

## Corpus y bases de datos (*Corpora and databases*)

Giovanni Parodi

*Pontificia Universidad Católica de Valparaíso, Chile*

Gina Burdiles

*Universidad Católica de la Santísima Concepción, Chile*

### 1. Introducción<sup>1</sup>

En los últimos años la enseñanza del español como segunda lengua (L2) se ha visto beneficiada enormemente por los desarrollos producidos tanto en lingüística teórica como en lingüística aplicada. En esta última línea, resultan de particular interés los avances y recursos implementados en lingüística de corpus y lingüística computacional, los cuales han posibilitado –entre otros– aproximarnos de modo más certero y fidedigno a los usos lingüísticos en situaciones reales, así como disponer de datos empíricos acerca de, por ejemplo, la variación de registros y géneros del discurso. Al mismo tiempo, estos avances han permitido el diseño e implementación de sofisticados sistemas computacionales como herramientas de gran utilidad para el apoyo de la investigación y de la docencia en procesos de enseñanza/aprendizaje de lenguas. En este contexto, este capítulo tiene como objetivo ofrecer un análisis crítico del panorama de los corpus y de las bases de datos disponibles para los procesos de enseñanza/aprendizaje del español L2. Para cumplir este objetivo, daremos cuenta de iniciativas que se ofrecen a modo de herramientas y recursos en línea, tales como corpus *de* y *para* aprendientes con soporte computacional y, de preferencia, disponibles en línea. Al mismo tiempo, y a modo de proyecciones, se esbozan pautas y sugerencias para la incorporación de estos recursos tanto en la investigación aplicada como en las actividades del aula.

Palabras clave: Corpus de aprendientes, base de datos, enseñanza del español, programas de enseñanza en línea, recursos lingüísticos

In recent years the teaching of Spanish has been greatly benefited by developments both in theoretical linguistics and applied linguistics. In this last line, advances and resources in corpus linguistics and computational linguistics are of particular interest because they have enabled a more accurate and reliable approach to language use in real situations. At the same time, they have offered the possibility to access robust empirical data on, for example, the variation in registers and discourse genres. These advances have allowed the design and implementation of sophisticated computer systems as useful tools to support research and teaching processes of languages. In this context, this chapter aims to provide a critical analysis of the panorama of the corpus and databases available concerning the processes of the teaching/learning of Spanish. To meet this objective, we will review some initiatives that are offered as on-line tools and resources, such as learner corpus based on computer support and preferably available on line. We also offer a critical analysis of the kind of corpora described. As part of the projections, teaching guidelines and suggestions of how to employ these resources as part of applied research and classroom activities are advanced.

Keywords: Learner corpus, data base, Spanish language teaching, on-line teaching programs, linguistic resources

## 2. Estado de la cuestión

En la actualidad, desde la perspectiva de la lingüística de corpus con soporte computacional, existe un interés creciente por desarrollar estudios innovadores basados o guiados por grandes muestras de textos auténticos cuyos fines son servir de apoyo al quehacer científico y didáctico de la enseñanza de lenguas (De Kock 2001a, Parodi 2007, Cruz 2012, Parodi 2015a, Parodi y Burdiles 2015). Esto ha derivado en un fortalecimiento del enfoque empírico con corpus de textos de variada naturaleza y procedencia y, junto a ello, ha generado un vertiginoso movimiento preocupado por la creación e implementación de herramientas computacionales en línea al servicio de estrategias didácticas cada vez más eficaces (De Kock 2001b, Bernardini 2004, Drange 2008, Campoy, Belles-Fortuno y Gea-Valor 2010, Ferreira, Salcedo, Kotz y Barrientos 2012).

En el vértice entre estas diversas vertientes, la enseñanza del español L2 se ha beneficiado de formidables avances teóricos, empíricos y tecnológicos. Así, la perspectiva educativa de aprovechar los materiales lingüísticos auténticos compilados en los corpus se inspiró en la propuesta original de Johns (1986), que se basaba en algunos de los principios clásicos de la lingüística de corpus. A partir de un conciso artículo, Johns (1986) sostenía que una de las ocupaciones de los aprendientes de una L2 era hacerse cargo del descubrimiento de la lengua meta, ya fuera en situaciones de clases o fuera de ellas (Tribble y Johns 1990). Este enfoque se ha conocido como *aprendizaje guiado por datos (data-driven learning)*, y supone que el computador cumple el rol de un informante capaz de apoyar el aprendizaje por medio de textos o fragmentos textuales originales, todo ello con el objetivo “of contextualising and demythologizing the language and of making available to the learner information about authentic usage” (Johns 1994, 296). Aunque inicialmente estos desarrollos estaban fuertemente orientados hacia el aprendizaje del inglés como segunda lengua, recientemente se detecta un foco en otras lenguas, tal como el español.

La potente idea de expandir el campo de utilización de los corpus más allá de los laboratorios de investigación y emplearlos en la enseñanza de lenguas no solo se deslinda de los fines contemplados originalmente para estos en la lingüística de corpus, sino que ha alcanzado importantes desarrollos en los llamados *corpus de/para aprendientes (learner corpora)* (Sinclair 2004; Aijmer 2009; Granger, Gilquin y Meunier 2013). Este aprovechamiento de los corpus para la enseñanza de una lengua, en palabras de Flowerdew (2015, 15), ha inspirado “a plethora of data-driven learning (DDL) initiatives across the globe”. Tanto interés se ha suscitado que, incluso en los últimos años, se ha constituido una asociación internacional que reúne a los investigadores del área: *Learner Corpus Association, LCA* (<http://www.learnercorpusassociation.org>).

Ahora bien, en la expresión del inglés que da origen a este enfoque existe un aspecto implícito que, al traducir la expresión al español, es necesario explicitar debido a la relevancia que representa: la procedencia o audiencia ideal de los corpus; esto es, si son corpus *de* aprendientes o corpus *para* aprendientes. En el primer caso, corpus *de*

aprendientes, se consideran materiales registrados por hablantes, preferentemente, no nativos de español que se encuentran en alguna fase de adquisición/aprendizaje de la lengua. El correspondiente empleo de estos corpus no se encuentra limitado y puede potencialmente utilizarse tanto por investigadores interesados en algún asunto particular como por estudiantes que buscan apoyo para su propio proceso de aprendizaje o, incluso, por docentes que aprovechan estas instancias de material auténtico para su explotación didáctica. En un segundo caso, cuando la traducción explicita la otra opción preposicional y llegamos al término corpus *para* aprendientes, ciertamente las posibilidades son un tanto más reducidas. Estos corpus deberían, conceptualmente hablando, estar definitivamente orientados a los estudiantes de la lengua meta del corpus y, por tanto, estar organizados de modo tal que constituyan un aporte significativo para el aprendizaje, orientando de modo debido al aprendiente. Por supuesto, que estas dos interpretaciones no son excluyentes y un corpus que se declara *de* aprendientes puede ser también un corpus *para* aprendientes. Desde una perspectiva más integradora, junto a estas dos alternativas de traducción, también es posible pensar en una tercera opción, tal como corpus *con* aprendientes. En ella sería posible imaginar la idea de construir corpus con estudiantes como participantes de la actividad y que, al mismo tiempo, también pudieran sacarles provecho para su propio aprendizaje.

### 3. Consideraciones metodológicas actuales

Con el objetivo de construir la panorámica prometida, en lo que sigue, describimos brevemente y clasificamos una muestra de corpus y bases de datos disponibles en línea, corpus *de/para* aprendientes. La búsqueda de estos materiales se orientó bajo el supuesto de consignar todo corpus que declare registrar usos del español y que ofrezca muestras orales o escritas con soporte tecnológico de diversa índole y grado de desarrollo. Para llevarla a cabo de modo amplio y versátil, se empleó diversos buscadores (Google, Altavista, Yahoo, entre otros) y se puso en práctica un conjunto de palabras clave, tanto en español como en inglés (corpus de aprendientes, corpus español, *learner corpora*, etc.). Así, se llegó a identificar una muestra compuesta por 39 corpus y bases de datos en funcionamiento en la actualidad y que cumplieran con el requisito de ofrecer, en general, corpus de español y, en particular, corpus *de/para* aprendientes. Cabe destacar, entonces, que esta muestra es aleatoria e intencionada y no pretende ser representativa ni exhaustiva.

Ahora bien, es relevante consignar que los términos *corpus* y *base de datos* no siempre presentan una definición operacional precisa y, en muchos contextos, se emplean indistintamente. En este capítulo, luego de realizada la búsqueda aleatoria, no estableceremos una distinción entre ambos.

Para organizar el análisis de los corpus recolectados, proponemos una clasificación en torno a dos criterios: modo (lengua oral y/o escrita) y procedencia (L1 o L2). Lo que buscábamos era una forma de hacer emerger agrupaciones y poder, de este modo, observar tendencias. Como se aprecia en la Tabla 1, la aplicación de estos criterios a los materiales encontrados nos condujo a cinco categorías.

**Tabla 1.** Corpus disponibles de español

<b>Corpus escritos EL1:</b> Base de Datos Sintácticos del Español Actual (BDS), Biblioteca Virtual Miguel de Cervantes, Corpus del Español (Mark Davis), CREA, CORDE, Corpus Diacrónico y Diatópico del Español de América (CORDIAM), Corpus Dinámico del Castellano de Chile (Codicach), esTenTen corpus, esAmTenTen corpus, El Grial, IULA Spanish LSP Treebank, Environment Spanish Monolingual Corpus (PANACEA), RST Spanish Treebank, Spanish Online (SOL), UAM Spanish Treebank, Corpus del Nuevo diccionario histórico (CDH).
<b>Corpus orales EL1:</b> Corpus Digital Giennense, Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC), Corpus Oral Didáctico Anotado Lingüísticamente (C-Or-DiAL), Corpus Oral de Lenguas Romances (C-ORAL-ROM), Corpus Oral del Lenguaje Adolescente (COLA), Corpus oral para la fonética y las tecnologías del habla en español, PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y América), Corpus Val.Es.Co, Corpus Oral y Sonoro del Español Rural.
<b>Corpus escrito/oral EL1:</b> Corpus del Español del Siglo XXI (CORPES XXI), Corpus Cumbre, Corpus del Español Mexicano Contemporáneo (CEMC), Corpus del Español en el Sur de Arizona (CESA), Corpus del Español en Texas, Corpus Oral y Sonoro del Español Rural.
<b>Corpus orales EL2:</b> Corpus de Aprendices Taiwaneses de Español (CATE), Spanish Corpus Proficiency Level Training (SPT), Spanish Learner Language Oral Corpora (SPLLOC), Spanish Learner Oral Corpus (SLOC), Fono.Ele.
<b>Corpus escritos EL2:</b> CAES (Corpus de Aprendices de Español), Corpus Escrito del Español como L2 (CEDEL2), SFN Corpus.

Tal como se aprecia en la Tabla 1, el rastreo logró identificar 16 corpus escritos de español L1, 9 corpus orales de español L1 y otros 3 corpus que incluyen tanto oralidad como escritura. Esto quiere decir que, de los 39 corpus analizados como operativos en la actualidad, en esta muestra el 80% se concentra en español L1. Según estas cifras, un porcentaje, comparativamente, muy menor se centra en español L2 y, ciertamente, solo unos pocos abordan decididamente los procesos de enseñanza/aprendizaje de español L2 con actividades o ejercicios didácticos, orientados al aprendiente. Así, hemos logrado identificar 5 corpus para español oral en L2 y 3 para español escrito L2.

A partir de en estos datos, resulta evidente que, sea porque existe mayor interés en la escritura o porque pueda resultar más accesible y expedita la compilación de fragmentos o textos escritos completos, se han invertido mayores esfuerzos en registrar el modo escrito de la lengua que el oral. Incluso, un análisis no sistemático de los corpus aquí revisados, lleva a pensar que son más robustos y contienen mayores muestras de textos o fragmentos textuales los corpus escritos que los orales. Queda pendiente un estudio más pormenorizado de este rasgo.

A pesar de los avances significativos que se han llevado a cabo en el área en un periodo de tiempo relativamente breve, los datos de la Tabla 1 muestran que aún es escasa y –aparentemente– poco sólida la disponibilidad de corpus para aprendientes de español L2 con un claro y definitivo enfoque didáctico. Lo que se detecta es la existencia de diversos corpus disponibles en línea de hablantes de español –de preferencia– como L1 y en el modo

escrito. En menor medida se observan iniciativas orientadas al español L2. Junto a lo anterior, la explotación educativa de los corpus, que ofrezca oportunidades efectivas y contextualizadas de enseñanza/aprendizaje para el aprendiente se revelan, de hecho, comparativamente, aún muy reducidas. Son escasos los corpus en los que se detecta apoyo y desarrollo de actividades orientadas a potenciar la conciencia metalingüística y metacognitiva de la lengua o de los procesos discursivos relacionados con las habilidades lingüísticas.

Al mismo tiempo, este análisis preliminar permite identificar 3 corpus orales de la denominada lengua de herencia (LH), cuya importancia ha ido progresivamente en aumento, sobre todo en los Estados Unidos. En ellos se ofrecen preferentemente muestras del español hablado, tal como por ejemplo en el Corpus del Español en el Sur de Arizona (CESA), en el Corpus del Español en Texas, y en el Corpus Oral y Sonoro del Español Rural.

Ahora bien, aunque nuestro foco de interés inicial no estaba en hacer una descripción cuantitativa, el examen de la muestra de corpus revela que la disponibilidad como herramientas de análisis y búsqueda para soporte de investigación o de preparación de materiales didácticos es mayoritaria. Por ejemplo, el CAES exhibe –principalmente– orientación de corte investigativo para el lingüista o académico interesado en llevar a cabo estudios empíricos inspirados en principios de la lingüística de corpus (Parodi, 2015b), más que mostrar una preocupación decidida y manifiesta por procesos de enseñanza/aprendizaje. Su valor reside principalmente en contar con una diversa y rica muestra de textos producidos por hablantes de diversas lenguas maternas y clasificados por nivel de competencia lingüística. Esto es, tal como se declara explícitamente por sus creadores, el CAES constituye un excelente corpus *de* aprendientes, aunque es solo potencialmente una herramienta *para* aprendientes.

Una excepción a la mayoría lo constituye el Spanish Corpus Proficiency Level Training (SPT), desarrollado por la University of Texas at Austin para aprendientes de español L2, preferentemente para angloparlantes. El SPT se puede consultar a través de una página web en inglés que explícitamente busca apoyar el desarrollo de la habilidad de producción oral en español por medio de un corpus de videos de diversos hablantes no nativos de español en diversos niveles de desarrollo lingüístico y en temáticas variadas. Cuenta con ejercicios graduados y clasificados en torno a, por ejemplo, gramática, léxico y fluidez. También permite la autoevaluación y el acceso al conocimiento y a la comprensión de los diversos niveles de competencia del español L2. Además de ello, en el sitio se declara que el corpus también se ofrece como una herramienta útil para investigadores y estudiantes de lingüística, aunque sus recursos no están orientados claramente hacia ellos, como sí lo hace el CAES.

Un propósito similar al SPT es el que persigue el corpus Fono.Ele, desarrollado por la Universidad de Alcalá en colaboración con otras siete universidades internacionales. Su foco central es la pronunciación del español L2. A pesar de contar con un amplio corpus en línea compuesto por muestras de habla procedentes de estudiantes de seis nacionalidades diferentes y con diferentes niveles de competencia de español, el componente didáctico está escasamente explotado.

Así, según revela nuestra pesquisa, son escasas las posibilidades de encontrar corpus para aprendientes que ofrezcan al estudiante herramientas o actividades que apoyen su

desarrollo lingüístico en el español L2. De hecho, lo que se aprecia es más bien la constitución y disponibilidad de corpus para investigadores que para aprendientes. O, incluso, sería posible catalogar algunos de estos recursos como bases de datos, ya que –en muchos casos– lo que se ofrece es un compendio no claramente organizado de materiales lingüísticos, aunque –tal como ya se indicó– tal distinción no queda suficientemente clara.

Este balance, a partir de una muestra aleatoria y claramente no definitiva, pone de manifiesto el estado actual de la cuestión. En su mayoría, como se adelantó, los corpus se revelan *de* aprendientes y no *para* aprendientes. Incluso, entonces, resultan ser más bien para investigación, más que con fines didácticos. No obstante ello, esta relativa carencia de corpus para aprendientes encierra –al mismo tiempo– la oportunidad de un tremendo nicho y brinda posibilidades infinitas para generar adelantos tecnológicos como soporte para apoyar el desarrollo de las diversas habilidades lingüísticas de los estudiantes, considerando –incluso– las variaciones disciplinares y de géneros discursivos (Parodi y Burdiles, 2015; Parodi, 2015a).

Dada la imposibilidad de proporcionar en este espacio un completo panorama de los corpus y de las bases de datos disponibles del español, tanto en lengua materna como extranjera, se recomienda consultar Briz y Albelda (2009), Pitkowski y Vásquez (2009), Parodi (2007, 2010a, 2015b) y Rojo (2015, 2016). Aquí nos hemos concentrado en los sitios disponibles en línea que emergieron producto de nuestras definiciones de búsqueda (véanse los anexos 1 y 2).

#### **4. Pautas para la enseñanza**

A partir de los años noventa, los resultados de la investigación sobre la lengua basada en corpus comienzan, de manera cada vez más creciente, a ser aplicados a la enseñanza de lengua (Benett 2010; Flowerdew 2012). Surge de ahí un nutrido conjunto de trabajos que se han ocupado del uso de los corpus en la enseñanza de primeras o segundas lenguas, sobre todo en inglés. Entre ellos, destacan los incorporados a las publicaciones de Sinclair (2004), Aijmer (2009), Bennett (2010) y Flowerdew (2012). El amplio desarrollo de investigaciones sobre la enseñanza del inglés contrasta considerablemente, tal como vimos más arriba, con el estado de la cuestión en el caso de la lengua española. Por lo mismo, la investigación de que se dispone es comparativamente menor y son aún más escasos los trabajos con propuestas para la enseñanza de español L2 con uso de corpus. Algunos de ellos son los trabajos de De Kock (2001a, 2001b), Pérez-Ávila (2006), Drange (2008), Mendikoetxea (2013) y Bailini (2014).

Un dato que refrenda el exiguo desarrollo de trabajos que vinculan la enseñanza de español L2 con uso de corpus es la casi nula presencia de propuestas o experiencias pedagógicas presentadas en los congresos anuales de ASELE, Asociación para la Enseñanza del Español como Lengua Extranjera. En las actas de cada congreso, en las tesis y monografías publicadas en su sitio oficial, la temática que nos ocupa está casi ausente, con excepción de los trabajos de Sánchez, Bernardos y López (2001), Cestero, Penadés, Blanco, Camargo y Simón (2001), Albelda (2010), López-García (2013) y Nicolás (2013). No obstante, la celebración del XXVIII Congreso Internacional de ASELE en 2017, “Léxico y cultura en LE/L2: corpus y diccionarios”, en la Universitat Rovira i Virgili de

Tarragona (España), corrobora que se trata de un tema cada vez más presente en la disciplina.

Para orientar las pautas de enseñanza con corpus, nos guiamos por el trabajo de Leech (1997), quien propuso tres maneras de usar los corpus con finalidad educativa. Según este autor, se puede hacer uso directo de los corpus; en ese caso, profesor y alumno colaboran en el descubrimiento de las razones que subyacen a los usos observados directamente en los corpus. Se puede, asimismo, hacer uso de ellos de manera indirecta; en este caso, los corpus se emplean como fuente para la creación y publicación de recursos para la enseñanza. Leech (1997) planteó, además, una tercera forma de uso, la que pretendía superar, hace dos décadas, la escasez de corpus para enseñar lengua como L2, pues los existentes estaban principalmente concebidos para la investigación y la enseñanza de lengua para fines específicos. En consecuencia, propuso desarrollar adicionalmente corpus orientados de manera especial a la enseñanza de la lengua. Dada la existencia de los corpus disponibles actualmente, en lo que sigue, nos referiremos al uso directo e indirecto de corpus orales y escritos en el proceso de enseñanza de español L2. Al final del capítulo se incluyen dos anexos con más información sobre los corpus y bases de datos analizados.

#### **4.1. *Uso directo de los corpus***

El uso directo de los corpus en la enseñanza de la lengua, tanto de los corpus de hablantes nativos como de aprendientes de L2, permite responder a propósitos educativos inspirados en perspectivas tanto normativas como descriptivas. Ello, debido a que se puede motivar la observación de los casos para inducir el descubrimiento de las reglas del español y sus contextos de uso, al mismo tiempo que se puede proponer la observación de la variedad lingüística y la riqueza de alternativas. La utilización directa de corpus en el caso de la clase de lengua extranjera implica la explotación de esos materiales por parte de profesores y alumnos, en el examen de los usos reales de la lengua, de modo que la actividad de aprendizaje esté centrada en el estudiante, quien formula hipótesis acerca de lo observado y, en consecuencia, aprende descubriendo. Este trabajo mancomunado impone al docente, en un primer momento, el desafío de conocer los corpus disponibles, como los mencionados a manera de ejemplo en los anexos 1 y 2, y seleccionar el que estima pertinente para sus objetivos dentro de una unidad particular de enseñanza/aprendizaje.

En esta modalidad de uso directo de los corpus, el profesor es “a facilitator of the learning process, helping the learners to interpret the data, and giving them advice on how best to search the corpus and analyse their search results” (Chambers 2010, 355). Como mediador pedagógico, el profesor debe planificar las experiencias educativas en función de la edad de los alumnos, de su nivel de competencia en la lengua y con consideración de los fines que persigue el aprendizaje de lengua. En ese punto, una sugerencia inicial es que las decisiones que tome tengan en consideración las directrices del *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (Consejo de Europa 2002), el *Plan curricular del Instituto Cervantes* (Instituto Cervantes 2006) y las *Proficiency Guidelines* del ACTFL (ACTFL 2012).

Los corpus que se mencionan en este artículo están alojados en sitios web que se caracterizan por presentar instrucciones claras y por tener un diseño que facilita el acceso a

los estudiantes para que observen y descubran por sí mismos las características propias de la lengua. Sin embargo, se advierte también que en varios otros corpus, que se presentan y proponen como sitios de aprendizaje, se usa un lenguaje excesivamente técnico a la vez que las instrucciones de uso y la navegación pueden resultar, en ocasiones, complejas para el usuario.

Dadas estas condiciones, probablemente, no sería sencillo para un estudiante de español L2 desarrollar de manera autónoma tareas como las sugeridas por Tolchinsky (2014) en el uso directo de corpus; esto es, comprobar directamente diferencias en riqueza léxica, computando las relaciones *type/token*, observar densidad léxica, calcular la relación entre palabras contenido y palabras función; calcular la media de longitud de palabra y debatir las razones por las cuales ciertos textos contienen palabras más largas que otras, descubrir las colocaciones más frecuentes en cada tipo de texto o las palabras de contenido más frecuentes en cada uno, entre otras posibles actividades. La complejidad en el diseño del sitio donde esté alojado el corpus y los tecnicismos asociados a la realización de las búsquedas pueden constituir un obstáculo importante a la autonomía en el uso del corpus. Cabe agregar que, aun cuando el estudiante consiguiera realizar búsquedas de menor complejidad, la relevancia de los hallazgos también podría estar limitada si carece de una explicación complementaria que, por ejemplo, confiera sentido a una expresión mostrada en la página del sitio.

Partiendo del hecho de que algunos corpus son más complejos de usar, una primera pauta para desarrollar la enseñanza de lengua con apoyo de corpus es la introducción y enseñanza gradual –por parte del profesor– del uso de las herramientas necesarias hasta que el alumno alcance un nivel de independencia que le permita trabajar de manera autónoma. En atención a esa gradualidad, Fligelstone (1993) describe tres etapas en el proceso de aplicación de los corpus a la enseñanza de lenguas por parte del docente: enseñar acerca de los corpus, enseñar para explotar los corpus y explotarlos para enseñar. De acuerdo con esto, en un primer momento, se recomienda enseñar a los aprendientes acerca de qué y cómo son los corpus y a usarlos como una fuente cercana para proveer ejemplos. En un segundo momento, se propone enseñar a los estudiantes a explotar los corpus, proporcionándoles la metodología y las herramientas (computacionales, por ejemplo) para estudiarlos y analizarlos. Finalmente, y como una posible consecuencia de lo anterior, se puede llegar a promover la explotación de un corpus para responder a una tarea de aprendizaje o de investigación a partir de los datos que, sobre diversas áreas de estudio de la lengua, están disponibles en los corpus.

El uso directo de los corpus, en consecuencia, puede ser mucho más probable y exitoso una vez que el estudiante ha adquirido los conocimientos necesarios y las técnicas de estudio de la lengua basada en el corpus. Esto nos lleva a sostener que un número importante de los corpus en español que hoy están disponibles en la red de Internet puede beneficiar grandemente no solo a estudiantes de español L2, sino que también puede reforzar la formación profesional de profesores de lengua española, de traductores o de quien requiera profundizar en los usos del discurso especializado.

#### 4.1.1. *Corpus orales*



El *MCER* y el *PCIC* subrayan la importancia que tiene para el aprendiente de lengua extranjera el desarrollo de la competencia pragmática y la sensibilización sobre los registros de la lengua, especialmente si el estudiante no se encuentra en situación de inmersión (Secchi 2014). En este sentido, es recomendable trabajar con corpus orales y facilitar el aprendizaje de la lengua a través del uso de conversaciones reales que sean frecuentes en los encuentros comunicativos del día a día, donde se pueda comprender los hábitos comunicativos culturales que comparten los hablantes de la lengua que estudian. Las conversaciones coloquiales incorporadas a los corpus son particularmente valoradas porque esta es la modalidad discursiva más empleada y prototípica en la comunicación humana, y por ser, según Briz (1998) y Albelda (2010), la que se adquiere de manera más natural. Otra actividad recomendada con corpus orales de distintos géneros discursivos dialógicos, es promover, en la clase, el aprendizaje de valores y comportamientos culturales, la observación y práctica de la gestión de los turnos de habla y de otros intercambios comunicativos como saludar, despedirse, pedir disculpas, reformular lo dicho, sugerir o pedir algo. Dentro de los corpus considerados en nuestro rastreo, existen algunos que poseen mayor potencial para ser usados de la manera recién descrita, a saber: Corpus Oral y Sonoro del Español Rural, esTenTen corpus, Fono.ele, SFN Corpus, Cumbre, CORLEC, PRESEEA y SPT.

Dado que los corpus conversacionales en español permiten acercar al aprendiente a la cultura de los países hispanohablantes, se aconseja exponerlo a la mayor cantidad posible de variedades del español, no solo para que aprenda su uso, sino más bien para que él “tome conciencia de la diversidad en la unidad y de la riqueza que esto conlleva, ya que la inmensa comunidad idiomática de los hispanohablantes reúne a una multiplicidad de comunidades de habla con perfiles propios y distintivos” (Moreno 2000, 55). Al respecto, Drange (2008) empleó el Corpus Oral del Lenguaje Adolescente (COLA) en clase de español L2, para explicar y mostrar el uso de palabras y expresiones específicas, como el empleo de vocativos o apelativos (Stenström y Jørgensen 2008) y verbos de percepción con función apelativa, como *oye* (Drange 2008). La autora concluye, finalmente, que el corpus se puede emplear para aumentar el nivel de comprensión de lengua informal y para incentivar la conversación informal en la clase.

#### 4.1.2 *Corpus escritos*

Por su parte, los corpus escritos permiten a los aprendientes de español explorar rasgos léxicos, gramaticales, discursivos y genéricos de los textos pertenecientes a distintas áreas del saber y del quehacer humano representado en los corpus. En este sentido, resulta inminente incorporar en corpus de español general, pero también desde géneros discursivos especializados y disciplinares, los rasgos multimodales o multisemióticos. No basta con focalizar la atención únicamente en el sistema verbal de la lengua, sino que también se debe atender, por ejemplo, al desarrollo de estrategias de lectura de relaciones intermodales o intersistemas, como es el caso de la comprensión de textos compuestos por palabras y gráficos, esquemas o figuras. Todos estos tipos de artefactos deberían, de hecho, ser identificados en los corpus de géneros disciplinares (Parodi 2010b; Parodi y Burdiles 2015; Parodi, Julio y Vásquez-Rocca 2015). Leer y aprender a leer no es solo leer palabras, sino

que leer todos los sistemas semióticos que componen un texto. En este sentido, se debe tender a superar el Principio Logocentrista (Parodi y Julio 2016) que ha primado en la conformación de los corpus orientados preferentemente a explotar un solo sistema y centrados –de preferencia– en el componente verbal de la lengua.

En relación con el desarrollo de la competencia léxica, se recomienda promover su observación más allá de las palabras en los corpus y, del mismo modo, observar patrones de combinación de palabras para identificar los rasgos contextuales que las acompañan, explorar las colocaciones oracionales para tener en cuenta la posición de los vocablos en el contexto de uso, registrar frecuencias de aparición, buscar ejemplos de uso, etc. (Yılmaz y Soruç 2014).

La enseñanza para el desarrollo de la competencia gramatical tiene también un aliado en cualquier corpus disponible. Entre algunas aplicaciones posibles, está identificar y analizar estructuras gramaticales específicas, comparar algún componente prescriptivo gramatical con el uso auténtico de la lengua proveniente de textos en la red o de corpus de aprendientes, esto implica la detección de errores, proponer y comprobar alguna hipótesis gramatical mediante la recolección de datos del corpus, relacionar los usos de estructuras gramaticales del español con las de la lengua materna del estudiante (Paz 2013). El resultado de las búsquedas puede promover que el alumno induzca una determinada regla gramatical o norma de uso. Asimismo, si se propone al estudiante una regla gramatical como punto de partida, la búsqueda en el corpus estaría dirigida a encontrar deductivamente ejemplos representativos. En suma, se trata de que los estudiantes aprendan mediante la exploración del corpus (Bernardini 2004).

Todos los rasgos léxicogramaticales de la lengua escrita y hablada podrían, de la misma manera, ser objeto de observación y análisis por parte del estudiante de lengua, utilizando las bases de datos y herramientas de los corpus para establecer variaciones diatópicas, diacrónicas y variaciones genéricas. Entre los corpus en español revisados, el CREA es uno de los más accesibles a la búsqueda por tramo temporal, áreas temáticas, géneros discursivos y áreas geográficas.

#### ***4.2. Uso indirecto de los corpus***

Como ya se ha señalado, los corpus también pueden ser usados de manera indirecta en la enseñanza de lenguas. En este caso, la investigación basada en corpus deriva en la generación de productos a partir de los datos, tales como publicaciones de referencia y diseño de materiales didácticos. Se dice que es un uso indirecto porque el estudiante y el profesor, esto es, los usuarios de estos materiales, no acceden a la exploración del corpus, sino que son beneficiarios terciarios de ellos.

De un lado, tenemos a los diccionarios que provienen de proyectos lexicográficos basados en corpus. En ellos se recogen las palabras que figuran en un corpus representativo de la lengua o variedad lingüística sobre la que se trabaja y reflejan los significados realmente presentes en los textos, incorporando las marcas de uso correspondientes en cada caso (Rojo 2009). Un buen ejemplo de esta relación entre la existencia de un corpus y la creación de material a partir de él es la conformación del Corpus del Español Mexicano Contemporáneo (CEMC). Esta iniciativa surge de la necesidad de contar con una colección

de contextos de uso de las palabras que sirvieran como base a la elaboración del *Diccionario del español de México*. El corpus fue necesario porque permitió deslindar y reconocer las características del español mexicano contemporáneo, sobre todo en cuanto a su léxico.

De otro, en 2009 la RAE y ASALE publicaron la *Nueva gramática de la lengua española*, obra colectiva a cargo de las veintidós Academias de la Lengua Española. Esta gramática se construyó sobre las bases de datos textuales del Corpus Diacrónico del Español (CORDE) y del Corpus de Referencia del Español Actual (CREA), los que, con casi quinientos millones de formas, integran el banco de datos del español de la Real Academia Española. De este modo, la gramática considera variaciones de registro y explota las diferencias entre el modo escrito y el oral. En los corpus en que se apoya, se puede comprobar la existencia, la difusión y el ámbito de uso de las construcciones que se analizan. De ahí se ha extraído, además, un subcorpus especial, el Corpus de la Nueva Gramática, del que se obtienen los ejemplos más ilustrativos sobre el uso del español (RAE y ASALE 2009).

Para el diseño de materiales didácticos, los corpus de hablantes nativos constituyen una fuente amplia y fiable de recursos lingüísticos, una sólida base para elegir las estructuras lingüísticas más frecuentes en las producciones reales de los hablantes de una lengua. De este modo, los ejemplos provistos en la clase de lengua superan los ejemplos elaborados *ad hoc* y la mera intuición (Pérez-Ávila 2006). Se supera así un aspecto constantemente criticado a los textos de estudio, en el sentido de que se ofrece a los aprendientes ejemplos de la lengua que se alejan de su uso real y que no reflejan las complejidades y matices de la lengua natural.

En esta línea metodológica, pero a menor escala, los corpus también son un recurso para que el profesor construya materiales de clase a partir de extracción de *input* sobre ciertos aspectos de la lengua en los cuales desea instruir a sus alumnos. Para la explotación del corpus con esta finalidad por parte del docente, se sugiere que realice análisis cuantitativos que le permitan identificar los rasgos (palabras, colocaciones, estructuras, etc.) más frecuentes en la variación de lengua que interese.

Ahora bien, disponer de corpus de aprendientes de español L2, posibilita el acceso a datos exhaustivos y fidedignos del uso real que hacen de la lengua aprendientes de diferentes lenguas de origen y distinto nivel de competencia. Las recomendaciones de su uso apuntan principalmente a usarlos para establecer diferencias en algún rasgo de la lengua observado en corpus de aprendientes y fuentes prescriptivas del español. De manera semejante, se recomienda también medir las diferencias entre la interlengua y la actuación del hablante nativo, a fin de desarrollar mejoras en la enseñanza de la lengua y en el conocimiento del proceso de aprendizaje de la misma (Hasselgard y Johansson 2011). Subrayamos, no obstante, que un corpus de aprendientes puede proporcionar muestras de uso de lengua con errores típicamente producidos por escritores aprendientes de español L2 en una fase de su desarrollo lingüístico, influido por la transferencia lingüística de su lengua materna (Parodi 2015b). Este hecho se debe considerar y se debe advertir oportunamente de cara al estudiante (para más información sobre la aplicabilidad del corpus a la enseñanza de la lengua, véase Sinclair [2004] Anderson y Corbett [2009], Aijmer [2009], Reppen [2010], Bennett [2010], Campoy, Belles-Fortuno y Gea-Valor [2010], Cruz [2012] y Flowerdew [2012]).

Como es posible observar en lo expuesto en este capítulo, hoy contamos con sofisticados sistemas computacionales para la investigación y la docencia en procesos de enseñanza/aprendizaje de lenguas. Estos desarrollos han dado pie a un amplio número y variedad de corpus y bases de datos de lengua española disponibles en la red. Sin embargo, en sus características se advierte que su origen y finalidad están ligados principalmente a la investigación lingüística y no directamente a la búsqueda de propósitos pedagógicos. Asimismo, se observa que los corpus se han transformado en una herramienta cada vez más útil para la enseñanza como L2, pero, en concordancia con lo observado por McEnery y Xiao (2010), principalmente para el inglés. Observamos así que hasta ahora los corpus en español parecen haber jugado un papel más importante para decidir qué enseñar y un rol menos importante para decidir cómo los estudiantes pueden gestionar y construir su conocimiento a partir de corpus de manera autónoma con o sin mediación del profesor.

Sobre la base de las consideraciones anteriores, es posible afirmar que aún queda mucho por hacer en relación con el tema tratado en este capítulo, tanto en la investigación como en las aplicaciones didácticas. Se hace evidente que estamos en los albores de un recorrido que se inaugura hace poco y que se revela prometedor. Uno de los desafíos que emergen de nuestro análisis se relaciona con las características deseables en el profesor de lengua. Se requiere incorporar en la formación inicial de los profesores de español L2 un fuerte componente de conocimientos sobre lingüística de corpus y sobre sus posibilidades de explotación en el aula. Otro de los desafíos que quedan pendientes en la incorporación de los resultados de la investigación con corpus a la enseñanza de español L2, es, precisamente, promover la creación de corpus de español que, tanto en su diseño como en su contenido, se encuentren motivados pedagógicamente para cumplir con una planificación de enseñanza del español tanto L1 como L2, a fin de que las actividades de aprendizaje basadas en corpus se conviertan en una parte integral de las clases de lengua.

#### **Anexo 1.** Corpus del español L2.

<b>Nombre</b>	<b>Dirección web</b>	<b>Oral</b>	<b>Escrito</b>	<b>Anotado</b>
CAES	<a href="http://galvan.usc.es/caes">http://galvan.usc.es/caes</a>		X	Sí
Corpus de Aprendices Taiwaneses de Español (CATE)	<a href="http://corpora.fild.ncku.edu.tw/">http://corpora.fild.ncku.edu.tw/</a>	X		Sí
Spanish Corpus Proficiency Level Training (SPT)	<a href="http://www.laits.utexas.edu/spt/intro">http://www.laits.utexas.edu/spt/intro</a>	X		No
Spanish	<a href="http://www.splloc.soton.ac.uk/">http://www.splloc.soton.ac.uk/</a>	X		No

Parodi, G. & Burdiles, G. (2018). Corpora and databases. *En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español.* Londres: Routledge (en prensa).

Learner Language Oral Corpora (SPLLOC)				
Spanish Learner Oral Corpus	<a href="http://cartago.llf.uam.es/corele/home_en.html">http://cartago.llf.uam.es/corele/home_en.html</a>	X		Sí
Fono.Ele	<a href="http://www3.uah.es/fonoele/">http://www3.uah.es/fonoele/</a>	X		No
The DIAZ corpus	<a href="http://search.language-archives.org/record.html?id=talkbank_org_SLA_Bank-Spanish-DiazRodriguez">http://search.language-archives.org/record.html?id=talkbank_org_SLA_Bank-Spanish-DiazRodriguez</a>	X		No
Corpus Escrito del Español como L2 (CEDEL2)	<a href="http://wdb.ugr.es/~zanos/cdw/index.php">http://wdb.ugr.es/~zanos/cdw/index.php</a>		X	No
SFN Corpus	<a href="http://spanishfn.org/corpus">http://spanishfn.org/corpus</a>		X	Sí

## Anexo 2. Corpus del español L1.

Nombre	Dirección Web	Oral	Escrito	Anotado
El Grial	<a href="http://www.elgrial.cl">http://www.elgrial.cl</a>		X	Sí
Corpus del Español	<a href="http://www.corpusdelespanol.org/">http://www.corpusdelespanol.org/</a>		X	Sí

Parodi, G. & Burdiles, G. (2018). Corpora and databases. *En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español.* Londres: Routledge (en prensa).

Corpus Oral y Sonoro del Español Rural	<a href="http://www.lllf.uam.es/coser/index.php">http://www.lllf.uam.es/coser/index.php</a>	X		No
Corpus Oral de Referencia de la Lengua Española Contemporánea (CORLEC)	<a href="http://www.lllf.uam.es/ESP/Corlec.html">http://www.lllf.uam.es/ESP/Corlec.html</a>	X		No
CREA	<a href="http://corpus.rae.es/creanet.html">http://corpus.rae.es/creanet.html</a>		X	No
CORDE	<a href="http://corpus.rae.es/cordenet.html">http://corpus.rae.es/cordenet.html</a>		X	No
Biblioteca Virtual Miguel de Cervantes	<a href="http://www.cervantesvirtual.com/controladores/busqueda_avanzada_form.php">http://www.cervantesvirtual.com/controladores/busqueda_avanzada_form.php</a>		X	No
Corpus Diacrónico y Diatópico del Español de América (CORDIAM)	<a href="http://www.cordiam.org/">http://www.cordiam.org/</a>		X	No
Corpus Oral Didáctico Anotado Lingüísticamente (C-Or-DiAL)	<a href="http://lablita.dit.unifi.it/corpora/cordial">http://lablita.dit.unifi.it/corpora/cordial</a>	X		Sí
C-ORAL-ROM	<a href="http://lablita.dit.unifi.it/coralrom/index.html">http://lablita.dit.unifi.it/coralrom/index.html</a>	X		Sí
Corpus Dinámico del Castellano de Chile (Codicach)	<a href="http://sadowsky.cl/codicach-es.html">http://sadowsky.cl/codicach-es.html</a>		X	Sí
Proyecto COLA (Corpus Oral del Lenguaje)	<a href="http://colam.org/om_prosj-espannol.html">http://colam.org/om_prosj-espannol.html</a>	X		No

Parodi, G. & Burdiles, G. (2018). Corpora and databases. *En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español.* Londres: Routledge (en prensa).

Adolescente)				
El Corpus del Español del Siglo XXI (CORPES XXI)	<a href="http://web.frl.es/CORPES/view/inicioExterno.vi&lt;br/&gt;ew">http://web.frl.es/CORPES/view/inicioExterno.vi ew</a>	X	X	Sí
Corpus del Español en Texas	<a href="http://corpus.spanishintexas.org/es">http://corpus.spanishintexas.org/es</a>	X		Sí
Corpus orales para la fonética y las tecnologías del habla en español	<a href="http://liceu.uab.es/~joaquim/language_resources/&lt;br/&gt;spoken_res/Corp_oral_esp.html">http://liceu.uab.es/~joaquim/language_resources/ spoken_res/Corp_oral_esp.html</a>	X		No
IULA Spanish LSP Treebank	<a href="http://www.iula.upf.edu/recurs01_tbk_uk.htm">http://www.iula.upf.edu/recurs01_tbk_uk.htm</a>		X	Sí
PANACEA Environment Spanish monolingual corpus	<a href="http://catalog.elra.info/product_info.php?product&lt;br/&gt;s_id=1192">http://catalog.elra.info/product_info.php?product s_id=1192</a>		X	Sí
PRESEEA (Proyecto para el Estudio Sociolingüístico del Español de España y América)	<a href="http://preseea.linguas.net/">http://preseea.linguas.net/</a>	X		No
RST Spanish Treebank	<a href="http://www.corpus.unam.mx/rst/index_es.html">http://www.corpus.unam.mx/rst/index_es.html</a>		X	Sí
Spanish Online (SOL)	<a href="http://spraakbanken.gu.se/konk/rom2/">http://spraakbanken.gu.se/konk/rom2/</a>		X	No

Parodi, G. & Burdiles, G. (2018). Corpora and databases. *En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español*. Londres: Routledge (en prensa).

Corpus del Español en el Sur de Arizona (CESA)	<a href="https://cesa.arizona.edu/">https://cesa.arizona.edu/</a>	X		No
UAM Spanish Treebank	<a href="http://www.llf.uam.es/ESP/Treebank.html">http://www.llf.uam.es/ESP/Treebank.html</a>		X	Sí
Base de datos sintácticos del Español actual (BDS)	<a href="http://www.bds.usc.es/">http://www.bds.usc.es/</a>		X	Sí
Val.Es.Co	<a href="http://www.uv.es/corpusvalesco/corpus.html">http://www.uv.es/corpusvalesco/corpus.html</a>	X		No
Cumbre	<a href="http://www.um.es/grupolacell/proyectos/proyecto/1">http://www.um.es/grupolacell/proyectos/proyecto/1</a> <a href="http://www.um.es/lacell/miembros/asp/cumbre.doc">www.um.es/lacell/miembros/asp/cumbre.doc</a>	X	X	Sí
Corpus del Nuevo diccionario histórico (CDH)	<a href="http://www.frl.es/Paginas/Corpusdiccionariohistorico.aspx">http://www.frl.es/Paginas/Corpusdiccionariohistorico.aspx</a>		X	Sí
Corpus del Español Mexicano Contemporáneo (CEMC)	<a href="http://www.corpus.unam.mx:8080/unificado/index.jsp?c=cemc">http://www.corpus.unam.mx:8080/unificado/index.jsp?c=cemc</a>	X	X	Sí
Corpus Oral y Sonoro del Español Rural	<a href="http://www.llf.uam.es/coser/index.php">http://www.llf.uam.es/coser/index.php</a>	X		No
esTenTen corpus	<a href="https://www.sketchengine.co.uk/estenten-corpus">https://www.sketchengine.co.uk/estenten-corpus</a>		X	Sí
esAmTenTen corpus	<a href="https://www.sketchengine.co.uk/esamtenten-corpus">https://www.sketchengine.co.uk/esamtenten-corpus</a>		X	Sí

## 5. Bibliografía citada

Aijmer, K. 2009. *Corpora and Language Teaching*. Amsterdam: John Benjamins.



- Albelda, M. 2010. "Rentabilidad de los corpus discursivos en la didáctica de lenguas extranjeras". En *Actas del XXI Congreso Internacional de ASELE. Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*, eds. J. de Santiago Guervós, H. Bongaerts, J. Sánchez y M. Seseña, 83-95. Salamanca: Universidad de Salamanca.
- Anderson, W. y J. Corbett. 2009. *Exploring English with Online Corpora*. Basingstoke: Palgrave Macmillan.
- Bailini, S. 2014. "Los corpus como recursos didácticos para la enseñanza de las variedades diatópicas del español". Trabajo presentado en el V Congreso internacional de FIAPE: ¿Qué español enseñar y cómo? Variedades del español y su enseñanza, Cuenca, 25-28 de junio.
- Bennett, G. 2010. *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor, MI: University of Michigan Press.
- Briz, A. 1998. *El español coloquial en la conversación. Esbozo de pragmatogramática*. Barcelona: Ariel.
- Briz, A. y M. Albelda. 2009. "Estado actual de los corpus de lengua española hablada y escrita: I+D". En *Anuario del Instituto Cervantes*, 165-225. Madrid: Instituto Cervantes.
- Bernardini, S. 2004. "Corpora in the Classroom: An Overview and Some Reflections on Future Developments". En *How to Use Corpora in Language Teaching*, ed. J. Sinclair, 15-36. Amsterdam: John Benjamins.
- Campoy, M., B. Belles-Fortunato y M. Gea-Valor. 2010. *Corpus-Based Approaches to English Language Teaching*. Londres: Continuum.
- Cestero, A., I. Penadés, A. Blanco, L. Camargo y J. Granda. 2001. "Corpus para el análisis de errores de aprendices de E/LE (CORANE)". En *Actas del XII Congreso Internacional de ASELE. Tecnologías de la información y de las comunicaciones en la enseñanza de ELE*, ed. A. Gimeno, 527-534. Valencia: Universidad Politécnica de Valencia.
- Chambers, A. 2010. "What is Data-Driven Learning". En *The Routledge Handbook of Corpus Linguistics*, eds. A. O'Keeffe y M. McCarthy, 345-358. Londres y Nueva York: Routledge.
- Consejo de Europa. 2002. *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: MECD.  
[http://cvc.cervantes.es/ensenanza/biblioteca\\_ele/marco/](http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/).
- Cruz, Mar. 2012. *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco/Libros.
- De Kock, J. 2001a. "Un corpus informatizado para la enseñanza de la lengua española. Punto de partida y término". *Hispanica Polonorum* 3: 60-86.
- De Kock, J., ed. 2001b. *Lingüística con corpus. Catorce aplicaciones sobre el español*. Salamanca: Universidad de Salamanca.
- Drange, E. 2008. "Un corpus oral en línea como recurso didáctico". En *Actas de ANPE. II Congreso Nacional: multiculturalidad y norma policéntrica: aplicaciones en el Aula de ELE*, eds. J. Izquierdo, K. Salkjelsvik, I. Hansejordet, J. Cosmes-Cuesta, A. Johnsen y M. Sánchez, 26-27. Bergen: ANPE.

Parodi, G. & Burdiles, G. (2018). Corpora and databases. En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), *The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español*. Londres: Routledge (en prensa).

Ferreira, A., P. Salcedo, G. Kotz y F. Barrientos. 2012. "La arquitectura de ELE-TUTOR: un sistema tutorial inteligente para el español como lengua extranjera." *Revista Signos. Estudios de Lingüística* 45, 79: 102-131.

Fligelstone, S. 1993. "Some Reflections on Teaching, from a Corpus Linguistics Perspective". *ICAME Journal* 17: 97-109.

Flowerdew, L. 2012. *Corpora and Language Education*. Londres: Palgrave Macmillan.

Flowerdew, L. 2015. "Data-Driven Learning and Language Learning Theories: Whither the Twain Shall Meet." En *Multiple Affordances of Language Corpora for Data-Driven Learning*, eds. A. Lenko-Szymanska y A. Boulton, 15-36. Amsterdam: John Benjamins.

Granger, S., G. Gilquin y F. Meunier, eds. 2013. *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead*. Louvain-la-Neuve: Presses Universitaires de Louvain.

Hasselgard, H. y S. Johansson. 2011. "Learner Corpora and Contrastive Interlanguage Analysis". En *A Taste for Corpora. In honour of Sylviane Granger*, editado por Fanny Meunier, S. De Cock, G. Guilquin y M. Paquot, 33-61. Amsterdam: John Benjamins.

Instituto Cervantes. 2006. *Plan curricular del Instituto Cervantes*. Madrid: Biblioteca Nueva.

Johns, T. 1994. "From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-Driven Learning". En *Perspectives on Pedagogical Grammar*, ed. T. Odlin, 293-313. Cambridge: Cambridge University Press.

Johns, T. 1986. "Micro-Concord: A Language Learner's Research Tool". *System* 14 (2): 151-162.

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.

López-García, M. 2013. "Aplicaciones de los corpus orales a la didáctica de ELE: repertorios léxicos y contextos pragmlingüísticos". En *Actas del XXIV Congreso Internacional de la ASELE. La enseñanza del español como LE/L2 en el siglo XXI*, ed. N. Contreras, 397-409. Málaga: ASELE.

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*, ed. Eli Hinkel, 364-380. Londres y Nueva York: Routledge.

Mendikoetxea, A. 2013. "Corpus-Based Research in Second Language Spanish". En *The Handbook of Spanish Second Language Acquisition*, ed. K. Geeslin, 9-29. Medford, MA: Wiley.

Moreno, F. 2000. *Qué español enseñar*. Madrid: Arco/Libros.

Nicolás, C. 2013. "Propuestas prácticas para el uso en el aula de C-Or-DiAL (Corpus Oral Didáctico Anotado Lingüísticamente)". En *Actas del XXIV congreso internacional de ASELE. La enseñanza del español como LE/L2 en el siglo XXI*, ed. N. Contreras, 903-912. Málaga: ASELE.

Parodi, G., ed. 2007. *Working with Spanish Corpora*. Londres: Continuum.

Parodi, G., 2010a. *Lingüística de Corpus: de la teoría a la empiria*. Fráncfurt: Iberoamericana/Veruert.

Parodi, G. & Burdiles, G. (2018). Corpora and databases. En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), *The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español*. Londres: Routledge (en prensa).

- Parodi, G. 2010b. "Multisemiosis y lingüística de corpus: Artefactos (multi)semióticos en los textos de seis disciplinas en el Corpus PUCV-2010". *Revista de lingüística teórica y aplicada* 48 (2): 33-70.
- Parodi, G. 2015a. "Variation across University Genres in Seven Disciplines: A Corpus-Based Study on Academic Written Spanish". *International Journal of Corpus Linguistics* 20 (4): 469-499.
- Parodi, G. 2015b. "Corpus de Aprendices de Español (CAES)". *Journal of Spanish Language Teaching* 2 (2): 194-200.
- Parodi, G. y G. Burdiles, eds. 2015. *Leer y escribir en contextos académicos y profesionales: géneros, corpus y métodos*. Santiago de Chile: Ariel.
- Parodi, G. y C. Julio. 2015. "Más allá de las palabras: ¿Puede comprenderse el género discursivo Informe de Política Monetaria desde un único sistema semiótico predominante?" *Revista ALPHA* 41: 133-158.
- Parodi, G., C. Julio y L. Vásquez-Rocca. 2015. "Los géneros del Corpus PUCV-UCSC-2013 del discurso académico de la economía: el caso del Informe de Política Monetaria". *Revista latinoamericana de estudios del discurso* 15 (2): 179-200.
- Parodi, G. y C. Julio. 2016. "¿Dónde se posan los ojos al leer textos multisemióticos disciplinares? Procesamiento de palabras y gráficos en un estudio experimental con eye tracker". *Revista Signos. Estudios de Lingüística* 49 (Supl.1): 149-183.
- Paz, Y. 2013. "Aplicación pedagógica de la lingüística de corpus a una clase de gramática avanzada de español como segunda lengua". *MarcoELE* 16.  
<http://marcoele.com/linguistica-de-corpus/>.
- Pérez-Ávila, E. 2006. "El corpus lingüístico en la didáctica del léxico en el aula de E/LE". Tesis de máster, Universidad Antonio de Nebrija.
- Pitkowski, E. y J. Vásquez. 2009. "El uso de corpus lingüísticos como herramienta pedagógica para la enseñanza y aprendizaje de ELE". *TINKUY* 11: 31-51.
- Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española*. Madrid: Espasa.
- Reppen, R. 2010. *Using Corpora in the Language Classroom*. Cambridge: Cambridge University Press.
- Rojo, G. 2009. "Sobre la construcción de diccionarios basados en corpus". *Tradumática* 7.  
<http://webs2002.uab.es/tradumatica/revista/num7/articles/02/02art.htm>.
- Rojo, G. 2015. "Hispanic Corpus Linguistics". En *The Routledge Handbook of Hispanic Applied Linguistics*, ed. Manel Lacorte, 371-386. Londres y Nueva York: Routledge.
- Rojo, G. 2016. "Los corpus textuales del español". En *Enciclopedia Lingüística Hispánica*, ed. Javier Gutiérrez-Rexach, 285-296. Londres y Nueva York: Routledge.
- Sánchez, A., S. Bernardos y A. López. 2001. "Enseñanza de español como lengua extranjera a través de un sistema informático de análisis textual disponible en Internet". En *Actas del XII Congreso Internacional de ASELE. Tecnologías de la Información y de las Comunicaciones en la Enseñanza de ELE*, ed. Ana Gimeno, 485-490. Valencia: Universidad Politécnica de Valencia.
- Secchi, D. 2014. "Los corpus discursivos orales como recurso didáctico: la oralidad a través del texto en contexto". *Foro de Profesores de E/LE* 10: 241-50.

Parodi, G. & Burdiles, G. (2018). Corpora and databases. En J. Muñoz-Bassols, E. Gironsetti & M. Lacorte (Eds.), *The Routledge Handbook of Spanish Language Teaching: Metodología, recursos y contextos para la enseñanza del español*. Londres: Routledge (en prensa).

- Sinclair, J. 2004. *How to Use Corpora in Language Teaching*. Ámsterdam: John Benjamins.
- Stenström, A.-B. y A. Jørgensen. 2008. “La función fática de los vocativos en la conversación juvenil de Madrid y Londres”. En *Cortesía y conversación: de lo escrito a lo oral, Actas del III Coloquio internacional de EDICE*, ed. A. Briz, A. Hidalgo, M. Albelda, J. Contreras y N. Hernández, 355-365. Valencia: Universidad de Valencia.
- American Council on the Teaching of Foreign Languages. 2012. *ACTFL Proficiency Guidelines*. Alexandria: American Council on the Teaching of Foreign Languages.
- Tolchinsky, L. 2014. “El uso de corpus lingüísticos como herramienta pedagógica” *Textos de didáctica de la lengua y de la literatura* 65: 9-17.
- Tribble, C. y G. Jones. 1990. *Concordances in the Classroom: A Resource Book for Teachers*. Londres: Longman.
- Yilmaz, E. y A. Soruç. 2015. “The Use of Concordance for Teaching Vocabulary: A Data-Driven Learning Approach”. *Procedia - Social and Behavioral Sciences* 191: 2626-2630.

## 6. Bibliografía recomendada

- Cantos Gómez, P. y A. Sánchez Pérez, eds. 2009. *A Survey on Corpus-Based Research/Panorama de investigaciones basadas en corpus*. Murcia: AELINCO.
- Hinkel, E., ed. 2010. *Handbook of Research in Second Language Teaching and Learning*. Londres y Nueva York: Routledge.
- Cifuentes Honrubia, J. y S. Rodríguez Rosique, eds. 2011. *Spanish Word Formation and Lexical Creation*. Ámsterdam: John Benjamins.
- Geeslin, K., ed. 2013. *The Handbook of Spanish Second Language Acquisition*. Medford, MA: Wiley.
- Parodi, G. y G. Burdiles, eds. 2015. *Leer y escribir en contextos académicos y profesionales: géneros, corpus y métodos*. Santiago de Chile: Ariel.
- Sánchez Rufat, A. 2015. “La investigación de corpus de aprendientes y el desarrollo de los estudios de la interlengua del español”. *Language Design* 17: 57-84.

---

<sup>1</sup> Investigación desarrollada en el marco del Proyecto FONDECYT 1170623.