

Corpus y enseñanza del español

(Corpora in Spanish language teaching)

Victoria Vázquez Rozas (Universidade de Santiago de Compostela)

Marta Blanco (Universidade de Santiago de Compostela)

1. Introducción

El capítulo ofrece un panorama actualizado de la lingüística de corpus aplicada a la enseñanza del español como primera y segunda lengua, atendiendo tanto a la vertiente investigadora como pedagógica. En uno y otro ámbito, se muestra la utilidad de los corpus para los distintos agentes implicados en el proceso de enseñanza-aprendizaje (investigadores, profesores, estudiantes, autores de materiales, evaluadores y formadores de profesores). La exposición pone de manifiesto que el tipo de corpus (oral, escrito, general, especializado, multilingüe, de aprendientes, entre otros) determina las opciones de uso y explotación no solo en lo que se refiere a los agentes involucrados sino también en cuanto al área específica de estudio (nivel fónico, morfosintáctico, léxico y pragmático-discursivo). En relación con este aspecto, se observa que el tratamiento y anotación que han recibido los materiales de los diferentes corpus es de crucial importancia para su aprovechamiento en la enseñanza del español, pues condiciona la información que puede ser objeto de búsqueda y recuperación, algo que resulta relevante desde la perspectiva del corpus como fuente de datos y también para su uso como herramienta de aprendizaje.

A partir del examen de los diferentes recursos, investigaciones y materiales desarrollados para la enseñanza del español, el capítulo da cuenta del alcance actual y las direcciones futuras de la lingüística de corpus aplicada a un campo de gran interés y proyección en los próximos años.

Palabras clave: corpus; aprendizaje; enseñanza; español primera lengua / segunda lengua / lengua extranjera; desarrollo de materiales

The chapter offers an updated view of corpus linguistics applied to the teaching of Spanish as a first and second language, encompassing both research and pedagogy. In the two areas, the usefulness of the corpora for the actors concerned in the teaching-learning process will be shown. Actors are researchers, teachers, learners, authors of educational materials, examiners and teacher trainers. The presentation reveals that the type of corpus (spoken, written, general, specialised, multilingual, learner corpus, among others) determines the use and exploitation of the resources not only as far as the involved actors are concerned but also depending on the specific area of study (phonic, morpho-syntactic, lexical, and discourse-pragmatic levels). With regard to those aspects, it has been noted that the treatment and annotation of the materials of the different corpora are of vital importance for their use in the field of Spanish language teaching, as they condition the information that can be searched and recovered from the texts. This is relevant from the perspective of the corpus as a data source and also for the use of corpus as learning tool.

By examining the different resources, investigations, and materials specifically developed for the teaching of Spanish, the chapter gives an account of the current issues and future directions of corpus linguistics applied to a growing field of study in the years to come.

Keywords: corpus; learning; teaching; Spanish as a first/second/foreign language; materials development

2. Conceptos fundamentales

Los términos segunda lengua (L2) y lengua extranjera (LE) se han empleado por igual para hacer referencia a la lengua o lenguas que un individuo adquiere después de su lengua materna. Ahora bien, atendiendo al contexto en que tiene lugar el aprendizaje, se puede establecer una diferenciación entre L2 y LE. Así, se habla de L2 si el nuevo código está presente en el entorno social del aprendiente y de LE si no lo está. En cuanto a los términos adquisición y aprendizaje, aunque se usan a veces como sinónimos para designar el proceso de interiorización de una segunda lengua, diversos estudiosos han establecido una distinción entre ambos en función del sistema de apropiación, “subconscious language acquisition” o “conscious language learning” (Krashen 1988). La adquisición remite a un proceso inconsciente y espontáneo caracterizado por su automaticidad y creatividad mediante el uso de estrategias universales, está orientado más hacia el significado que hacia la forma y exige la participación en situaciones de comunicación real. El aprendizaje, en cambio, es un proceso consciente y voluntario que implica explicitación de reglas, retroalimentación y corrección de errores, y que tiene lugar normalmente en un contexto formal e institucional.

Las lenguas no son ajenas a las modificaciones que se producen en las sociedades en las que se hablan. Debido a los movimientos migratorios y cambios sociales de naturaleza diversa el estatus de las lenguas puede variar. En este contexto, el término lengua de herencia (LH) puede entenderse desde un punto de vista abarcador para hacer referencia a lenguas originarias, coloniales, de inmigrantes o minoritarias (el español, por ejemplo, ha pasado históricamente por varias de esas categorías en Estados Unidos). Esta concepción abierta de LH dificulta la interpretación del término hablantes (o aprendientes) de lengua de herencia para el que se han propuesto distintas definiciones, si bien todas ellas giran en torno a dos elementos distintivos, por una parte, la conexión personal o familiar con un grupo concreto y, por otra, el dominio de competencia en la lengua (Belpoliti y Gironzetti 2019).

En las últimas décadas, las demandas planteadas por la sociedad exigen una enseñanza especializada de lenguas extranjeras, de ahí que junto a la enseñanza de lenguas para fines generales se encuentre también la enseñanza de lenguas con fines específicos o enseñanza de lenguas de especialidad. La primera orientación permite a los aprendientes adquirir conocimientos para desenvolverse de forma correcta y adecuada en cualquier situación comunicativa; la segunda, en cambio, atiende a necesidades y objetivos particulares de un grupo de aprendientes que necesitan desarrollar competencias para comunicarse en diferentes contextos concretos ya sean académicos o profesionales, de ahí que se suele distinguir en su interior entre lengua extranjera para fines profesionales (español jurídico, español de negocios, etc.), lengua extranjera para fines académicos y lengua extranjera científica y tecnológica.

El ámbito disciplinar de la enseñanza de lenguas cuenta en la actualidad con nuevos recursos tecnológicos, entre los que destacan por sus posibilidades didácticas y de investigación las aportaciones derivadas del desarrollo de los corpus lingüísticos desde finales del siglo XX. Se emplea aquí el término corpus en un sentido restringido para hacer referencia a aquellos recursos computacionales que dan acceso a una recopilación de muestras de discurso –oral, escrito, signado– reunidas y clasificadas de acuerdo con determinados criterios previstos en su diseño. La adecuada codificación de tales criterios permite al usuario manejar y combinar en sus consultas la información referida a las propiedades externas de los textos,

como el lugar y el tiempo de producción, el género textual, las características de los participantes o el medio de transmisión, es decir, los llamados metadatos. Adicionalmente, los materiales pueden enriquecerse con información lingüística de diverso tipo mediante la anotación morfosintáctica (*POS tagging*), sintáctica (*parsing*) y también semántica y pragmática, aunque estas últimas con un desarrollo todavía incipiente en español.

Dependiendo de las particularidades de los corpus y de los objetivos que han guiado su elaboración, los textos pueden incorporar, por ejemplo, marcas o etiquetas que codifican rasgos propios de la oralidad, como solapamientos entre intervenciones, palabras cortadas o alargamientos de sonidos, o bien, en los corpus de habla infantil y corpus de aprendientes, anotación específica que refleje adecuadamente los distintos estadios de adquisición (L1) o aprendizaje (L2). En este último caso se maneja el concepto de interlengua para hacer referencia al sistema que va construyendo el aprendiente en el proceso de apropiación de la L2. En los corpus de aprendientes resulta muy útil la anotación de las características de los sistemas intermedios para facilitar el análisis contrastivo tanto entre las interlenguas como entre interlenguas y L1.

En cuanto al sistema de representación, los corpus orales pueden seguir las convenciones ortográficas, como los escritos, pero con algunas adaptaciones, o bien incluir una transcripción fonética con mayor o menor grado de detalle. Algunos recursos ofrecen asimismo la opción de acceso al audio y al vídeo alineados con la transcripción, lo que amplía las vías de uso y explotación de los materiales en la enseñanza y aprendizaje de la lengua.

Tanto la estructura de metadatos de los documentos que integran un corpus como la incorporación de distintos niveles de etiquetación y representación determinan las posibilidades de recuperación de los datos contenidos en el recurso. Para facilitar la explotación es necesario, además, contar con un motor de búsqueda que permita la combinación flexible de los parámetros de clasificación y los tipos de anotación que configuran el corpus. La utilidad de un corpus para la enseñanza de la lengua no reside tanto en su tamaño como en la riqueza de opciones que ofrece para extraer la información que contiene.

3. Estado de la cuestión

Desde sus inicios la lingüística de corpus (LC) se aplicó provechosamente a la enseñanza de lenguas, especialmente en el campo lexicográfico a partir de los datos de frecuencias léxicas y en estudios de léxico-gramática. A pesar de que los primeros corpus textuales, en el sentido moderno del término, fueron construidos hace solo unas décadas, su uso en la enseñanza-aprendizaje de lenguas extranjeras se ha consolidado en los últimos años, especialmente para el inglés (Sinclair 2004; Aston *et al.* 2004; O’Keeffe *et al.* 2007; O’Keeffe y McCarthy 2010; Timmis 2015; Poole 2018).

En lo que se refiere al español, en general el contacto de los agentes de la enseñanza con los recursos y métodos de la LC ha sido bastante limitado hasta época reciente. No obstante, en la actualidad se dan unas condiciones muy favorables para que la enseñanza del español se beneficie del uso de corpus. Por una parte, se ha ampliado el conocimiento de los corpus disponibles, de sus características y de sus posibilidades de aplicación didáctica, tanto a través de publicaciones panorámicas, como Cruz Piñol (2012), Mendikoetxea (2014) y Parodi y Burdiles (2019), como por la cada vez más frecuente difusión de trabajos metodológicos, descriptivos y analíticos basados en corpus. Por otra, se han creado nuevos corpus y se han mejorado las opciones de acceso y recuperación de la información textual y contextual que contienen.

De especial interés en el área de ELE son los corpus de aprendientes, entre los que están CAES (Palacios *et al.* 2019), CEDEL2 (Lozano y Mendikoetxea 2013) y el más reciente CINEAS como corpus escritos, y CORELE (Campillos Llanos 2014) y SPLLOC (Mitchell *et al.* 2008) como ejemplo de corpus orales (para más información sobre estas y otras compilaciones véase Rojo y Palacios en este mismo volumen). Pero también los corpus de L1 constituyen un instrumento de gran utilidad para los estudiantes de español, pues no solo los grandes corpus de referencia (CREA, CORPES, CdEGH) han incorporado herramientas de anotación cada vez más completas para facilitar su consulta, sino que se ha iniciado la construcción de recursos específicos para su explotación en la enseñanza de ELE, como es el caso de C-Or-DiAL, un corpus anotado constituido por registros orales naturales de hablantes nativos diseñado para la enseñanza del español como LE (Nicolás Martínez 2012).

La LC suele identificarse con el estudio de los hechos lingüísticos basado en datos de corpus, pero se ocupa también del diseño y desarrollo de los corpus y de los problemas metodológicos y técnicos relacionados con la anotación y recuperación de los datos, aspectos cruciales para una explotación adecuada de los recursos (Rojo 2015). El primer enfoque de la LC coincide con dos de las líneas de confluencia entre corpus lingüísticos y enseñanza indicadas por Leech (1997), respectivamente, el uso indirecto de los corpus como fuente de datos y su uso directo en la práctica docente. El segundo enfoque de la LC se vincula con la tercera línea de Leech (1997), el desarrollo de corpus orientados a un aprovechamiento didáctico (*cf.* también McEnery y Xiao 2010). Las tres líneas señaladas permiten organizar la presentación que se ofrece en este apartado sobre el panorama de los corpus textuales y la enseñanza del español en la actualidad.

i. Uso indirecto de los corpus en la enseñanza

El acceso al uso de la lengua en su contexto natural es una necesidad en diferentes ámbitos, muy especialmente en la investigación para la enseñanza de L1/L2/LE/LH y en el diseño de materiales didácticos. Los corpus proporcionan información cuantitativa (frecuencias) y cualitativa (uso contextualizado) de las unidades lingüísticas, tanto más refinada cuanto más detallada y completa haya sido la anotación de los textos.

Algunos trabajos tienen un objetivo básicamente metodológico, por ejemplo, al comparar las descripciones y las muestras de lengua que ofrecen los manuales de ELE con los datos de corpus (*cf.* Colen Meniconi 2009). En otros casos ponen de relieve las ventajas del uso combinado de varios corpus, especialmente mediante el contraste entre datos ‘nativos’ y el uso registrado en corpus de aprendientes (Caballero Benavente *et al.* 2012). La necesidad de confrontar las producciones de los aprendientes con las de los hablantes de español L1 se refleja en el diseño del recurso CEDEL2, que incluye, junto al grueso del corpus formado por textos escritos por aprendices de español con L1 inglés, un subcorpus de textos comparables producidos por hablantes nativos. Esta configuración de CEDEL2 ha permitido investigar el uso que hacen hablantes de español L2 respecto al que realizan los nativos en aspectos como las colocaciones (García Salido y García 2017) o las construcciones verbales (Sánchez Rufat 2016). De manera similar, Campillos (2014) incluye la producción oral de algunos hablantes de español L1 como grupo de control en su análisis contrastivo de interlenguas basado en muestras orales de aprendientes de español con diferentes L1.

Resultan también útiles para la enseñanza del español los análisis que se puede llevar a cabo con corpus comparables, como muestran, por ejemplo, Jørgensen y Olza (2019), al contrastar los usos registrados en el corpus COLA con los correspondientes en corpus similares de inglés y noruego.

Los corpus paralelos son asimismo una valiosa fuente de datos para diferentes aplicaciones relacionadas con la enseñanza del español como L2, si bien su uso primario se ha relacionado más con los ámbitos del procesamiento del lenguaje natural y la traducción. Estudios

recientes, como los llevados a cabo a partir del corpus ACTRES inglés-español, muestran el interés de este tipo de recursos para la investigación contrastiva y la enseñanza del ELE (cf. Labrador 2015; Ramón 2015). Igualmente, algunos de los trabajos publicados en Doval y Sánchez Nieto (2019) apuntan nuevas posibilidades de uso de los corpus paralelos para el aprendizaje de idiomas.

Un enfoque netamente didáctico del uso indirecto de corpus en la enseñanza es el del diseño de actividades para el trabajo en el aula. Una muestra de este aprovechamiento la encontramos en Campillos Llanos, Gozalo y Moreno (2007), que elaboran una propuesta didáctica sobre distintos aspectos gramaticales y comunicativos basándose en el C-ORAL-ROM. Por su parte, Albelda y Briz (2017) ilustran las posibilidades de explotación del corpus Valesco para la enseñanza de la pragmática en español LE.

Como para otras lenguas, en el desarrollo de la LC aplicada a la enseñanza del español es de crucial importancia la creciente disponibilidad de corpus anotados en diversos niveles. Mientras los estudios iniciales con datos de corpus partían casi invariablemente de la extracción de concordancias de palabras ortográficas, en la actualidad es posible un aprovechamiento más refinado de los datos a partir de la información léxica, gramatical, semántica y pragmática que ya incorporan algunos corpus. Así, la etiquetación morfosintáctica de corpus de referencia como el CREA (versión anotada 0.1) y el CORPES XXI (versión 0.91) facilita la recuperación de información a un nivel más abstracto que el de la representación gráfica, de tal modo que es posible obtener con una consulta simple todas las formas del paradigma de un elemento léxico variable, como un verbo o un adjetivo, o bien todas las ocurrencias de una determinada clase de palabras o categoría gramatical con independencia del lema o lemas particulares que las instancien.

ii. Uso directo de los corpus en la enseñanza

Abundan los trabajos que exponen argumentos a favor del uso directo de los corpus en la enseñanza de la lengua española (casi siempre LE/L2) (cf. Gabrielatos 2005; Blake y Zyzik 2016), pero son escasas las publicaciones con aplicaciones y desarrollos concretos en el ámbito del español (véase, sin embargo, Cruz Piñol 2012, capítulos 9 y 12; Cao, da Cunha e Iruskietta 2016). Se ha señalado además que, al menos para ciertos fines, es preciso que los materiales incluidos en los corpus reciban un tratamiento didáctico para ser realmente útiles a los aprendientes. A tal necesidad responde, por ejemplo, la etiquetación de funciones comunicativas a la que se han sometido las transcripciones del corpus C-Or-DiAL, así como la clasificación temática y por niveles de las conversaciones de hablantes nativos que lo componen. Todo ello confluye en publicaciones específicas como Nicolás Martínez y Hernández Toribio (2015), que aporta material complementario para ELE a partir de las transcripciones y audios de dicho corpus.

Otro recurso creado igualmente con un objetivo didáctico es el *Columbia corpus de conversaciones para ELE*, integrado por conversaciones entre hablantes nativos de español de diferentes orígenes geográficos, videograbadas y transcritas ortográficamente. El corpus Columbia incluye entre los metadatos los temas sobre los que se habla y la indicación del nivel mínimo de español que se considera necesario para un aprovechamiento adecuado de cada conversación. Además, se proporciona una guía de uso de los materiales con indicaciones y propuestas de actividades que resultan útiles tanto en el aula como para orientar el trabajo autónomo de los estudiantes.

iii. Desarrollo de corpus orientados a la enseñanza

Junto con recursos específicos para el estudio del español LE, como los citados C-Or-DiAL y Columbia, ambos corpus monolingües, el interés por el diseño de corpus como instrumentos para la enseñanza del español se plasma especialmente en la creación de corpus de

aprendientes, algunos de los cuales se mencionaron anteriormente (cf. Rojo y Palacios en este mismo volumen para más detalles sobre los corpus de aprendientes).

Una necesidad creciente en la enseñanza del español, tanto L1 como L2, es el acceso a datos de lengua de especialidad. En este contexto debe citarse El Grial, una plataforma para la anotación textual que permite consultar un amplio corpus de lenguaje académico. En Parodi (2007) se reúnen algunos trabajos que muestran el aprovechamiento de los recursos y materiales de El Grial para L1, tanto en enseñanza secundaria como en el nivel universitario. No obstante, son de momento escasos los recursos textuales para la enseñanza del español con fines específicos. Entre los corpus que se integran en esta línea se encuentran el *Corpus técnico del IULA*, constituido por textos escritos de distintas áreas de especialidad (derecho, medio ambiente, economía, etc.) y el *Corpus Iberia*, que contiene documentos referidos al español científico-técnico.

Como adelantábamos en la Introducción, una línea fundamental de desarrollo de la LC se encuentra en el enriquecimiento de los corpus mediante recursos de lematización, anotación morfosintáctica, sintáctica, semántica y pragmática. El seguimiento de los estándares internacionales en la codificación de los textos favorece, además, la reutilización de los corpus para fines diferentes de los que motivaron su creación, lo cual amplía las posibilidades de explotación pedagógica de corpus que no fueron diseñados original y específicamente con un objetivo didáctico. Rabadán (2019) muestra, por ejemplo, las posibilidades y beneficios de combinar la información extraíble de los corpus paralelos con los datos de corpus comparables y monolingües. En el siguiente apartado se apuntan mediante ejemplos concretos algunas de estas vías de aprovechamiento de los corpus etiquetados para la enseñanza del español.

4. Consideraciones metodológicas

Los métodos de trabajo con corpus textuales en la enseñanza del español varían dependiendo de los objetivos que se persiguen y de las condiciones y características del corpus o corpus utilizados en cada caso. Desde una perspectiva general, destacan dos vías de aprovechamiento que están estrechamente ligadas a dos orientaciones que en la actualidad se siguen en la investigación y aplicaciones de corpus en la enseñanza de idiomas, *corpora for learners* y *corpora with learners* (Aston *et al.* 2004). Aunque estas dos líneas están asociadas fundamentalmente al ámbito de la didáctica de una lengua como lengua extranjera, también son aplicables a la enseñanza de una lengua como L1 o LH. Los corpus pueden ser vistos, por tanto, como una fuente de información (*corpora for learners*) o como una herramienta de aprendizaje (*corpora with learners*).

i. Los corpus como fuente de información

La primera orientación se centra en el uso de los corpus como repositorio de datos de uso al que investigadores, docentes, examinadores o autores de materiales pueden acudir para acceder a manifestaciones reales y contextualizadas del discurso producido por nativos o por aprendientes con la finalidad de obtener información cuantitativa y cualitativa sobre una amplia variedad de fenómenos lingüísticos.

En su uso más básico, los corpus textuales sirven primariamente para la extracción de muestras auténticas de lengua con las que ilustrar contenidos teóricos en gramáticas generales y gramáticas pedagógicas, o con las que diseñar actividades en libros de texto y materiales complementarios de español L2/LE. En la Nueva gramática de la lengua española (Academia 2009) se combinan ejemplos contruidos por sus redactores con otros procedentes de textos de los corpus académicos CORDE, CDH, CREA y CORPES, y en algunas gramáticas

pedagógicas se incluyen como ilustración fragmentos extraídos de diferentes corpus de español (cf. Penadés Martínez *et al.* 2008).

Una forma más elaborada de aprovechamiento de los corpus como fuente de información incluye la obtención de datos reales de frecuencias de uso de determinadas estructuras gramaticales, elementos léxicos (palabras, colocaciones) y unidades discursivas (marcadores, conectores), con el fin de manejarlos en la confección de programas de enseñanza y de materiales didácticos. La disponibilidad creciente de corpus textuales enriquecidos con información léxica y morfosintáctica propicia innovaciones metodológicas en el uso y explotación de los datos, ya que se multiplican las opciones de consulta al combinar las características incorporadas en los metadatos con las clases lingüísticas y unidades determinadas por el sistema de etiquetas y el lexicón.

En el apartado anterior se aludió a las posibilidades que ofrecen el CREA, el CORPES y el CdEGH de recuperar categorías gramaticales, como palabras de una cierta clase o determinadas formas de un paradigma flexivo, sin restricciones léxicas. La flexibilidad de las opciones de búsqueda de un corpus anotado gramaticalmente como el CORPES permite, por ejemplo, obtener todas las secuencias en las que los adverbios terminados en *-mente* van seguidos de un adjetivo cualquiera. Esta consulta, restringida para una ilustración rápida a los textos orales de América, devuelve 328 casos en 143 documentos. El cuadro 1 recoge los primeros ítems de la lista de concordancias resultante, que muestra asociaciones léxicas frecuentes de la construcción adverbio_{MODIFICADOR} – adjetivo_{NÚCLEO} (*completamente libres, verdaderamente importante*) junto con algunas ocurrencias de adverbio seguido de adjetivo que no corresponden a una estructura de MODIFICADOR + NÚCLEO (v. gr. en “las enfermedades oncológicas son *aproximadamente responsables* del treinta por ciento de la muerte y enfermedad”), lo que da pie a introducir nuevas reflexiones lingüísticas de interés para el aprendizaje del español.

ruinas de Pachacamac / ahm	totalmente	inspirada	va ser la cosa / digo yo
después las enfermedades aah oncológicas // que son	aproximadamente	responsables	del treinta por ciento de la muerte y enfermedad
no se puede simplemente / crear centros y dejarlos	completamente	libres	
la presencia léxica americana / es /	verdaderamente	/ importante	
porque era / un léxico	absolutamente	obsoleto	/ un léxico muerto que no conocía nadie
las academias todas han hecho un trabajo /	verdaderamente	extraordinario	
los Cerros Tutelares de Bogotá están	totalmente	despejados	arriba
	totalmente	impreciso	Colombia / en la mitad va Colombia
	totalmente	lícita	/ la / la / la repetición es totalmente clara
la repetición es	totalmente	clara	/ él se levanta // las manos arriba
cuando se incrementará a	parcialmente	nublado	y con escasa ocurrencia de lluvias
y yo he entregado pruebas //	sumamente	/ graves	// graves contra algunos alcaldes //

el crecimiento de la economía estadounidense será	ciertamente	negativo	/ en el periodo julio a septiembre
---	-------------	----------	------------------------------------

Cuadro 1. Muestra parcial de los resultados del CORPES para la búsqueda de adverbios en *-mente* seguidos de adjetivo (Oral América).

Las áreas de especial dificultad en la enseñanza del español/LE constituyen un terreno idóneo para comprobar la utilidad de los corpus textuales anotados para el análisis lingüístico y su aplicación en el diseño curricular y en la elaboración de materiales. Los corpus no solo indican si una cierta unidad o combinación de unidades o categorías aparece registrada, sino que proporcionan información sobre la frecuencia de su uso en el conjunto de los textos o en un determinado subconjunto. Dado que las listas de frecuencias léxicas aportan un criterio objetivo para la selección del vocabulario, no es de extrañar que los diccionarios de frecuencias hayan sido un recurso tradicionalmente empleado con fines pedagógicos desde los inicios de la LC. Pero el interés de la frecuencia no se limita a la enseñanza del vocabulario en sentido estricto, sino que se extiende al estudio de las construcciones sintácticas de las que participan los elementos léxicos concretos.

Un corpus etiquetado morfosintácticamente como ESLORA permite, por ejemplo, obtener la lista de frecuencia de los adjetivos y participios que aparecen a continuación de cualquier forma del paradigma de los verbos *ser* y *estar*, como muestran los cuadros 2 y 3. Obviamente, una consulta tan simple y limitada contextualmente deja fuera datos de uso imprescindibles si se pretende abordar un análisis real de las construcciones de *ser* y *estar* con adjetivos y participios, puesto que habría que tener en cuenta también, entre otras varias posibilidades de realización sintagmática y secuencial, aquellas combinaciones en las que el adjetivo o participio va acompañado de un adverbio cuantificativo, como ocurre, por ejemplo, con *ser + muy + bonito*, que se registra en ESLORA con una frecuencia de 51 casos por millón de palabras, *ser + muy + bueno*, registrada con 46 casos por millón, *ser + más + fácil*, con 24 por millón o *estar + muy + contento* con 14 casos por millón. Con todo, la información recogida en los cuadros 2 y 3 no deja de ser una muestra objetiva de las diferencias constructivas de *ser* y *estar*, que junto con la posibilidad de acceso inmediato a las secuencias contextualizadas en la aplicación de consulta del corpus, constituye una base empírica útil para la enseñanza teórica y práctica de estas estructuras a los aprendientes de español LE.

	<i>ser</i> + adj/part	Casos por millón
1	pequeño	115
2	bueno	66
3	capaz	57
4	normal	52
5	difícil	51
6	mayor	47
7	distinto	47
8	malo	43
9	impresionante	36
10	fácil	32
11	raro	30
12	importante	27
13	joven	26

14	curioso	26
15	diferente	26
16	necesario	25
17	feliz	25
18	grande	25
19	bonito	23
20	*	23

Cuadro 2. Combinaciones más frecuentes de *ser* + adjetivo o participio en el corpus ESLORA (v. 2.0). Frecuencia total = 1706 casos.

	<i>estar</i> + adj/part	Casos por millón
1	acostumbrado	37
2	casado	36
3	contento	26
4	bueno	21

5	encantado	19
6	hecho	19
7	preparado	14
8	solo	14
9	cerrado	13
10	metido	12
11	loco	11
12	embarazado	10
13	cómodo	10

14	ocupado	9
15	sentado	9
16	seguro	9
17	feliz	9
18	libre	9
19	enfermo	8
20	muerto	8

Cuadro 3. Combinaciones más frecuentes de *estar* + adjetivo o participio en el corpus ESLORA (v. 2.0). Frecuencia total = 816 casos.

El uso de los datos auténticos y contextualizados que proporcionan los corpus anotados representa, por un lado, una rica base documental para el diseño e implementación de actividades por parte de los docentes y autores de materiales, y por otro, un estímulo para la reflexión y el aprendizaje autónomo de los estudiantes, que en función de su nivel y necesidades pueden acceder por sí mismos a una variedad cada vez más amplia de recursos lingüísticos representativos del uso real del español.

ii. Los corpus como herramienta de aprendizaje

El uso de los corpus como herramienta de aprendizaje hace hincapié en cómo debe llevarse a cabo el proceso de aprendizaje y entronca con el denominado *Data-Driven Learning* ('aprendizaje guiado por datos') (Johns 1997), que fomenta el desarrollo de mecanismos cognitivos (predecir, observar, pensar, razonar, analizar, interpretar, reflexionar, inferir deductiva o inductivamente, hipotetizar, etc.), incide positivamente en la motivación y promueve el aprendizaje autónomo.

Los aprendientes, solos o con ayuda de los docentes, emplean los corpus para avanzar en su proceso de aprendizaje. Para ello deben saber qué información contienen, cómo se organiza y cómo se puede extraer. Algunas muestras de este enfoque las encontramos en el material complementario para ELE elaborado por Nicolás Martínez y Hernández Toribio (2015) basado en C-Or-DiAL.

En la misma línea es relevante el estudio de Lu & Chu (2013), que muestra la necesidad de graduar el uso directo de los diferentes tipos de corpus según el nivel de competencia de los aprendientes de español. Sus conclusiones se basan en la evaluación realizada por un grupo de estudiantes taiwaneses de español tras hacer varias actividades con tres corpus, de L1, L2 y paralelo. La secuencia más recomendable es empezar con el corpus paralelo, seguir con el de aprendientes y finalizar con el de español L1.

Las posibilidades de explotación de los corpus como herramienta de aprendizaje son diversas. Dependiendo del tipo de corpus manejado, los aprendientes de una L1, L2/LE o LH pueden comprobar hipótesis sobre el funcionamiento del nuevo código a partir de corpus generales; aproximarse a las características de distintos tipos de textos (narrativos, académicos, ensayísticos, etc.) a partir de corpus especializados; analizar rasgos propios del español como L1 o L2/LE con fines profesionales o académicos a partir de corpus de lengua de especialidad. El manejo de corpus permite a los estudiantes identificar errores en sus producciones y autocorregirse y también establecer comparaciones entre usos de hablantes nativos y no nativos a partir de corpus generales y de aprendientes, o registrar y describir diferencias entre las interlenguas de aprendientes de una misma L2/LE con distintas lenguas maternas a partir de corpus de aprendientes multi-L1. Asimismo, en la enseñanza del español los corpus generales sirven para implementar el aprendizaje en el aula de las destrezas de comprensión, expresión e interacción escrita y oral. Utilizando estos recursos los estudiantes pueden registrar, describir y analizar características del uso escrito y oral de la lengua.

5. Direcciones futuras y conclusiones

En las últimas décadas la investigación en el campo de la enseñanza del español ha incorporado avances y mejoras sustanciales provenientes de la LC, innovaciones que se van reflejando progresivamente en la elaboración de materiales didácticos y en la práctica docente. Además de lo expuesto en los apartados precedentes, es previsible que en un futuro inmediato las sinergias entre ambos enfoques impulsen el desarrollo de líneas de trabajo incipientes o totalmente nuevas. A continuación se resumen algunas de las perspectivas que se estiman más fructíferas en el aprovechamiento de los corpus para la enseñanza de la lengua.

En primer lugar, cabe esperar la incorporación efectiva a la enseñanza de los hechos de variación lingüística en sus distintas manifestaciones. La inclusión de materiales cada vez más diversos en los corpus permite ofrecer una base sólida para prestar la atención necesaria a la variación geográfica, social y estilística del español. La enseñanza del español como L1/LE/L2/LH sitúa a los agentes implicados, y en especial a docentes y formadores, ante el dilema de decidir qué variedad o variedades de la lengua deben ser objeto preferente de atención en el aula (*cf.*, por ejemplo, Blake y Zyzik 2016, cap. 6). Sea cual sea la solución adoptada, es preciso contar con datos de uso auténtico en cantidad y variedad suficientes para abordar una enseñanza empíricamente fundamentada.

La composición de los grandes corpus textuales (CREA, CORPES XXI, CdEGH) presenta un sesgo marcado hacia la representación del modo escrito y la norma o normas de prestigio que se le suelen asociar, el llamado estándar o norma culta. El peso mucho menor del habla, y especialmente del habla conversacional, en dichos corpus no se compensa con la existencia de otros exclusivamente orales, que, salvo excepciones, son de pequeño tamaño y de ámbito más restringido. Sin embargo, son precisamente estos últimos los que pueden aportar datos para que los estudiantes se familiaricen con la amplia gama de variedades geográficas, sociales y situacionales del español, una lengua hablada como L1 por muchos millones de personas en más de veinte países. Teniendo en cuenta que las mayores diferencias entre variedades se observan en el uso oral, ya que la escritura cumple una función uniformizadora, se hace aun más patente la necesidad de acceder a los recursos que aportan materiales de habla para estudiar sistemáticamente las características fónicas, léxicas, gramaticales y pragmáticas de los diferentes geolectos, sociolectos y registros en los que se manifiesta el español.

Independientemente del modelo de lengua preferente en cada contexto de enseñanza, el desarrollo adecuado de las destrezas comunicativas debe prestar atención a la mejora de la competencia receptiva hacia variedades diferentes de la dominante. La globalización de las comunicaciones y el uso generalizado de internet fomentan la conciencia de la diversidad lingüística, pero no son suficientes para alcanzar un conocimiento explícito del valor y distribución de las variantes que sí se puede lograr, en cambio, a través del empleo de los corpus. Por ejemplo, CREA, CORPES y CdEGH permiten condicionar las consultas con criterios geográficos, y aunque las áreas son excesivamente amplias para estudios especializados, sí resultan útiles para una aproximación general a diferentes fenómenos de variación, especialmente léxica y gramatical. Siguiendo con el ejemplo de la variación espacial, si se requieren datos específicos sobre el uso oral en diferentes ciudades tanto de España como de América, los materiales de entrevistas semidirigidas reunidos en PRESEEA son una buena fuente de información. La aplicación de consulta permite búsquedas ortográficas en parte de las transcripciones de cada uno de los corpus y ofrece también la posibilidad de consultar de forma independiente las grabaciones de audio, si bien estas no están alineadas con los textos. Además, dado que los corpus de PRESEEA recogen en los

metadatos referidos a los hablantes la diferencia entre mujeres y hombres y la información de su edad y el nivel de estudios alcanzado, es posible condicionar las consultas según estos parámetros y analizar, por tanto, cómo se reflejan esas variables sociolingüísticas en el uso registrado en las entrevistas.

En cuanto a la variación situacional o diafásica, junto con la tipología textual de los corpus de referencia, para la enseñanza de español LE/L2 resulta de especial interés el acceso a los registros de carácter más informal y conversacional, escasamente representados en los corpus generales. Además del C-Or-DiAL y del Columbia ya mencionados, reúnen materiales coloquiales, entre otros, los corpus Valesco y ESLORA, que comparten el objetivo de documentar el español hablado en comunidades bilingües, tradicionalmente desatendido en los estudios de dialectología española. El multilingüismo y la alternancia de lenguas en el ámbito hispánico no son en absoluto hechos excepcionales, por lo cual resulta de gran interés la recopilación de interacciones orales en situaciones de contacto lingüístico, de las que ofrecen también buenas muestras los corpus construidos en época reciente en Nuevo México (NMSEB) y Texas (Corpus del Español en Texas / Spanish in Texas Corpus).

En segundo lugar, es previsible que el desarrollo creciente de corpus multimodales abra nuevas perspectivas en el ámbito de la enseñanza. El empleo de materiales auténticos, con acceso a la imagen y el sonido de eventos comunicativos reales, enriquecerá progresivamente los recursos didácticos. Así, el uso creciente de corpus orales alineados contribuirá a la mejora de las destrezas orales y a la elaboración de estudios comparados sobre los distintos niveles lingüísticos, incluyendo el componente prosódico (pensemos, por ejemplo, en el campo de la enseñanza de la entonación del español a sinohablantes). Para ello es preciso que los corpus faciliten un acceso cómodo y amplio a los datos, lo que requiere herramientas computacionales cada vez más sofisticadas. Por otra parte, como ya se indicó en el apartado 3, la utilización de los corpus no tiene por qué limitarse al objetivo con que cada uno de ellos fue creado, sino que puede ampliarse de forma ilimitada con las herramientas de anotación idóneas y los estándares adecuados. La reutilización y aprovechamiento de los corpus de diferente tipo para la enseñanza es una línea de trabajo que se beneficia de los continuos avances en el conocimiento que se producen en el ámbito del procesamiento del lenguaje natural (PLN).

Una tercera línea de conexión entre la LC y la enseñanza se relaciona con la implementación de la docencia de español mediante la aproximación del *Data-Driven Learning* (DDL). Se ha señalado que este método aplicado al aprendizaje potencia la autonomía y la motivación del alumno, ya que favorece una práctica lingüística significativa al trabajar e investigar sobre muestras auténticas de lengua (cf. Gilquin y Granger 2010). El desarrollo del enfoque DDL es, no obstante, todavía incipiente en español. Será preciso abordar estudios específicos de aplicación del uso directo de los corpus en el aula para superar las reservas de los docentes hacia esta práctica y aprovechar todas sus posibilidades para el aprendizaje (cf. Timmis 2015; Lenko-Szymanska y Boulton 2015; Crosthwaite 2019).

La LC constituye un enfoque de gran relevancia potencial en la enseñanza del español, si bien se requieren todavía avances en varias direcciones para favorecer un uso generalizado y un aprovechamiento efectivo de los diferentes corpus disponibles. Por una parte es preciso contar con suficientes corpus a disposición pública en las condiciones adecuadas de variedad y representatividad. El acceso abierto y universal a los materiales recopilados y anotados, en la medida en que sea aceptable legal y éticamente, permite “democratizar” la información de los corpus para que pueda ser compartida tanto por investigadores y docentes como por estudiantes, tanto por hablantes de español L1 y LH como por hablantes de LE/L2, estimulando así la cooperación entre los diferentes usuarios. Por otra parte, los corpus han de ofrecer interfaces de consulta de uso ágil e intuitivo que permitan rentabilizar al máximo la

información que contienen. La utilidad pública de un corpus no depende únicamente de la cantidad de información incorporada –metadatos, etiquetación, anotación, alineación–, sino en gran medida de los instrumentos que proporciona para recuperar dicha información de forma flexible y adaptada a cada finalidad. Por último, es imprescindible introducir en la LC la perspectiva de los docentes, aprendientes, investigadores, autores de materiales y formadores, para lo cual deben establecerse canales de comunicación entre los lingüistas de corpus y los participantes en los procesos de enseñanza-aprendizaje que permitan a los primeros conocer las necesidades y objetivos de los segundos. Complementariamente, es también indispensable promover el conocimiento y la difusión del uso de los corpus y sus múltiples posibilidades de explotación entre los agentes de la enseñanza del español.

Lecturas complementarias recomendadas

- Crosthwaite P. (2019) *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. El volumen muestra el interés y las posibilidades de uso de los métodos del *data driven learning* a diversas lenguas y contextos docentes, con la novedad de que las aportaciones se aplican a los niveles de enseñanza primaria y secundaria.
- O’Keeffe, A., M. J. McCarthy y R. Carter (2007) *From Corpus to Classroom: Language Use and Language Teaching*. El libro relaciona de una manera fructífera los avances de la lingüística de corpus con la enseñanza de lenguas en diferentes niveles y con perspectivas complementarias. Entre los temas tratados destacan el aprendizaje del léxico y las expresiones fraseológicas en contexto, las relaciones entre gramática, discurso y pragmática, y el discurso especializado.

Recursos electrónicos citados

- ACTRES: Análisis Contrastivo y Traducción Inglés-Español, Corpus Paralelo P-ACTRES 2.0. <https://actres.unileon.es>.
- C-Or-DiAL: Corpus Oral Didáctico Anotado Lingüísticamente. <http://lablita.dit.unifi.it/corpora/cordial>.
- C-ORAL-ROM: Integrated reference corpora for spoken romance languages. <http://www.llf.uam.es/ESP/Coralrom.html>.
- CAES: Corpus de Aprendices de Español. <http://galvan.usc.es/caes>.
- CdEGH: Corpus del Español: Genre/Historical.. <https://www.corpusdelespanol.org/hist-gen/>.
- CDH: Corpus del Nuevo diccionario histórico del español. <https://www.rae.es/recursos/banco-de-datos/cdh>.
- CEDEL2: Corpus Escrito del Español como L2. <http://cedel2.learnercorpora.com/>.
- CINEAS: Corpus de INterlengua Española de Aprendices Sinohablantes. <https://cineas.udl.cat/>.
- COLA: Corpus Oral de Lenguaje Adolescente. <http://www.colam.org/>.
- Columbia: Columbia Corpus de Conversaciones para ELE. <https://edblogs.columbia.edu/corpusdeconversaciones/>.
- CORDE: Corpus Diacrónico del Español. <http://rae.es/recursos/banco-de-datos/corde>.
- CORELE: Spanish Learner Oral Corpus. http://cartago.llf.uam.es/corele/home_en.html.
- CORPES: Corpus del Español del Siglo XXI. Real Academia Española. <http://rae.es/recursos/banco-de-datos/corpes-xxi>.
- Corpus del Español en Texas / Spanish in Texas Corpus. <http://corpus.spanishintexas.org>.
- CREA: Corpus de referencia del español actual. <http://rae.es/recursos/banco-de-datos/crea>.
- El Grial: Interfaz computacional para anotación e interrogación de corpus en español. <http://www.elgrial.cl/index2015.php>.

[Borrador] Vázquez Rozas, Victoria y Blanco, Marta. Corpus y enseñanza del español Parodi, Giovanni, Lewis C. Howe y Pascual Cantos (eds.): *Lingüística de corpus en español/ The Routledge Handbook of Spanish Corpus Linguistics*. Londres y Nueva York: Routledge, 2022. 342-356. DOI:10.4324/9780429329296-26

ESLORA: Corpus para el estudio del español oral. <http://eslora.usc.es>.

Iberia: Corpus textual panhispánico de discurso científico-técnico.

<http://www.investigacion.cchs.csic.es/elci/node/8>.

IULA Corpus: Corpus técnico del Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra. <http://bwananet.iula.upf.edu/indexes.htm>.

NMSEB: New Mexico Spanish-English Bilingual corpus. <https://nmcode-switching.la.psu.edu/nmseb>.

PRESEEA: Proyecto para el Estudio Sociolingüístico del Español de España y América. <http://preseea.linguas.net/>.

SPLLOC: Spanish Learner Language Oral Corpora. www.splloc.soton.ac.uk.

Val.Es.Co: Corpus de conversaciones coloquiales. <http://www.valesco.es/corpus>.

Referencias citadas

- Academia, Real ____ Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Albelda, M. y A. Briz. 2017. "Teaching Spanish pragmatics through colloquial conversations". *MarcoELE* 25. https://marcoele.com/descargas/25/albelda-briz-colloquial_conversations.pdf.
- Aston, G., S. Bernardini y D. Stewart, eds. 2004. *Corpora and language learners*. Amsterdam: John Benjamins.
- Belpoliti, F. y E. Gironzetti. 2019. "Habla de herencia". En *The Routledge handbook of Spanish language teaching: metodologías, contextos y recursos para la enseñanza del español L2*, eds. J. Muñoz-Basols, E. Gironzetti y M. Lacorte, 447-462. Londres y Nueva York: Routledge.
- Blake, R. J. y E. C. Zyzik. 2016. *El español y la lingüística aplicada*. Washington D.C.: Georgetown University Press.
- Caballero Benavente, M., L. Díaz y M. Taulé. 2012. "Cómo prever y organizar los contenidos lingüísticos y sociopragmáticos de un escenario comunicativo en clase de E/LE a partir de un corpus". En *Monográfico. Estudios sobre aprendizaje y adquisición de lenguas y sus contextos*, ed. L. Díaz Rodríguez, 6-26. *MarcoELE* 15. <http://marcoele.com/descargas/15/1.aprendizaje-adquisicion.pdf>.
- Campillos Llanos, L. 2014. "A Spanish learner oral corpus for computer-aided error analysis". *Corpora* 9(2): 207-238.
- Campillos, L., P. Gozalo y A. Moreno. 2007. "El corpus C-ORAL-ROM en la enseñanza de ELE". En *Actas del XVII congreso internacional de ASELE: Las destrezas orales en la enseñanza del español*, ed. E. Balmaseda, 1115-1128. Logroño: Servicio de Publicaciones de la Universidad de La Rioja.
- Cao, S., I. da Cunha y M. Iruskieta. 2016. "A Corpus-based Approach for Spanish-Chinese Language Learning". En *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, eds. H.-H. Chen, Y.-H. Tseng, V. Ng y X. Lu, 97-106. Osaka: The COLING 2016 Organizing Committee.
- Colen Meniconi, F. 2009. "Los directivos en los libros didácticos de ELE". *MarcoELE* 9: 1-13. http://marcoele.com/descargas/9/colen_directivos.pdf
- Crosthwaite P. ed. 2019. *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Londres: Routledge.
- Cruz Piñol, M. 2012. *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco/Libros. 2ª ed. 2017.

[Borrador] Vázquez Rozas, Victoria y Blanco, Marta. Corpus y enseñanza del español Parodi, Giovanni, Lewis C. Howe y Pascual Cantos (eds.): *Lingüística de corpus en español/ The Routledge Handbook of Spanish Corpus Linguistics*. Londres y Nueva York: Routledge, 2022. 342-356. DOI:10.4324/9780429329296-26

- Doval, I. y M. T. Sánchez Nieto, eds. 2019. *Parallel Corpora for Contrastive and Translation Studies: New resources and applications*. Amsterdam: John Benjamins.
- Gabrielatos, C. 2005. "Corpora and language teaching: Just a fling or wedding bells?". *Teaching English as a Second Language, Electronic Journal* 8(4): 1-35.
- García Salido, M. y M. García. 2017. "Comparing learners' and native speakers' use of collocations in written Spanish". *International Review of Applied Linguistics in Language Teaching (IRAL)* 56(4): 401-426.
- Gilquin, G y S. Granger. 2010. "How can data-driven learning be used in language teaching?". En *The Routledge Handbook of Corpus Linguistics*, eds. A. O'Keeffe y M. McCarthy, 359-370. Londres y Nueva York: Routledge.
- Johns, T. 1997. "Contexts: The Background, Development and Trialling of a Concordance-based CALL Program". En *Teaching and Language Corpora*, eds. A. Wichmann, S. Fligelstone, T. McEnery y G. Knowles, 100-15. Londres: Longman.
- Jørgensen, A. M. e I. Olza Moreno. 2019. "Phraseology in teenage language in Spanish, English and Norwegian: Notes on a number of fixed expressions that articulate disagreement". *Languages in Contrast: International Journal for Contrastive Linguistics* 20(1): 58 – 83.
- Krashen, S. D. 1988. *Second Language Acquisition and Second Language Learning*. Londres: Prentice Hall.
- Labrador, B. 2015. "Translations as an aid to ELT: Using an English-Spanish parallel corpus. P-ACTRES) to study English *both* and its Spanish counterparts". *Digital Scholarship in the Humanities* 30: 71-87.
- Leech, G. 1997. "Teaching and language corpora: a convergence". En *Teaching and language corpora*, eds. A. Wichmann, S. Fligelstone, T. McEnery y G. Knowles, 1-23. Londres: Longman.
- Lenko-Szymanska, A. y A. Boulton, eds. 2015. *Multiple Affordances of language Corpora of Data-driven Learning*. Amsterdam: John Benjamins.
- Lozano, C. y A. Mendikoetxea. 2013. "Learner corpora and SLA: the design and collection of CEDEL2", en *Automatic Treatment and Analysis of Learner Corpus Data*, eds. A. Díaz-Negrillo, N. Ballier y P. Thompson, 65-100. Amsterdam: John Benjamins,
- Lu, H.-C. y Y.-H. Chu. 2013. "Evaluation of Corpus-Assisted Spanish Learning". En *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*, 467-473. Taipei: Department of English, National Chengchi University.
- McEnery, T. y R. Xiao. 2010. "What corpora can offer in language teaching and learning". En *Handbook of Research in Second Language Teaching and Learning*, Vol. 2, ed. E. Hinkel, 364-380. Londres y Nueva York: Routledge.
- Mendikoetxea, A. 2014. "Corpus-based Research in Second Language Spanish". En *The Handbook of Spanish Second Language Acquisition*, ed. K. L. Geeslin, 11-29. Oxford: Wiley-Blackwell.
- Mitchell, R., L. Domínguez, M. Arche, F. Myles y E. Marsden. 2008. "SPLLOC: A new database for Spanish second language acquisition research". *EuroSLA yearbook* 8(1): 287-304.
- Nicolás Martínez, C. 2012. *C-Or-DiAL. Corpus Oral Didáctico Anotado Lingüísticamente*. Madrid: Liceus.
- Nicolás Martínez, C. y M. I. Hernández Toribio. 2015. *Del oído al habla*. Barcelona: Octaedro.
- O'Keeffe, A., M. J. McCarthy y R. Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

[Borrador] Vázquez Rozas, Victoria y Blanco, Marta. Corpus y enseñanza del español Parodi, Giovanni, Lewis C. Howe y Pascual Cantos (eds.): *Lingüística de corpus en español/ The Routledge Handbook of Spanish Corpus Linguistics*. Londres y Nueva York: Routledge, 2022. 342-356. DOI:10.4324/9780429329296-26

- O’Keeffe, A. y M. McCarthy, eds. 2010. *The Routledge Handbook of Corpus Linguistics*. Londres y Nueva York: Routledge.
- Palacios, I. M., F. M. Barcala y G. Rojo. 2019. “El *Corpus de Aprendices de Español*. CAES) y sus aplicaciones para la enseñanza/aprendizaje del español como lengua extranjera”. En *Corpus y construcciones. Perspectivas hispánicas*, Anexo 79 de *Verba*, eds. M. Blanco, H. Olbertz y V. Vázquez Rozas, 273-301. Santiago de Compostela: Universidade de Santiago de Compostela.
- Parodi, G., ed. 2007. *Lingüística de corpus y discursos especializados: puntos de mira*. Valparaíso: Ediciones Universitarias de Valparaíso.
- Parodi, G. y G. Burdiles. 2019. “Corpus y bases de datos”. En *The Routledge handbook of Spanish language teaching: metodologías, contextos y recursos para la enseñanza del español L2*, ed. J. Muñoz-Basols, E. Gironzetti y M. Lacorte, 596-612. Londres y Nueva York: Routledge.
- Penadés Martínez I., M. Martí Sánchez y A. M. Ruiz. 2008. *La gramática española por niveles. De inicial a superior en un solo volumen*: Madrid: Edinumen.
- Poole, R. 2018. *A guide to using corpora for English language learners: empowering language learners to become independent learners through corpus-aided discovery and investigation*. Edimburgo: Edinburgh University Press.
- Rabadán, R. 2019. “Working with parallel corpora: Usefulness and usability”. En *Parallel corpora for contrastive and translation studies: new resources and applications*, eds. I. Doval y M. T. Sánchez Nieto, 57-78. Amsterdam: John Benjamins.
- Ramón, N. 2015. “Comparing original and translated Spanish: A corpus-based analysis of adjective position”. *Babel* 61(4): 527-551.
- Rojo, Guillermo. 2015. “Hispanic Corpus Linguistics”. En *The Routledge Handbook of Hispanic Applied Linguistics*, ed. M. Lacorte, 371-387. Londres y Nueva York: Routledge.
- Sánchez Rufat, A. 2016. “El uso transitivo y ditransitivo de *dar* en un corpus escrito contrastivo”. *Revista Fuentes* 18(2): 117-133.
- Sinclair, J. ed. 2004. *How to use corpora in language teaching*. Amsterdam: John Benjamins.
- Timmis, I. 2015. *Corpus Linguistics for ELT. Research and Practice*. Londres y Nueva York: Routledge.