemos que con este tipo de estudios se pueden obtener unos resultados de gran importancia que puedan aportar nuevos datos a las teorías de adquisición/aprendizaje de E/LE, reforzar diversos aspectos de la enseñanza de E/LE y ayudar a preparar un material didáctico específico.

En esta ocasión hemos hecho una aproximación al corpus de aprendices serbios de E/LE, que fue recopilado en dos universidades de Serbia, en los cuatro niveles de estudios académicos reconocidos. Describimos la situación de la recogida de datos, el mismo proceso de aprendizaje y los informantes. Por el momento, el corpus está informatizado en el formato de *Word* y su propósito comprende estar a disposición de todos los interesados en los estudios del corpus. A lo largo del trabajo, con numerosas directrices para la explotación del corpus en cuestión, recordamos que los estudios del corpus permiten unas nuevas líneas de investigación, y que, a veces, resaltan con más exactitud unos nuevos datos sobre el uso de la lengua, lo que no podría ser anticipado de otro modo. De esa manera, resaltamos la importancia de la creación y constante recreación de un corpus codificado, puesto que eso es una demanda no solo de la lingüística aplicada sino también de la era de las nuevas tecnologías en la que vivimos.

Los datos que hemos expuesto no son concluyentes puesto que todavía queda ampliar el diccionario electrónico que sirve de base de codificación, y luego etiquetar los datos con el programa computacional. Este artículo, además de servir como una presentación de la labor realizada hasta el momento, también ha de servir de estímulo a los colegas interesados en los estudios del corpus, cuyo interés podrá añadir nuevos impulsos a la futura informatización de este corpus o a las investigaciones lingüísticas basadas en el estudio del corpus. Por último, hemos de decir que esperamos ver pronto las huellas de esta trayectoria, que poco a poco se creará una base sólida tan necesaria actualmente en el ámbito de adquisición del español como lengua extranjera.

Bibliografía

- ADOLPHS, Svenja, *Corpus and context*, Amsterdam, John Benjamins, 2008.
- BARNBROOK, George, *Language and computers: a practical introduction to the computer analysis of language*, Edinburgh, Edinburgh University Press, 1996.
- BELZ, A. Julie, VYATKINA, Nina, «The pedagogical mediation of a developmental learner corpus for classroom-based language instruction», en *Language Learning & Technology*, 12 (3), 2008: 33-52.

- BIBER, Douglas, «Corpus-based and corpus-driven analyses of language variation and use», en HEINE, Bernd, NARROG, Heiko (eds.), *The Oxford handbook of Linguistic Analysis*, Oxford, Oxford University Press, 2012: 159-191.
- BLANCO, Xavier, «Dictionnaires électroniques et traduction automatique français-espagnol», en *Langages 143* (Lexicologie contrastive espagnol-français), Larousse, Paris, 2001a: 49-70.
- BLANCO, Xavier, «Les dictionnaires électroniques de l'espagnol (DELASs et DELACs)», en *Linguisticae Investigationes XXIII: 2*, Amsterdam/Philadelphia: Benjamins, 2001b: 201-218.
- BUCKINGHAM, Louisa, *Las construcciones con verbo de soporte en un corpus de especialidad*, Frankfurt am Main, Peter Lang, 2009.
- DOBRIĆ, Nikola, «Korpusna lingvistika kao osnovna paradigma istraživanja jezika», en *Naučno-stručni časopis za jezik, književnost i kulturu Philologia*, 7, 2009: 47-57.
- FUERTES OLIVERA, Pedro A., «El lenguaje de la ciencia y la tecnología», en ALCARAZ VARÓ, Enrique, MATEO MARTÍNEZ, José, YUS RAMOS, Francisco (eds.), *Las lenguas profesionales y académicas*, Barcelona, Ariel, 2007: 205-217.
- GRANGER, Sylviane, «The contribution of learner corpora to second language acquisition and foreign language teaching: a critical evaluation», en AIJMER, Karin (ed.), *Corpora and Language Teaching*, 8, 2009: 13-33.
- GRANGER, Sylviane, «A Birds-eye view of learner corpus research», en GRANGER, Sylviane, HUNG, Joseph, PETCH-TYSON, Stephanie (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, Amsterdam-Philadelphia: John Benjamins, 2002: 3-30.
- GRANGER, Sylviane, HUNG, Joseph, PETCH-TYSON, Stephanie (eds.), *Computer learner corpora, second language acquisition and foreign language teaching*, Amsterdam-Philadelphia: John Benjamins, 2002.
- CESTERO MANCERA, Ana María et al., «Corpus para el análisis de errores de aprendices de E/LE (CORANE)», en GIMENO SANZ, Ana María (ed.), *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE: Actas del XII Congreso Internacional de ASELE*, Valencia, Universidad Politécnica de Valencia, 2001: 527-534.
- HUNSTON, Susan, *Corpora in applied linguistics*, Cambridge, Cambridge University Press, 2002.
- JOHANSSON, Stig, «Some thoughts on corpora and second-language acquisition», en AIJMER, Karin (ed.), *Corpora and Language Teaching*, 8, 2009: 33-44.
- JOVANOVIĆ, Ana, ZEČEVIĆ KRNETA, Gorana, «Abductive reasoning and second language acquisition», en *Journal of Language Teaching and Research*, 3(2), Academy Publisher, 2012: 306-313.
- LEISTYNA, Pepi, MEIER, Charles (eds.), *Corpus analysis: Language structure and language use*, Amsterdam, Rodopi, 2003.
- O'KEEFFE, Anne, MCCARTHY, Michael, CARTER, Ronald, *From corpus to classroom: Language use and language teaching*, Cambridge, Cambridge University Press, 2007.
- OSBORNE, Jonathan, «What can students learn from a corpus? Building bridges between data and explanation», en BURNARD, Lou, MCENERY, Tony (eds.), *Rethinking Language pedagogy from a corpus perspective*, Frankfurt, Peter Lang, 1999: 165-185.
- RINGBOM, Hakan, *Cross-linguistic similarity in foreign language learning*, Clevedon-Buffalo-Toronto, Multilingual Matters, 2007.
- SÁNCHEZ, Álvaro, «Language teaching before and after digitalized corpora. Three main issues», en *Cuadernos de Filología inglesa*, 9(1), 2000: 5-37.

- SINCLAIR, John (ed.), *How to use corpora in language teaching*, Amsterdam, John Benjamins, 2004.
- VITAS, Duško, NAKAMURA, Takuya, VOYATZI, Stavroula, «Using Aligned Corpora to Construct Large-Scaled Linguistic Resources: The Electronic Dictionary of Spanish Compound Nouns», en Book of abstracts: *eLexicography in the 21th century: New challenges, new applications*, 2009. http://www.uclouvain.be/cps/ucl/doc/cecl/documents/eLEX_book_abstracts_FINAL_New.pdf [10/05/2013]
- ZEČEVIĆ KRNETA, Gorana, «El uso del corpus electrónico en la clase de E/LE», en *Actas del XLIV Congreso internacional de la Asociación europea de profesores de español*, AEPPE: EDICIÓN EN CD, 2010: 008B.