



El uso de corpus lingüísticos como herramienta pedagógica

Liliana Tolchinsky
Universidad de Barcelona

La lingüística de corpus es un enfoque metodológico para el estudio de las lenguas. Por medio de la compilación y el procesamiento de corpus lingüísticos ofrece la posibilidad de trabajar con datos reales y exhaustivos que permiten acercarse lo más posible a cómo usan su lengua o sus lenguas los hablantes. Veremos las principales características de los corpus lingüísticos, algunos de los recursos informáticos orientados a su procesamiento y las maneras directas e indirectas en que pueden ser utilizados en el trabajo didáctico en la enseñanza de lenguas primeras y segundas.

Palabras clave: *lingüística de corpus, lengua oral, lengua escrita, usos pedagógicos de corpus.*

The use of language corpora as a teaching tool

Language corpora offer a methodological focus for studying languages. By compiling and processing language corpora, we can real work with exhaustive, real-life data that brings us as close as possible to speakers of the language. In this paper we look at some the main characteristics of language corpora, some of the computing resources aimed at processing them and the direct and indirect means of using them in teaching first and second languages.

Keywords: *language corpora, oral language, written language, educational use of corpora.*

La posibilidad de trabajar con datos reales y exhaustivos que permitan acercarse lo más posible a cómo usan su lengua los hablantes ha provocado algunas sorpresas tanto a lingüistas como a profesores de lengua. Es muy conocida la experiencia que tuvo Geoffrey Sampson, un famoso lingüista inglés, mientras explicaba en una clase que hay ciertas frases que son imposibles para los hablantes. Según Sampson (2001), frases como la del ejemplo [1], en la cual hay sucesivos anidamientos, eran rechazadas por los hablantes:

[1] ¿Que no sabes que la llave que perdimos de la puerta verde de la casa de Blanes no la encontraremos jamás?

Mientras Sampson estaba explicando y dando ejemplos, una de las estudiantes presentes hizo la siguiente intervención:

[2] ¿Pero no cree [que las frases [que dice la gente [que usted conoce] son más fáciles de entender]? (*But don't you find [that sentences [that people [you know] produce] are easier to understand]?*).

La intervención de la alumna contenía sucesivos anidamientos, pero era perfectamente comprensible y la había producido espontáneamente. Esta intervención llevó a Sampson a buscar en diferentes tipos de corpus frases con múltiples anidamientos y descubrió que eran muy natura-

les entre los hablantes de inglés y mucho más frecuentes de lo que los lingüistas suponían. Sucede que muchos de ellos habían llegado a la conclusión de la dificultad de ese tipo de frases en base a ejemplos aislados y fuera de contexto.

Como sucede con los lingüistas, también profesores de lengua y, sobre todo, algunos libros de texto o bases de recursos didácticos aportan a veces ejemplos artificiales que no reflejan la lengua viva ni el conocimiento lingüístico de los hablantes. Para ilustrar esta situación, hemos buscado ejemplos de oraciones de relativo en el corpus CesCa –que presentaremos luego– y en una base de datos de recursos didácticos. Comparemos las diferentes maneras como informantes del corpus de entre 8 y 16 años definen la palabra *bacteria* en [3], [4] y [5] con el ejemplo [6], extraído de la base de recursos didácticos:

[3] es una cosa que es te a les dents. (Corpus CesCa, 8 años)

Es una cosa que se tiene en los dientes.

[4] es un ésser viu de la natura microscòpic que afecta a altres éssers vius. (Corpus CesCa, 16 años)

Es un ser vivo de la naturaleza microscópico que afecta a otros seres vivos.

[5] Un bacteri es com un vix que surt Quan no ia molta neteja «son vixets que o be xuclen sang,o picen els dents». (Corpus CesCa, 16 anys)

Una bacteria es como un bicho que sale cuando no hay mucha limpieza, «son bichos que o bien chupan la sangre o pican los dientes»

[6] El ganivet amb el qual han matat el botiguer és d'un model rar. ([http://f-eines.alaxarxa.cat/cat/images/file/ESQUEMA%20ADJECTIVES\(1\).pdf](http://f-eines.alaxarxa.cat/cat/images/file/ESQUEMA%20ADJECTIVES(1).pdf))

El cuchillo con el que han matado al tendero es de un modelo raro.

Si quisiéramos trabajar este tipo de construcciones en la clase de lengua, con las definiciones auténticas de [3], [4] y [5] tendríamos usos mucho más naturales que en el ejemplo [6]. Pero, además, al presentar las definiciones podríamos trabajar no solamente las construcciones de relativo, sino también el conocimiento del mundo que transmiten, los criterios taxonómicos que se aplican y, por qué no, los errores ortográficos. En cambio, el ejemplo [6] impide cualquier reflexión más allá de la construcción de relativo. Sucede que cuando nos acercamos a muestras auténticas de usos lingüísticos aumentan las posibilidades de trabajo.

Los corpus lingüísticos son un instrumento útil para entender mejor cómo usan una lengua los hablantes, e intentaré mostrar por qué es así. Para ello me referiré, en primer lugar, a qué entendemos por *corpus lingüísticos*, cuáles son las características generales que los definen, qué tipos de corpus existen y qué herramientas de análisis pueden utilizarse con miras a un posible aprovechamiento en el aula. En segundo lugar explicaré muy sucintamente en qué consiste la lingüística de corpus y cómo puede relacionarse con el trabajo pedagógico.

■ ¿Qué es un corpus lingüístico?

El término *corpus* es ambiguo y puede referirse tanto a colecciones o inventarios de textos como a corpus informatizados. Sinclair (2004) define así la noción de *corpus*:

Una colección de piezas de una lengua que se seleccionan y ordenan según criterios lingüísticos explícitos para ser utilizados como una muestra de esa lengua.

Los corpus lingüísticos son un instrumento útil para entender mejor cómo usan una lengua los hablantes

La relación entre corpus e informatización es actualmente imprescindible, ya que ésta posibilita un uso dinámico de la información contenida en el corpus

En esta definición no es obligatorio que el corpus esté en formato digital. Pero la relación entre corpus e informatización es actualmente imprescindible, ya que ésta posibilita un uso dinámico de la información contenida en el corpus. Por lo tanto, nos centraremos en corpus informatizados. Pitkowski y Vásquez Gamarra (2009) definen así un corpus informatizado:

Un corpus informatizado es una colección extensa de diferentes tipos de textos, orales o escritos, en formato electrónico, que se codifican y clasifican adecuadamente. Los mismos se guardan y se procesan en medios de almacenamiento masivo (por ejemplo CD, DVD, memorias USB o las memorias flash) y, de esta manera, permiten al usuario hacer diferentes búsquedas entre grandes cantidades de textos electrónicos.

Los avances tecnológicos posibilitan buscar, recuperar, ordenar y hacer cálculos sobre cantidades masivas de texto. Estas técnicas computacionales nos brindan la oportunidad de comprender y de explicar el contenido de estos corpus.

Cuando en la definición de Pitkowski y Vásquez Gamarra se utiliza el término *usuario*, éste puede ser aplicable a un número más o menos restringido de personas que podrán realizar efectivamente las búsquedas y los ordenamientos posibles. Por ejemplo, el *Corpus del patrimonio culinario catalán* contiene 1.125 recetas numeradas y ordenadas alfabéticamente. Pero está publicado en forma de libro y, por lo tanto, no puede acceder a él cualquier usuario.

Aunque uno de los rasgos básicos de un corpus moderno es estar disponible para el público, en muchos casos resulta difícil difundir públicamente un corpus debido a posibles problemas legales con respecto a los derechos de propiedad intelectual y de privacidad. Por eso, para poder acceder al formato digitalizado, en muchos casos habrá que solicitar una autorización a la institución desde la cual se administra el corpus.

En suma, las características generales de un corpus informatizado son: el almacenamiento en soporte electrónico, la posibilidad de recuperación por el usuario, el establecimiento de reglas para poder acceder a la información, la codificación estándar, la clasificación tipológica de los textos y el almacenamiento y mantenimiento a cargo de una institución responsable que facilite su consulta y uso (Marcos Marín, 1994).

■ Tipos de corpus

En principio conviene distinguir entre corpus textuales y corpus orales. Los corpus textuales recopilan material lingüístico de lengua escrita; en cambio, los corpus orales recopilan muestras de lengua oral. Entre los corpus orales debemos diferenciar entre los que presentan solamente una transcripción ortográfica del habla y los que acompañan esa transcripción con un registro sonoro en formato digital.

Según cuales sean los objetivos de su uso será necesario recurrir a un tipo de corpus u otro.

Los avances tecnológicos posibilitan buscar, recuperar, ordenar y hacer cálculos sobre cantidades masivas de texto. Estas técnicas computacionales nos brindan la oportunidad de comprender y de explicar el contenido de estos corpus

Si el objetivo es explorar la frecuencia de uso de términos y construcciones en la lengua escrita de los niños para elaborar un dictado, convendrá consultar un corpus textual. Por ejemplo, el *Corpus de vocabulario del niño de 6 a 14 años (Diccionario de frecuencias)*, de la Universidad de Granada, que está basado en muestras de lenguaje escrito. Si, en cambio, estamos interesados en descubrir diferencias léxicas en lengua oral, se podrá consultar C-ORAL-ROM, un corpus de habla espontánea que incluye muestras de varias lenguas romances, entre ellas el español, y que está disponible en la Universidad Autónoma de Madrid (Moreno y otros, 2005).

Adicionalmente, los corpus se pueden distinguir según el porcentaje y la distribución de los textos que los componen; según la especificidad de los textos que los integran, distinguiéndose los corpus generales de los especializados en alguna variante específica de lengua. Los corpus también se diferencian por la cantidad de texto incluida en cada documento, ya que hay algunos que comprenden fragmentos de diversos tipos de textos, mientras que otros contienen textos completos de un tipo o más.

Otro parámetro de diferenciación tiene que ver con la codificación y las anotaciones añadidas a los textos. Un corpus anotado es aquel en el cual se ha añadido posteriormente algún tipo de información lingüística, es decir, que contiene etiquetas con información morfológica, semántica o sintáctica. Un corpus sin anotar contiene sola-

mente el material lingüístico sin añadidos. Finalmente, otro parámetro de diferenciación tiene que ver con la documentación que acompaña al corpus. El *Corpus del patrimonio culinario catalán*, por ejemplo, ofrece documentación adicional de las recetas a quien la solicite.

■ Recursos informáticos

Existe un gran número de recursos informáticos orientados al procesamiento de corpus de todo tipo. Internet es una fuente inagotable de recursos para buscar corpus de distinto tipo y recursos informáticos para su exploración (véase, por ejemplo, <http://webs2002.uab.es/tradumatica/revista/num7/articles/10/10central.htm>). Algunos de estos recursos solo se pueden utilizar en una lengua determinada o en un sistema de escritura, otros tratan solo corpus de textos planos y otros, tales como el sistema operativo UNIX, permiten tratar cualquier tipo de texto con todo tipo de anotación. Entre los programas más usados para la extracción de datos a partir de corpus cabe mencionar AntConc y WordSmith. Si bien estos programas pueden tratar corpus anotados, las restricciones que imponen los hacen mucho más adecuados para el tratamiento de textos planos. Paralelamente, existen programas que permiten procesar corpus y anotar automáticamente a diferentes niveles: fonético-fonológico, morfológico, léxico, sintáctico, semántico y pragmático.

El procesamiento automático de corpus nos permite obtener el número de palabras que contienen, el número de palabras diferentes, su frecuencia, los grupos de palabras (*n-gramas*) más frecuentes, las concordancias de una palabra. Si los corpus están enriquecidos con información lingüística, podemos observar los datos en función de la categoría, la estructura sintáctica, los sentidos de las palabras y las funciones discursivas, entre otras cosas.

Existe un gran número de recursos informáticos orientados al procesamiento de corpus de todo tipo. Internet es una fuente inagotable de recursos para buscar corpus de distinto tipo y recursos informáticos para su exploración

El procesamiento automático de corpus nos permite obtener los datos cuantitativos que de manera implícita o explícita están contenidos en los corpus y que facilitan y mejoran su análisis cualitativo

Con este utillaje podemos identificar, por ejemplo, las construcciones multipalabra que caracterizan los corpus de especialidad, las expresiones más frecuentes y su distribución según el nivel escolar, los usos lingüísticos no convencionales, las variantes morfológicas más usadas, el porcentaje de nombres, verbos y adjetivos en un corpus, las colocaciones... En síntesis, podemos obtener los datos cuantitativos que de manera implícita o explícita están contenidos en los corpus y que facilitan y mejoran su análisis cualitativo.

■ La lingüística de corpus (LC)

Constituye un enfoque metodológico para el estudio de las lenguas que ofrece oportunidades revolucionarias para analizar cómo utilizan los alumnos su lengua o sus lenguas, dado que se centra en el estudio de las producciones lingüísticas de los hablantes en una situación concreta (Leech, 1997). Un supuesto básico de este enfoque es que a partir de la observación y el análisis de la actuación, de lo que la gente hace con la lengua, es posible desarrollar teorías explicativas. Es decir, teorías que explican cómo funciona el lenguaje y cómo se adquiere.

■ Usos pedagógicos de corpus

Hay por lo menos tres maneras diferentes de aprovechar el uso de corpus en la enseñanza:

de manera indirecta, de manera directa y a través de la recopilación de corpus de aprendices.

De manera indirecta, la existencia de corpus lingüísticos ya ha influido notablemente en la enseñanza de lenguas, aunque no seamos completamente conscientes de ello. Los diccionarios en uso, herramienta indispensable en la enseñanza, están basados en corpus y por esto contienen usos auténticos que se renuevan permanentemente. Estos diccionarios incluyen información sobre colocaciones (las habituales co-ocurrencias de elementos léxicos) y frecuencia que han posibilitado la definición de los vocabularios básicos en diferentes lenguas (Xiao, 2010). Como los corpus en los que se basan los diccionarios contienen información sobre el registro, el género y la variedad lingüística del material lingüístico que recopilan, los diccionarios pueden incluir también esta información en sus entradas.

También de manera indirecta, la utilización de corpus ha influido en el diseño de materiales didácticos y en el desarrollo de programas, sobre todo, por ahora, en la enseñanza de segundas lenguas. Los datos provenientes del corpus permiten descubrir los puntos de divergencia entre las estructuras gramaticales, las expresiones y la organización del discurso tal y como aparece en los libros de texto y las que emergen del análisis de los corpus (Mindt, 1996). En muchos casos estas divergencias podrían explicar las dificultades de comunicación entre hablantes nativos y

La lingüística del corpus ofrece oportunidades revolucionarias para analizar cómo utilizan los alumnos su lengua o sus lenguas, dado que se centra en el estudio de las producciones lingüísticas de los hablantes en una situación concreta

Otro aspecto de la pedagogía de la lengua en el cual se utiliza el enfoque basado en corpus es el de la evaluación

aprendices de segundas lenguas. Como vimos antes, los corpus proporcionan ejemplos más realistas de uso de la lengua, que reflejan la complejidad y los matices del lenguaje natural.

Otro de los efectos indirectos del uso de corpus lingüísticos es la creciente importancia que se da a la aproximación léxica, en particular a la enseñanza de las colocaciones, las expresiones idiomáticas y las llamadas «unidades prefabricadas». Existe consenso sobre la importancia del conocimiento de estas unidades para el desarrollo de habilidades lingüísticas tanto en L1 como en L2 (Cowie, 1994; Hoey, 2000, 2004; McEnery y Wilson, 2001; McAlpine y Myles, 2003; Nesselhauf, 2003).

Otro aspecto de la pedagogía de la lengua en el cual se utiliza el enfoque basado en corpus es el de la evaluación. La información sobre estructuras gramaticales, giros idiomáticos y palabras clave en contexto se utiliza para la construcción de pruebas, así como para el diseño de ejercicios, la confección de textos orales y escritos que se incluyen en las evaluaciones y la definición de los criterios de valoración (Alderson, 1996).

En aquellos entornos en los cuales se han difundido enfoques pedagógicos basados en corpus (por ejemplo, en Estados Unidos, Australia o Gran Bretaña) se insiste en que la utilización de corpus debería formar parte de la formación del profesorado. En efecto, para que los alumnos se beneficien de la utilización de corpus, en primer lugar los profesores deben estar familiarizados con este enfoque. Ya hay numerosas experiencias en la formación continua de profesores de lengua (Allan, 1999, 2002;

Conrad, 2000; Seidlhofer, 2000, 2002) y propuestas para incluir la lingüística de corpus en la formación inicial del profesorado de lengua (O’Keeffe y Farr, 2003).

■ Usos directos

La cuestión de cómo utilizar los corpus en la clase de lengua se ha debatido ampliamente, sobre todo en la enseñanza de segundas lenguas, y está comenzando a introducirse en otras áreas del currículo. En términos generales, para la clase de lengua se propone que profesores y alumnos miren juntos los datos reales, los discutan e intercambien opiniones sobre ellos y traten de descubrir juntos las razones que subyacen a los usos detectados (Carter y McCarthy, 1995, p. 155). Se trata de que alumnos y profesores asuman una actitud investigadora respecto a los hechos lingüísticos que detectan. Esta actitud de investigación conjunta de datos lingüísticos puede ser promovida en los diferentes niveles educativos; la clave de su éxito radica en la guía del maestro o la mediación pedagógica en función de la edad de los alumnos y el nivel de competencia (Leech, 1997).

Además de la enseñanza de inglés como segunda lengua o lengua extranjera, en general, un enfoque pedagógico basado en corpus puede ser especialmente útil para adentrarse en los usos específicos de la lengua y de la comunicación profesional. Las comparaciones entre textos académicos y conversacionales o entre textos académicos de distintas disciplinas iluminan acerca de las características distintivas de cada registro y tipo de texto.

Para que los alumnos se beneficien de la utilización de corpus, en primer lugar los profesores deben estar familiarizados con este enfoque

A diferencia de una comparación realizada por medio de la lectura de diferentes tipos de textos, el trabajo con corpus informatizados posibilita interactuar con los textos. Es decir, los alumnos podrán comprobar directamente las diferencias en riqueza léxica, computando las relaciones *type/token*, y en densidad léxica, calculando la relación entre palabras contenido y palabras función; podrán calcular la media de longitud de palabra y debatir las razones por las cuales ciertos textos contienen palabras más largas que otras; podrán descubrir las colocaciones más frecuentes en cada tipo de texto, las palabras contenido más frecuentes en cada uno y una larga lista de otras posibilidades.

La ventaja de este tipo de exploración es que a partir de medidas superficiales, como la longitud media de las palabras, es posible debatir aspectos mucho más profundos del funcionamiento de una lengua, como la complejidad morfológica. Constituye una magnífica oportunidad para que los alumnos formulen hipótesis de trabajo (por ejemplo, ¿en qué tipo de texto encontraremos mayor densidad léxica? ¿Por qué?) y las comprueben en los análisis que realizan. A través de este trabajo, el material lingüístico en los textos se convierte en un espacio de problema.

■ Corpus de aprendices

Uno de los desafíos más interesantes en el aprovechamiento de corpus en la enseñanza surge a partir de la creación y el uso de los llamados *corpus de aprendices*. Hasta hace poco, la mayor parte de los corpus que se compilaban correspondían a muestras de lengua normativa. Actualmente se recopilan corpus procedentes de entornos web donde se producen situaciones de comunicación muy abiertas y espontáneas en las cuales se alteran las convenciones normativas de la lengua escrita.

Al igual que los geólogos, que exploran las rocas porque están interesados en los procesos geológicos a los que no tienen acceso directo, los investigadores en adquisición de lenguas y los profesores de lengua pueden explorar los datos de los aprendices para inferir lo inaccesible: los procesos mentales de adquisición de las lenguas

Esta alteración de las convenciones de la lengua escrita también la encontramos en los corpus producidos por aprendices de segundas lenguas, por escolares que están en proceso de aprendizaje de la escritura o por cualquier hablante en un contexto de comunicación informal. En principio, cualquier profesor de lengua podría compilar corpus de aprendices y explotarlos didácticamente.

En este tipo de corpus suelen realizarse varias versiones, ya que los textos no normativos no se pueden procesar automáticamente, pero todo el proceso de preparación de los textos origina situaciones de aprendizaje muy ricas tanto para profesores como para alumnos. Por ahora, los corpus de aprendices se han utilizado sobre todo en enseñanza de segundas lenguas, pero podemos prever su generalización a la enseñanza de primeras lenguas y también de otras asignaturas. Si los corpus de hablantes nativos de la lengua diana y los corpus de textos de escritores profesionales o especialistas proporcionan un enfoque *de arriba hacia abajo* para el uso de corpus, los corpus de aprendices proporcionan un enfoque *de abajo hacia arriba* para la enseñanza (Osborne, 2002).

Stubbs (2001) establece un paralelismo entre el uso de los corpus de aprendices en lingüística y en pedagogía y la utilización de las rocas en geología. Al igual que los geólogos, que

exploran las rocas porque están interesados en los procesos geológicos a los que no tienen acceso directo, los investigadores en adquisición de lenguas y los profesores de lengua pueden explorar los datos de los aprendices para inferir lo inaccesible: los procesos mentales de adquisición de las lenguas.

Referencias bibliográficas

- ALDERSON, C. (1996): «Do corpora have a role in language assessment?», en THOMAS, J.; SHORT, M. (eds.): *Using corpora for language research*. Londres. Longman, pp. 248-259.
- ALLAN, Q. (1999): «Enhancing the language awareness of Hong Kong teachers through corpus data». *Journal of Technology and Teacher Education*, vol. 7(1), pp. 57-74.
- (2002): «The TELEC secondary learner corpus: a resource for teacher development», en GRANGER, S.; HUNG, J.; PETCH-TYSON, S. (eds.): *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam. John Benjamins, pp. 195-212.
- CARTER, R.; MCCARTHY, M. (1995): «Grammar and the spoken language». *Applied Linguistics*, vol. 16(2), pp. 141-158.
- CONRAD, S. (2000): «Will corpus linguistics revolutionize grammar teaching in the 21st century?». *TESOL Quarterly*, vol. 34, pp. 548-560.
- COWIE, A. (1994): «Phraseology», en ASHER, R. (ed.): *The encyclopaedia of language and linguistics*. Vol. 6. Oxford. Pergamon Press, pp. 3168-3171.
- HOEY, M. (2000): «A world beyond collocation: new perspectives on vocabulary teaching», en LEWIS, M. (ed.): *Teaching collocations*. Hove. Language Teaching Publications, pp. 224-245.
- (2004): «Lexical priming and the properties of text», en PARTINGTON, A.; MORLEY, J.; HAARMAN, L. (eds.): *Corpora and discourse*. Berna. Peter Lang, pp. 385-412.
- LEECH, G. (1997): «Teaching and language corpora: a convergence», en WICHMANN, A., y otros (eds.): *Teaching and language corpora*. Londres. Longman, pp. 1-23.
- MARCOS MARÍN, F. (1994): *Informática y Humanidades*. Madrid. Gredos (Grandes Manuales, 44).
- MCALPINE, J.; MYLES, J. (2003): «Capturing phraseology in an online dictionary for advanced users of English as a second language: a response to user needs». *System*, núm. 31, pp. 71-84.
- MCENERY, T.; WILSON, A. (2001): *Corpus linguistics*. 2.^a ed. Edimburgo. Edinburgh University Press.
- MINDT, D. (1996): «English corpus linguistics and the foreign language teaching syllabus», en THOMAS, J.; SHORT, M. (eds.): *Using corpora for language research*. Harlow. Longman, pp. 232-247.
- MORENO, A., y otros (2005): «The Spanish Corpus», en CRESTI, E.; MONEGLIA, M. (eds.): *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam. John Benjamins, pp. 135-161.
- NESSELHAUF, N. (2003): «The use of collocations by advanced learners of English and some implications for teaching». *Applied Linguistics*, vol 24(2), pp. 223-242.
- O'KEEFFE, A.; FARR, F. (2003): «Using language corpora in initial teacher education: pedagogic issues and practical applications». *TESOL Quarterly*, vol. 37(3), pp. 389-418.
- OSBORNE, J. (2002): «Top-down and bottom-up approaches to corpora in language teaching», en CONNOR, U.; UPTON, T. (eds.): *Applied corpus linguistics: A multidimensional perspective*. Amsterdam. Rodopi, pp. 251-265.
- PITKOWSKI, E.F.; VÁSQUEZ GAMARRA, J. (2009): «El uso de los corpus lingüísticos como herramienta pedagógica para la ense-

- ñanza y aprendizaje de ELE». *Tinkuy: Boletín de Investigación y Debate*, núm. 11, pp. 31-51.
- SAMPSON, G. (2001): *Empirical linguistics*. Londres. Continuum.
- SEIDLHOFER, B. (2000): «Operationalizing intertextuality: using learner corpora for learning», en BURNARD, L.; MCENERY, T. (eds.): *Rethinking language pedagogy from a corpus perspective*. Nueva York. Peter Lang, pp. 207-24.
- (2002): «Pedagogy and local learner corpora: working with learning driven data», en GRANGER, S.; HUNG, J.; PETCH-TYSON, S. (eds.): *Computer learner corpora, second language acquisition and foreign language teaching*. Philadelphia. John Benjamins, pp. 213-234.
- SINCLAIR, J. (ed.) (2004): *How to use corpora in language teaching*. Amsterdam. John Benjamins.
- STUBBS, M. (2001): «Texts, corpora, and problems of interpretation: A response to Widdowson». *Applied Linguistics*, vol. 22(2), pp. 149-172.
- XIAO, R. (2010): «What corpora can offer in language teaching and learning», en HINKEL, E. (2010): *Handbook of research in second language teaching and learning*. Londres. Routledge.

Bibliografía

- APARICI, M. (2010): *El desarrollo de la conectividad discursiva en diferentes géneros y modalidades de producción*. Tesis doctoral no publicada. Universidad de Barcelona.
- BERMAN, R.A.; VERHOEVEN, L. (2002): «Crosslinguistic perspectives on the development of text production abilities in speech and writing». *Written Language and Literacy*, vol. 5(1), pp. 1-44.
- FILLMORE, Ch. (1992): «Corpus linguistics or computer-aided armchair linguistics», en SVARTVIK, J. (ed.): *Directions in corpus linguistics*. Berlín. Mouton de Gruyter, pp. 13-38.

- GRIES, S. (2010): «Behavioral profiles», en *The mental lexicon*. Amsterdam. John Benjamins.
- LLAURADÓ, A.; MARTÍ, M.A.; TOLCHINSKY, L. (2013): «Corpus CesCa: Compiling a corpus of written Catalan produced by school children». *International Journal of Corpus Linguistics*, vol. 17(3).
- STEFANOVITX, A.; GRIES, S. (2008): «Corpora and Grammar», en *Corpus linguistics*. Edimburgo. Edinburgh University Press.
- TORRUELLA, J.; LLISTERRI, J. (1999): «Diseño de corpus textuales y orales», en BLECUA, J.M., y otros (eds.): *Filología e informática: Nuevas tecnologías en los estudios filológicos*. Barcelona. Universidad Autónoma de Barcelona / Milenio, pp. 45-77.

Referencias de la autora

Liliana Tolchinsky

Universidad de Barcelona. Grupo de Investigación para el Estudio del Repertorio Lingüístico (GRERLI) ltolchinsky@ub.edu
Lineas de trabajo: desarrollos tardíos del lenguaje, aprendizaje de la lengua escrita, lingüística de corpus.

Este artículo fue solicitado por TEXTOS DE DIDÁCTICA DE LA LENGUA Y DE LA LITERATURA en julio de 2013 y aceptado en marzo de 2014 para su publicación.