

**As línguas estrangeiras no
ensino superior:
propostas didáticas
e casos em estudo**

Maria Ellison
Marta Pazos Anido
Pilar Nicolás Martínez
Sónia Valente Rodrigues
ORGS.

Porto, FLUP, 2018

FICHA TÉCNICA

TÍTULO: As línguas estrangeiras no ensino superior: propostas didáticas e casos em estudo

ORGANIZAÇÃO: Maria Ellison, Marta Pazos Anido, Pilar Nicolás Martínez e Sónia Valente Rodrigues

EDIÇÃO: Faculdade de Letras da Universidade do Porto e APROLÍNGUAS - Associação Portuguesa de Professores de Línguas Estrangeiras do Ensino Superior

ANO DE EDIÇÃO: Impresso em fevereiro de 2018

COLEÇÃO: FLUP e-DITA

EXECUÇÃO GRÁFICA: Gráfica Firmeza Lda. / Porto

TIRAGEM: 100 exemplares

DEPÓSITO LEGAL: 437121/18

ISBN: 978-989-54030-7-3

ISSN: 1646-1525

Este trabalho é financiado pela APROLÍNGUAS - Associação Portuguesa de Professores de Línguas Estrangeiras do Ensino Superior.

LOS CORPUS DE APRENDICES: UN TERRENO EN EXPANSIÓN PARA LA ENSEÑANZA DE ESPAÑOL¹

ABSTRACT

The first decades of this century are witnessing a new impulse in the design and exploitation of different types of linguistic corpora, from which learner corpora are becoming the most relevant and interesting for foreign language teachers, i.e., corpora built on texts produced by people learning a new language. Having access to the interlanguage of those who are in the process of learning Spanish as a foreign language is becoming a tool of great interest both for teachers and language researchers, since it allows them to contrast the difficulties in the process of learning a language and the vocabulary employed in its production.

In this chapter we will describe the current projects concerning Spanish learner corpora by succinctly reviewing the results of the research carried out from them. We will conclude by emphasising the main contributions, as well as the interest in extending the information coming from the learners' productions of initial diverse languages in oral texts and in longitudinal samples.

Keywords: learner corpora, second language acquisition, Spanish as a foreign language, teaching Spanish as a foreign language

¹ Este texto tiene su origen en una comunicación presentada por las autoras en el Colóquio Internacional APROLÍNGUAS (Oporto, Portugal, 16-18 de junio de 2016), con el título «Los corpus de aprendices de español y su explotación». La participación de Inmaculada Mas Álvarez contó con la financiación del proyecto COMBIDIGILEX, *La combinatoria en paradigmas léxico-semánticos en contraste* (FFI2015-64476-P MINECO/FEDER UE), de la Universidade de Santiago de Compostela (España).

RESUMEN

Los años transcurridos del siglo actual están suponiendo un nuevo impulso a la creación y explotación de corpus lingüísticos de muy distinto tipo. Entre los que ofrecen más interés para el profesorado de lenguas extranjeras hay que destacar los corpus de aprendices, aquellos en los que se recogen textos producidos por personas que aprenden una nueva lengua, diferente de la(s) inicial(es) o familiar(es). Tener acceso a la interlengua de quienes están en el proceso de aprendizaje del español es una herramienta de gran interés para el profesorado, pues permite contrastar las dificultades de aprendizaje y el vocabulario empleado en la producción.

En este capítulo describimos los proyectos de corpus de aprendices de español existentes, haciendo una somera revisión de las investigaciones llevadas a cabo a partir de ellos. Concluimos destacando las aportaciones esenciales, así como el interés de ampliar los datos a partir de las producciones de aprendices de lenguas iniciales diversas, en textos orales y en recogidas longitudinales.

Palabras clave: corpus de aprendices, adquisición de segundas lenguas, Español como lengua extranjera, enseñanza del Español como lengua extranjera

1 – CORPUS DE APRENDICES: CONCEPTO Y CARACTERÍSTICAS

En lingüística, se entiende por *corpus* un conjunto, más o menos amplio, de (fragmentos de) textos naturales, almacenados en formato electrónico, que han sido reunidos en una aplicación informática, según un determinado diseño, para facilitar el estudio de la lengua o variedad lingüística de la que esos textos han sido extraídos como representativos de ella (Rojo, 2016). Entre los muchos tipos y subtipos de corpus existentes en la actualidad, los llamados *corpus de aprendices* contienen textos producidos por personas que están aprendiendo una determinada lengua, la lengua objeto o meta, y que tienen diferentes lenguas iniciales o familiares (diferentes L1). La definición de referencia es la que formula Granger (2002, p. 7), basándose en la definición de *corpus* de Sinclair (1996):

Computer learner corpora are electronic collections of authentic F[oreign] L[anguage] / S[econd] L[anguage] textual data assembled according to explicit designed criteria for a particular SLA[cquisition]/FLT[eaching] purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance.

Esta definición requiere, según explica su autora — y discute en detalle también Sinclair —, comentarios sobre los términos siguientes: *authentic, FL/SL varieties, textual data, explicit design criteria, SLA/FLT purpose, standarization and documentation*. En efecto, la autenticidad en relación con

el discurso dentro de situaciones de aprendizaje de lenguas cubre un abanico de producciones más o menos genuinas; la autenticidad será mayor cuando estemos ante emisiones naturales propias de una clase de lengua extranjera. Estos corpus cubren, por una parte, a los aprendices de una lengua como lengua extranjera y, por otra, a los de una lengua como segunda lengua, ambas variedades no nativas entran en consideración, por lo tanto. Además, cuando se habla de datos textuales debe quedar claro que se trata de fragmentos de discurso continuos, no de palabras o secuencias aisladas. Por este motivo, los corpus de aprendices no reúnen solo errores, o secuencias erróneas, sino que de manera natural contienen usos correctos e incorrectos de la lengua en cuestión. Como vemos más abajo, los criterios que guían el diseño de estos corpus resultan decisivos para la aplicabilidad de la investigación a que den lugar, pues deben atender, además de a los habituales para corpus nativos, tanto a las variables que afectan a los aprendices como a las condiciones de las tareas que desempeñan para las producciones que se pretenden recoger.

1.1 – Características principales de los corpus de aprendices

La importancia del diseño de un corpus de aprendices reside en lo fundamental en tres factores: los objetivos que se persigan, las variables que se tengan en cuenta y la metodología seguida en la anotación. Naturalmente los objetivos condicionan los otros factores, pero cada vez se hace más evidente la utilidad de corpus bien diseñados, accesibles en línea, que permitan consultas y estudios para finalidades diversas. Sea como sea, la sistematicidad en la recogida de datos, la rigurosidad en la metodología y la explicación detallada de las características de cada colección son condiciones esenciales para un buen aprovechamiento. En general, los corpus de aprendices suelen ser monolingües, tienden a contener ejemplos de lengua general (no especializada), son normalmente sincrónicos y recogen, en su mayor parte, textos escritos (Granger, 2002). Pero en los desarrollos recientes el panorama se amplía con corpus bilingües/paralelos y multilingües, para enseñanza con fines específicos, con recogidas longitudinales y de textos orales. No nos cabe duda de que el abanico de posibilidades continuará enriqueciéndose y diversificándose. Como elementos principales a tener en cuenta en los corpus de aprendices destacamos los siguientes:

- **Variables referidas a los informantes.** Además de las variables clásicas, como edad, nivel de lengua y L1, se han apuntado otras de indudable interés para la enseñanza-aprendizaje de lenguas, como motivación, actitud hacia el aprendizaje de lenguas, estilo de aprendizaje, contexto de aprendizaje o conocimiento de otras lenguas extranjeras (Cestero Mancera et al., 2001). Con respecto al nivel de lengua, se ha insistido

en la conveniencia de una estandarización de la asignación del nivel, ya que los años u horas de estudio, o el curso de escolarización, no siempre son criterios fiables. Por ello, Lozano (2015, p. 189) defiende como mejor indicador de la competencia en la lengua meta la aplicación de una prueba de nivel estándar. En el momento actual se recurre, en la mayor parte de los casos, a los niveles establecidos como referencia común en Europa (desde A1, usuario principiante, hasta el C2, usuario competente). Nos gustaría subrayar la relevancia de tener en cuenta el conocimiento de otras lenguas extranjeras, así como la posible condición de hablantes bilingües de los informantes (con dos L1, por tanto), algo que es esencial, por ejemplo, para Buyse & González Melón (2013, p. 251).

- **Diseño de las pruebas.** Esta categoría incluye los métodos usados para la recogida de muestras de lengua, sincrónico frente a longitudinal, las formas en que se obtienen las producciones — textos espontáneos o textos con diferentes grados de control — y la posibilidad de usar materiales de referencia como diccionarios o gramáticas. Se incluye asimismo en este apartado la duración de cada una de las pruebas, puesto que es un factor a tener en cuenta a la hora de analizar las muestras. Aunque normalmente la recogida es sincrónica, esto es, se hace en un momento dado del aprendizaje, es indudable que lo más interesante tanto desde la psicolingüística, como desde la didáctica, es obtener corpus longitudinales, es decir, pruebas similares en distintos momentos, pues cada persona que aprende, cada grupo de estudiantes, cuenta con una interlengua en cada estadio de su aprendizaje. Por otra parte, se suele apuntar, en cuanto al diseño, la conveniencia de contar con un corpus nativo comparable al no nativo para abordar los aspectos contrastivos que cada estudio particular pueda requerir (Granger, 2002; Lozano, 2015). Con todo, no es sencillo decidir qué corpus puede resultar válido en cada caso, ya que las producciones de informantes nativos pueden verse condicionadas y presentar fluctuaciones y errores respecto a la norma de la lengua meta (Sánchez Rufat & Jiménez Calderón, 2013). Por último, a propósito del grado de control en la obtención de los datos, los estudios de adquisición prefieren recogidas mixtas, con muestras espontáneas y muestras experimentales, mientras que los de enseñanza de lenguas extranjeras aprecian la espontaneidad de las producciones, que de todas formas siempre es relativa en contextos de enseñanza-aprendizaje.
- **Variables referidas a las muestras de lengua.** Esta categoría distingue entre corpus orales (expositivos o interactivos) y corpus escritos (cartas,

diarios, ensayos, etc.), tipo de texto (narración, argumentación, etc.) y tema. Hay que tener en cuenta las ventajas y desventajas del tipo de corpus a la hora de analizar los resultados. Se suele señalar que en los corpus escritos hay menos errores en la producción, los aprendices tienden a emplear estructuras más complejas —el lenguaje se presenta simplificado, se dice, en la producción oral de los no nativos— y se considera un medio particularmente apto para aprendices avanzados, especialmente en comparación con corpus de hablantes nativos. Los proyectos de corpus orales destacan, por su parte, la espontaneidad de la interacción cara a cara como premisa para obtener pruebas más directas de la interlengua, a la vez que minimizan los efectos de la autocorrección y monitoreo. Como un reto para futuras compilaciones y análisis, se ha apuntado también el interés de poner el foco además sobre textos multimodales, atendiendo a una presencia cada vez más destacada, en los medios de comunicación y en entornos académicos, de gráficos, esquemas, tablas, figuras e ilustraciones (Parodi, 2010, 2015). Otras categorías que igualmente sería deseable cubrir son las correspondientes a nuevas formas de interacción mediadas por ordenador y dispositivos móviles: la de los textos espontáneos, producidos por aprendices —en comparación con los producidos por nativos—, que se manifiestan en forma escrita, pero están más o menos imbuidos de oralidad, como los mensajes de correo electrónico y, sobre todo, los mensajes de chat móvil; o la de formas de diálogo en situación de videoconferencia, en especial destinadas a mejorar la competencia oral de los aprendices.

1.2 – Describiendo la lengua de los aprendices

Los datos observables que resultan en las producciones de las personas que aprenden una nueva lengua constituyen una vía esencial para conocer la interlengua. Como es sabido, el concepto de interlengua, desarrollado en el seno de los estudios de adquisición de segundas lenguas (o lenguas extranjeras), hace referencia al sistema específico de una lengua no materna, y los estudios sobre la interlengua se ocupan del conjunto de elementos que intervienen en la construcción del conocimiento interlingüístico (Baralo Ottonello, 2004). El acceso a muestras de la interlengua que proporcionan los corpus de aprendices supone una verdadera revolución en el terreno de los estudios de adquisición de segundas lenguas, pero también pueden emplearse con gran aprovechamiento en la investigación dedicada a cómo mejorar el aprendizaje y la enseñanza de lenguas extranjeras en general y el aprendizaje y la enseñanza de una determinada lengua en particular.

Las ventajas que ofrecen los corpus de aprendices como fuente de acceso a la interlengua se deben al desarrollo de las herramientas tecnológicas, ya

que la compilación a gran escala, el análisis y la presentación de los datos se convierten en procesos asumibles en buena medida de manera automática. Podemos resumir las oportunidades que ofrecen los corpus de aprendices actuales en los siguientes puntos (Granger 2002; Granger 2009; Granger, Gilquin & Meunier, 2015):

- a. Es posible recopilar gran cantidad de datos —procedentes de un número elevado de aprendices—, con la consiguiente representatividad de los fenómenos observados.
- b. Los datos recogidos se pueden considerar auténticos, en el sentido de que no se han obtenido en entornos experimentales o clínicos.
- c. Es posible abordar el análisis semiautomático gracias a los lematizadores y a las herramientas informáticas que aceleran el trabajo de análisis y obtención de concordancias, lo cual facilita el estudio del léxico y la fraseología.
- d. Ofrecen datos sobre frecuencia de palabras y frases, y presentan todos los ejemplos de un ítem en su contexto lingüístico inmediato. Permiten la anotación automática de errores.
- e. Permiten el acceso a la interlengua en forma de discurso contextualizado, lo que supone hacer factible el estudio de aspectos hasta hace poco no tenidos en cuenta, como elementos léxicos, fraseología, información estructural, discurso, y no solo morfología o sintaxis, que han sido los centros de interés más abordados hasta el siglo actual.

Todo ello está permitiendo el desarrollo de una nueva disciplina aplicada, la de la investigación en corpus de aprendices; un campo de naturaleza intrínsecamente interdisciplinar para el que hace pocos años se auguraba un futuro brillante y al que se otorga el protagonismo en la conexión entre los estudios de corpus lingüísticos, la adquisición de segundas lenguas, la enseñanza de lenguas y el procesamiento del lenguaje natural (Granger 2009; Granger, Gilquin & Meunier, 2015).

Desde la didáctica de lenguas extranjeras los estudios clásicos han puesto el énfasis en la descripción de la lengua meta, en la caracterización de los aprendices —respecto a estilos de aprendizaje, motivación, etc.— y en los aspectos de corte pedagógico, como la formación del profesorado, las tareas de aprendizaje, el currículo y las programaciones. Lo que, sin embargo, se echaba de menos a finales del siglo pasado era algo tan esencial y básico como el conocimiento de la lengua empleada por los aprendices en sus producciones (Mark, 1998). Y el acceso a la lengua de los aprendices, a la interlengua, es lo que nos proporcionan los corpus de aprendices. En la tradición de las investigaciones sobre producciones de estudiantes el foco de

atención se ha dirigido al análisis de errores. Sin embargo, tanto el concepto de *interlengua* como propiamente el de *error* resultan controvertidos a lo largo de los estudios lingüísticos de las producciones de aprendices. Por una parte, los motivos que llevan al estudio de nuevas lenguas y las expectativas de las personas que las aprenden muestran que el ideal de todo aprendizaje de lenguas extranjeras no es ya en todos los casos el del dominio lo más próximo posible al nativo. Ello supone que el sistema lingüístico de los aprendices de una determinada lengua pueda ser considerado un lugar de llegada, y no tanto un punto de paso en el proceso de adquisición/aprendizaje. Por otra parte, las investigaciones actuales han ampliado su óptica para estudiar no solo errores, sino también casos de infrarrepresentación o sobrerrepresentación de palabras, frases y estructuras.

Los estudios de la interlengua se vienen explotando con análisis contrastivos en tres direcciones: las comparaciones de escrituras y hablas nativas frente a otras no nativas, las comparaciones entre varias escrituras y hablas no nativas —con diferentes L1, por ejemplo— y los estudios longitudinales, aquellos que tienen en cuenta las producciones de los mismos informantes en diversos momentos de su aprendizaje. El caudal de información que aportan los corpus de aprendices para análisis contrastivos de estos tipos es muy valioso, pero es necesario tener en cuenta cada una de estas finalidades en la recogida de los textos y el diseño de la presentación y las posibilidades de búsqueda, como ya hemos visto. Es fundamental, pues, considerar los criterios que guían la selección de informantes y textos de cara a la explotación de los corpus. Por ejemplo, a la hora de elegir un corpus nativo como grupo de control, será necesario decidir la variante dialectal y diafásica, así como el nivel de conocimiento de la lengua que se considera adecuado en los informantes, o en todo caso, cuál será la norma tenida en cuenta (Sánchez Rufat & Jiménez Calderón, 2013). Por supuesto, los resultados de los estudios de carácter contrastivo presentan un indudable atractivo para la investigación en adquisición de segundas lenguas así como para la didáctica de lenguas.

2 – PROYECTOS DE CORPUS DE APRENDICES DE ESPAÑOL

El interés por los corpus de aprendices de español ha ido en aumento durante los últimos 15 años y son varios los proyectos que se han puesto en marcha en diferentes universidades, así como diversos son los criterios que los guían. Como reseñan los trabajos que revisan el panorama (Mendikoetxea, 2014, p. 15, por ejemplo), estamos en estos años asistiendo a la proliferación de investigaciones sobre la compilación y explotación de los corpus de aprendices en general y de los de aprendices de español en particular. Aquí los presentamos brevemente, ordenados según el carácter

oral o escrito de las producciones recogidas y, dentro de cada grupo, en orden alfabético².

2.1 – Corpus de aprendices de español orales

2.1.1 – CORELE (Corpus Oral de Español como Lengua Extrajera). Se trata de un corpus compuesto por muestras de hablantes ($N = 40$) de trece L1 diferentes, de niveles A2 y B1, recogido para un estudio de análisis de errores de la producción oral. Consta de cerca de 14 horas de entrevistas en un diálogo semiespontáneo entre el investigador y cada estudiante. Es posible acceder al corpus en línea a través de un cuidado sitio con información completa y clara, que incluye las transcripciones, el etiquetado de errores (fonéticos, léxicos, gramaticales y pragmáticos), recuentos de palabras y listas de frecuencias. Además, se ofrece una guía didáctica y actividades para profesores, cuyo objetivo es la explicación del error en cada grupo de estudiantes y el adiestramiento en la corrección de los errores (Campillos Llanos, 2012, 2014).

2.1.2 – CORINÉI (Corpus Oral Interlengua Español Italiano). Es un corpus en proceso, de momento no accesible, resultado del proyecto *Teletándem* de la Universidad de Alicante, que pretende conseguir que los estudiantes de español (L1 = italiano) interaccionen con nativos en condiciones de autenticidad. Interesa propiciar en los aprendices la reflexión metalingüística a través de la autotranscripción y la autoevaluación. Además, de cara a la investigación se pretende que los materiales que constituyen el corpus sean válidos para analizar las estrategias empleadas por los estudiantes italianos y españoles en la interacción nativo/no nativo y para comprender la evolución de su interlengua.

2.1.3 – DIAZ Corpus. Se trata de un corpus oral longitudinal, construido durante los años 1995-2000, a partir de las muestras de habla de estudiantes de español de diferentes L1 (alemán, chino, coreano, islandés y sueco). Las pruebas consistieron en entrevistas (datos espontáneos) y cuestionarios estructurados (datos experimentales). La compilación de este conjunto de textos se hizo tomando como base los estudios desarrollados en el ámbito del proyecto *Beyond Parameters* (Universidad de Ottawa) en la última década

² No pretendemos ser exhaustivas, ya que de algunos de los proyectos apenas hemos recabado unos pocos datos (de hecho, hemos dejado fuera los corpus multilingües de aprendices que incluyen el español entre las lenguas meta). Para más información pueden visitarse las direcciones que recogemos al final del capítulo, antes de las referencias, y el sitio del *Centre for English Corpus Linguistics* de la Universidad Católica de Lovaina, que ofrece una tabla con los datos principales de los corpus de aprendices existentes (*Learner corpora around the world*): <https://www.uclouvain.be/en-cecl-lcworld.html> (última actualización 15/09/2016) [visita 30/10/2016].

del siglo pasado (Díaz, 2007). Es accesible a través del catálogo de recursos lingüísticos del sitio de OLAC (*Open Language Archives Community*).

2.1.4 – SpinTX (Corpus vídeo del español hablado en Texas). Todavía en construcción, recoge muestras orales en español (entrevistas y conversaciones sobre 18 temas) de hablantes bilingües español-inglés de diferentes regiones del estado de Texas (actualmente $N = 97$). El objetivo de este proyecto es definir el español hablado en Texas y proporcionar herramientas de aprendizaje a estudiantes, profesorado y público general para explorar la variación de la lengua española en este contexto concreto. Es accesible a través del archivo de vídeos e incluye transcripciones completas y anotaciones de categoría gramatical que se ofrecen como recurso auténtico para ejemplificar explicaciones gramaticales del español. Todo ello se presenta en un portal bilingüe de espléndido diseño y completa información. Las directoras del proyecto desarrollado en el seno de COERLL (*Centre for Open Educational Resources and Language Learning*) son B. E. Bullock y A. J. Toribio, de la Universidad de Texas en Austin (Bullock & Toribio, 2011).

2.1.5 – SPLLOC (Spanish Learner Language Oral Corpus I y II). Se trata del corpus oral del español accesible más grande recogido hasta el momento. Reúne muestras de aprendientes de español ($N = 120$), de nivel principiante a avanzado, de escuelas y universidades del Reino Unido (L1 = inglés). Los datos proceden de narraciones, entrevistas y descripciones de fotografías y discusión entre iguales, recogidas en dos fases (2006-2008 y 2009-2010) por un equipo de la Universidad de Southampton. El objetivo de este corpus es profundizar en la adquisición del español como L2, por lo que ha seguido en su diseño criterios adecuados para la investigación en adquisición de segundas lenguas (según las pautas de su precedente FLLOC, *French Learner Language Oral Corpus*) (Mitchell et al., 2008). Es una de las colecciones de textos orales que cuenta con más publicaciones derivadas y constituye un hito en la investigación de la adquisición del español como segunda lengua y en el ámbito de la adquisición de segundas lenguas en general (Mendikoetxea, 2014, pp. 16-17).

2.1.6 – SPT (Spanish corpus Proficiency level Training). Este corpus audiovisual está construido a partir de grabaciones en vídeo de aprendientes de español de diferentes niveles (principiante a avanzado) ($N = 25$), con L1 inglés y el español como lengua de herencia (16 de 25), que contestan a diferentes preguntas que se les plantean sobre nueve temas diferentes. El corpus ofrece la transcripción de cada una de las muestras orales y ejercicios —con sus soluciones—, que guían de forma interactiva para indicar

diferentes rasgos acerca de la interlengua de los informantes. El proyecto responde a la demanda en la formación de profesores y acreditadores de niveles de competencia en español. El objetivo es, por tanto, enseñar a evaluar los niveles de conocimiento del español (Koike, 2007). El corpus es accesible en línea a través de COERLL (*Centre for Open Educational Resources and Language Learning*), de la Universidad de Texas, en Austin.

2.2 – Corpus de aprendices de español escritos

2.2.1 – Aprescrliv (*Aprender a Escribir en Lovaina*). Se trata de un amplio corpus de español como segunda lengua (L1 = holandés), que contiene un millón de palabras. Recoge muestras escritas (2700 textos) de estudiantes de lengua española de la Facultad de Letras de la Universidad Católica de Lovaina y de la Lessius Hogeschool, en los niveles A1 a C1. Su acceso es restringido a través de la solicitud a los investigadores, que han mostrado las posibilidades de explotación con un ejemplo de análisis de errores con el uso de *ser* y *estar*, una investigación sobre verbos de cambio y otra sobre preposiciones (Buyse & González Melón, 2013; Buyse, Fernández Pereda & Verveckken, 2015, p. 2016).

2.2.2 – CAES (*Corpus de aprendices de español como lengua extranjera*). Constituye el proyecto más reciente y ambicioso de corpus escritos de aprendices de español. Promovido y financiado por el Instituto Cervantes, se puede consultar en línea desde octubre de 2014 (versión 1.0). Ha sido diseñado y desarrollado íntegramente por un equipo de investigación de la Universidad de Santiago de Compostela (USC), bajo la dirección de G. Rojo e I. M. Palacios Martínez, gracias a la colaboración de un gran número de profesores de diferentes centros del Instituto Cervantes y de diversas universidades de muy dispares lugares del mundo. En la versión actual comprende muestras de aprendices con seis L1: árabe, chino mandarín, francés, inglés, portugués y ruso. La recogida de muestras se llevó a cabo a partir de unas pruebas de expresión escrita, que requerían de los aprendices, dependiendo del nivel, la producción de diferentes tipos de textos, desde descriptivos a argumentativos (algunos ejemplos son: mensajes de correo electrónico para amigos o familiares, solicitudes de trabajo, reseña crítica de un libro, hacer una queja, escribir una anécdota divertida, reservar una habitación en un hotel). El resultado final son los 575.000 elementos lingüísticos que incluye en la actualidad, con una distribución que abarca casi todos los niveles de dominio lingüístico: del A1 al C1 ($N = 1423$). La compilación de este corpus tiene como objetivo principal, según se explica en la presentación de la aplicación de consulta, ofrecer una herramienta que permita a los profesionales de la enseñanza del español como lengua extranjera

—profesorado, investigadores, evaluadores, creadores de materiales didácticos, responsables y equipos de centros e instituciones lingüísticas, etc— llevar a cabo investigaciones aplicadas sobre la base de datos sólidos y objetivos, ya que puede proporcionar información sobre dificultades de aprendizaje, errores más comunes, vocabulario más o menos empleado, etc., que se podrá aplicar con facilidad en las aulas o integrar en los manuales o materiales (Instituto Cervantes, 2014). El diseño ha seguido criterios unitarios y las muestras han recibido anotación morfosintáctica automática y desambiguación manual, según un sistema desarrollado por el equipo de la USC. El resultado se ha volcado en una aplicación informática de consulta que permite búsquedas simples y combinadas de acuerdo con variables lingüísticas, personales y sociales (Rojo & Palacios Martínez, 2016).

2.2.3 – CATE (*Corpus de Aprendices Taiwaneses de Español*). Se trata de un corpus construido con ensayos de estudiantes taiwaneses que aprenden español (L1 = chino), con un total de cerca de dos mil redacciones (340.000 palabras). Sigue en construcción en la Universidad Nacional Cheng-Kung de Taiwán (Lu 2012).

2.2.4 – CEDEL2 (*Corpus Escrito de Español L2*). La composición de este corpus comenzó en 2006 como parte del proyecto WOSLAC (*Word Order in Second Language Acquisition Corpora*) en la Universidad Autónoma de Madrid, con la participación también de la Universidad de Granada. Se trata de un corpus que recoge las muestras escritas de hablantes ingleses que aprenden español desde el nivel A1 al C1. El objetivo inicial fue doble: explorar el papel de la Hipótesis de la interfaz en las gramáticas de L2 usando datos de corpus y compilar dos corpus comparables de aprendices que fueran adecuados para la investigación en adquisición de L2 (CEDEL2 para español segunda lengua y *WriCLE*, *Written Corpus of Learner English*, para inglés segunda lengua). Aunque sigue en elaboración, en la actualidad tiene ya aproximadamente 750.000 palabras procedentes de cerca de 2400 participantes, y es accesible para la investigación. La explotación de CEDEL2 —ahora también junto a la de *WriCLE*— cuenta con numerosas investigaciones publicadas y en curso de gran interés para la adquisición del español como segunda lengua (Lozano, 2009; Lozano & Mendikoetxea, 2013; Mendikoetxea, 2014).

2.2.5 – CORANE (Corpus para el Análisis de errores de aprendices de E/LE). Este corpus contiene 1091 textos escritos por 321 estudiantes de los cursos de español como lengua extranjera de la Universidad de Alcalá de Henares (los niveles de dominio lingüístico van del A1 hasta el C1). Ha sido

publicado en formato CD-ROM (Cestero Mancera et al., 2001; Cestero Mancera & Penadés, 2009).

2.2.6 – ESPALEX para el estudio de la adquisición del español como lengua extranjera se ha construido sobre la base de los exámenes para la obtención de los Diplomas de Español como Lengua Extranjera (nivel B2). En la actualidad el corpus recoge 12.576 textos distintos referidos a 50 temas diferentes agrupados en tomo a 5 grandes tipos de textos: carta formal, carta informal, texto narrativo, texto descriptivo y texto expositivo. Los textos tienen una longitud media de 175 palabras. El proyecto fue llevado a cabo por la Universidad de Salamanca y el Instituto Cervantes (Bustos Gisbert & Sánchez Iglesias, 2011).

2.2.7 – SCIL (Spanish Corpus of Italian Learners). Corpus nacido de la necesidad de obtener material para analizar la interlengua de estudiantes de español con lengua nativa italiano, así como para el desarrollo de materiales didácticos. Es un corpus longitudinal que recoge 457 composiciones, sobre un único tema, de estudiantes universitarios italianos de español como lengua extranjera ($N = 43$), con niveles de competencia de A1 a B2. Las muestras fueron recogidas a lo largo de 7 meses (de diciembre de 2008 a julio de 2009) en grupos de e-tándem entre estudiantes españoles que aprendían italiano y estudiantes italianos que aprendían español, aunque el corpus solo recoge las muestras en español de los estudiantes italianos (cartas informales a través del correo electrónico) (Bailini, 2013).

2.3 – Corpus de aprendices de español oral y escrito

LANGSNAP (*Language and Social Networks Abroad Project*). Se trata de un corpus longitudinal oral y escrito con producciones recogidas a lo largo de un periodo de 21 meses de estudiantes universitarios de nivel avanzado de español y francés (L1 = inglés). Llevado a cabo por miembros de los equipos de SPLLOC y FLLOC, tiene la intención de documentar el desarrollo de la competencia y uso de la lengua meta por los estudiantes en un periodo que incluye una estancia en el extranjero, en situación de inmersión, de nueve meses (Tracy-Ventura, Mitchell & McManus, 2016). El repositorio accesible incluye tres tipos de datos para cada lengua, recogidos en seis ocasiones: entrevistas orales semidirigidas por un miembro del equipo investigador, narración de historias guiada por una secuencia de dibujos y escritura argumentativa como respuesta a una pregunta-estímulo. Además de las grabaciones de audio están disponibles las transcripciones de todos los datos.

3 – EXPLOTACIÓN DE LOS CORPUS DE APRENDICES DE ESPAÑOL

Si pensamos en la explotación de los corpus de aprendices, una cuestión es qué datos podemos extraer de ellos y otra muy diferente es cómo usar de manera eficiente, y con qué finalidad, la información que obtengamos en los análisis. Puesto que se trata de un campo de investigación aún en proceso de consolidarse en el caso del español, la situación es desigual, en buena medida a causa de las características de los corpus existentes, que con frecuencia condicionan la aplicación de los resultados. Por una parte, como hemos visto, la mayoría de los corpus de aprendices de español se orientaron en principio hacia la enseñanza del español como lengua extranjera, al menos se compilaron con ese interés y se situaron en el terreno del análisis de errores, por lo que tuvieron en cuenta informantes de diversas L1.

La metodología seguida fue en sus inicios característicamente de tipo descriptivo, y en muchos casos sigue siéndolo: toma las producciones de los corpus como punto de partida para detectar los errores, que, una vez identificados y descritos, se ubican en una clasificación elaborada para tal finalidad (CORANE y CORELE, por ejemplo). La aportación para la didáctica del español como segunda lengua no siempre se hace explícita, pero en ciertos trabajos sí se destacan algunas de las implicaciones pedagógicas de los hechos estudiados, encaminadas sobre todo a mejorar la elaboración de materiales o a la orientación al profesorado sobre lo que es más conveniente de cara a evitar determinados errores en la lengua de los aprendices. Tomemos como ejemplo de esta vía el estudio contrastivo entre un corpus nativo y un corpus de aprendices —de L1 inglés—, ambos de lengua escrita —se trata de CEDEL2—, sobre las dificultades de aprendizaje de colocaciones del español (Orol González & Alonso Ramos, 2013). Las autoras parten de la evaluación de la producción escrita de colocaciones españolas en ambos corpus para determinar la riqueza de las colocaciones empleadas. En su conclusión hacen referencia a las implicaciones para la didáctica:

[...] the results point the way towards the elaboration of useful guidelines that will make learning collocations easier. On the one hand, the work with corpora has allowed us to reject some common assumptions. For instance, we have proved that learners do, indeed, use collocations, but their choices do not have the degree of variety, sophistication and correction of native speakers. Therefore, some assumptions such as the notion that learners use few collocations are shown to be false. On the other hand, the parameters that define native speakers' collocational richness indicate that error is the key factor in terms of producing correct collocations. This could certainly be taken as an incentive for placing greater emphasis on the relevance of collocations in didactic material. (p. 570)

En el mismo terreno de las colocaciones, aunque centrado en las construcciones con verbo soporte a partir del estudio contrastivo de los mismos datos, el trabajo de García Salido (2014) concluye en una dirección semejante, pero en este caso no menciona la utilidad para la enseñanza de estas construcciones españolas:

Se desde o punto de vista da frecuencia das CVS [construccións con verbo soporte] non se pode considerar que exista unha infrautilización deste recurso por parte dos aprendices, polo menos no corpus analizado, si que se pode afirmar que existen diferenzas con respecto á variedade de CVS que se rexistra nas producións de nativos e aprendices. Os primeiros empregan un maior número de CVS diferentes que os segundos, o que se traduce nun maior número de lemas para un conxunto equivalente de ocorrencias. Non hai, polo tanto, infrautilización, senón menor riqueza léxica e maior repetición de CVS coñecidas por parte dos aprendices. (p. 198)

Desde luego, la detección, descripción y análisis de errores es un recorrido esencial para la enseñanza de lenguas, pero requiere una fase ulterior en la que se afronte la aplicabilidad de los hallazgos.

En cuanto al ámbito de la adquisición de segundas lenguas, los estudios sobre el español se han centrado en analizar los corpus de aprendices —sobre todo SPLLOC y CEDEL2— a la luz de las hipótesis formuladas en el seno de la adquisición de segundas lenguas, por ello emplean un subcorpus de hablantes nativos como elemento de control, se refieren en exclusiva a informantes de L1 inglés y, sobre todo, se proponen evaluar algunas de las hipótesis actuales de la teoría de la adquisición de segundas lenguas. Se trata de trabajos que en general no se plantean una aplicabilidad directa en la didáctica de lenguas, recordemos que son de carácter psicolingüístico, por lo que el acceso a los logros y conclusiones que aportan se presenta como una tarea ardua desde la enseñanza de español como lengua extranjera. Para ilustrar este hecho, consideremos, como ejemplo, uno de los trabajos basados en las colecciones de datos de lengua oral de SPLLOC, el de Arche & Domínguez (2011). Estas autoras investigan la adquisición de los pronombres clíticos españoles —marcados para caso, género y número— por estudiantes cuya L1 es el inglés, a través de pruebas de tipo experimental, con tareas centradas en producción y en interpretación, diseñadas específicamente para la evaluación de hipótesis de adquisición de segundas lenguas. Tal y como está expresado en su resumen:

We examine the relationship between morphology and syntax in L2 learners' grammars in order to assess two current acquisition hypotheses: the Impaired Representation Hypothesis (IRH) and the Missing Surface Inflection Hypothesis (MSIH). Data from a production and a comprehension task suggest that learners have an unimpaired narrow syntax, despite apparent inflectional variability. We propose that absent or inaccurate morphology can be

explained by a deficit in the mapping to PF [Phonetic Form]. This supports a dissociation between syntactic representation and surface inflection. (p. 291)

Como vemos, es imprescindible conocer los presupuestos teóricos de la adquisición de segundas lenguas y estar familiarizados con las varias hipótesis propuestas para sopesar el interés y relevancia de cada estudio concreto. En este caso, los datos obtenidos de sus tareas experimentales muestran que la precisión en la producción e interpretación de los clíticos está, como era de esperar, en relación directa con el nivel de competencia: en el nivel avanzado la precisión es muy buena, mientras que en el nivel de principiante las puntuaciones son muy bajas, tanto en producción como en comprensión. Ello les permite a las autoras afirmar que la adquisición de formas con rasgos no representados en la L1 es posible, es decir, que la adquisición de los clíticos españoles es asequible para los hablantes ingleses. Destacan una matización interesante en el caso de los informantes de nivel intermedio: aunque puntúan muy alto en la tarea de comprensión, presentan un escaso uso de clíticos en la tarea de producción. Esta desigualdad les sirve a las investigadoras para sugerir que la representación interna (sintáctica) se encuentra en perfectas condiciones, lo cual se deja traslucir en la tarea de comprensión, por lo tanto las dificultades, patentes en las tareas de producción, surgen en un módulo más externo, el de la morfología (Arche & Domínguez, 2011, pp. 315-6). Como se percibe en esta breve síntesis, toda la investigación está dirigida a evaluar, justificar y apoyar algunos de los presupuestos teóricos de la adquisición de segundas lenguas y no resulta fácil, entre otros motivos porque no es lo que este estudio pretende, extraer los aspectos aprovechables para la enseñanza de los clíticos del español a estudiantes cuya lengua inicial es el inglés. Por otra parte, el conjunto de datos que ha servido de base para la investigación de Arche & Domínguez (2011) no es propiamente lo que llamamos un corpus de aprendices, pues se ha extraído de tareas experimentales diseñadas específicamente con un objetivo concreto. Otros aspectos estudiados a partir de SPLLOC incluyen el orden de palabras (SV-VS), la progresión en la adquisición del léxico (con una comparación entre hablantes de L1 inglés y L1 francés) o la evolución en la adquisición del sistema de tiempo-aspecto³, siempre en la línea del trabajo reseñado sobre los clíticos, aunque tomando como base nuevos datos (correspondientes también al proyecto SPLLOC II).

En lo que se refiere a los corpus de textos escritos, ya hemos visto que CEDEL2 constituye un proyecto sólido que cuenta con numerosos estudios

³ Hay un somero resumen de metodología y conclusiones en Mendikoetxea (2014, pp. 16-17), donde es posible encontrar reunidas todas las referencias bibliográficas.

publicados e investigaciones en curso. Entre los factores de adquisición estudiados a partir de él se encuentran los sujetos pronominales, el uso del pronombre *se* con predicados inacusativos, el orden de palabras o las colocaciones empleadas en el discurso de los aprendices⁴. Como en el caso de los estudios basados en SPLLOC, buena parte de las publicaciones están encaminadas a comprobar las hipótesis de la teoría general de la adquisición de segundas lenguas, en especial la Hipótesis de la interfaz (*Interface Hypothesis*). Subrayamos, también en este caso, la dificultad de hacer un trasvase efectivo de los hallazgos concretos a la enseñanza del español como segunda lengua.

Aún en vías el proceso de explotación, y con perspectiva de futuros incrementos —de muestras orales en videograbaciones, ampliando al nivel de competencia C2 y con informantes de L1 no representadas de momento— el CAES se presenta como una plataforma de referencia para todo tipo de aprovechamiento didáctico. Por ahora apenas contamos más que con sugerencias de explotación, que recogen ejemplos de determinados fenómenos obtenidos a través de la aplicación de consulta: los falsos amigos del español para hablantes nativos de inglés, francés y portugués o las preposiciones del español más usadas por los aprendices y las que entrañan más dificultad, así como muestras de cambio de código (Palacios Martínez, 2016), el empleo del pronombre *ello* por aprendices y por nativos (Parodi, 2015), son algunos ejemplos. La verdad es que el CAES ofrece datos de aprendices con lenguas iniciales apenas representadas en el panorama de investigación en corpus de aprendices de español, como el árabe o el ruso, por lo que ofrece interesantes oportunidades al profesorado de español correspondiente.

Como subraya Parodi en su reseña (2015), el CAES se define como un apoyo a la investigación de especialistas profesionales, es decir, se trata de un «corpus de aprendientes y para profesionales expertos» (p. 195). El CAES destaca por dos aspectos: en primer lugar, por su herramienta de búsqueda, la sencilla y práctica aplicación de consulta a través de la que es posible seleccionar las variables relacionadas con los informantes; en segundo lugar, por la valiosa información disponible sobre las características del corpus que «permite construir una visión de conjunto de la procedencia de los textos y de las fortalezas y limitaciones que ellos encierran» (Parodi, 2015, p. 195), así como por su *Guía de consulta de CAES*, que es la que indica cómo moverse a través de la plataforma de una manera sencilla y manejar una herramienta

⁴ Véanse los trabajos recogidos y comentados en Mendikoetxea (2014, pp. 19-20). En concreto sobre el orden de palabras pueden consultarse los resúmenes y conclusiones de Lozano (2014) y, a propósito de la competencia léxica de los aprendices, hay una evaluación de los trabajos sobre el uso de los verbos copulativos (*ser/estar*), la distribución de pronombres plenos y nulos, las colocaciones y otros aspectos —marcadores del discurso y preposiciones— en Lozano (2015).

que nos permitirá profundizar y abordar desde otra perspectiva problemas lingüísticos de distinta índole en el campo de la enseñanza-aprendizaje del español como lengua extranjera. Las potencialidades para la investigación —infinitas, al decir de Parodi (2015, p. 198)— están aún por desplegarse, pero es indudable que el CAES constituye una aportación que marcará un antes y un después en el terreno de la investigación de corpus de aprendices y la enseñanza de español como lengua extranjera.

4 – PERSPECTIVAS DE FUTURO

El panorama que hemos trazado a lo largo de estas páginas brinda una perspectiva prometedora en el futuro inmediato de la investigación en corpus de aprendices y, en concreto, en las posibilidades de continuar con estudios de fenómenos ya revisados u otros no considerados hasta ahora. Los nuevos datos ofrecidos por CAES, además de las ampliaciones previstas para algunos de los proyectos presentados más arriba —en el apartado 2— permitirán conocer el comportamiento interlingüístico de aprendices de español que tienen como lengua nativa alguna de las nunca consideradas anteriormente. A nuestro modo de ver, en sociedades multilingües —especialmente europeas, pero no solo— es bueno ver que este campo se nutre de un número más amplio de L1, no del inglés en exclusiva. Por otra parte, destacamos también la importancia de tener en cuenta la competencia plurilingüe, pues el conocimiento lingüístico previo de los aprendices incluye la L1, pero también las otras lenguas que los estudiantes hayan adquirido o estén en proceso de adquirir. Con toda probabilidad, los estudiantes universitarios actuales se encuentran en proceso de aprendizaje de dos o más lenguas y, con toda probabilidad, parten de una situación de bilingüismo. Al estudio del español se llega en Europa desde un conocimiento de inglés, francés o alemán como segundas lenguas, promovido por la movilidad estudiantil y el fomento del Espacio Europeo de Educación Superior.

Como es natural, nuestro conocimiento de la adquisición del español como segunda lengua mejorará cuando contemos con más colecciones, especialmente más corpus orales —en formato audiovisual a ser posible— y más de carácter longitudinal, que permitan corroborar o contrastar los hallazgos actuales. Es esencial hacer hincapié en el carácter interdisciplinar de la lingüística de corpus de aprendices, así como en la relevancia de cuidar los criterios en el diseño de los corpus, la metodología en la anotación y la rigurosidad en las investigaciones. La disponibilidad de corpus que reúnan estas características supondrá un fuerte espaldarazo para explotar de manera eficiente, y en múltiples direcciones, las colecciones de datos.

Ello nos lleva, por último, a destacar la fuerte orientación aplicada de la investigación de corpus de aprendices. Afrontar una integración o,

al menos, un trasvase fructífero de las aportaciones de la investigación psicolingüística —en el ámbito de la adquisición de segundas lenguas— y las de la investigación en enseñanza de español como lengua extranjera constituye un verdadero reto, imprescindible para un avance definitivo en la didáctica del español. El diseño de materiales y programaciones, así como la formación del profesorado, se beneficiarán enormemente de los nuevos hallazgos aportados por este interesante terreno en expansión. Los corpus de aprendices se revelan, pues, como una pieza clave que marcará la hoja de ruta de las investigaciones encaminadas a conocer los mecanismos de la adquisición y el aprendizaje de lenguas.

Sitios de los corpus con información o acceso en línea [visita 30/10/2016]

CAES <http://galvan.usc.es/caes>

CATE <http://corpora.fld.ncku.edu.tw/#>

CEDEL2 http://wdb.ugr.es/~cristoballozano/?page_id=64

<http://www.uam.es/proyectosinv/woslac> (última actualización 19/01/2009)

CORELE <http://cartago.llf.uam.es/corele/index.html>

CORINÉI <https://dti.ua.es/es/teletandem-corinei/corinei/corinei.html>

DIAZ Corpus http://dla.library.upenn.edu/dla/olac/record.html?id=talkbank_org_SLABank-Spanish-DiazRodriguez

LANGSNAP <http://langsnap.soton.ac.uk/>

SpinTX <http://spanishintexas.org/es/>

SPLLOC <http://www.splloc.soton.ac.uk/>

SPT <http://www.laits.utexas.edu/spt/intro>

Referencias bibliográficas

Alonso Ramos, M. (Ed.) (2016). *Spanish learner corpus research. Current trends and future perspectives*. Amsterdam: John Benjamins.

Arche, M. J. & Domínguez, L. (2011). Morphology and syntax dissociation in SLA. A Study on clitic acquisition in Spanish. En Galani, A., Hicks, G. & Tsoulas, G. (Eds.), *Morphology and its Interfaces* (pp. 291-320). Amsterdam & Philadelphia: John Benjamins.

Bailini, S. (2013). SCIL: A Spanish Corpus of Italian Learners. *Procedia — Social and Behavioral Sciences, Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013)* 95, 542-549.

Baralo Ottonello, M. (2004). La interlengua del hablante no nativo. En Sánchez Lobato, J. & Santos Gargallo, I. (Dirs.), *Vademécum para la formación de profesores. Enseñar Español como segunda lengua (L2)/ Lengua extranjera (LE)* (pp. 369-389). Madrid: SGEL.

Bullock, B. & Toribio, A.J. (2011). Creating a linguistic landscape of Spanish speakers in the US, starting in Texas. From an August 2011 interview with COERLL [*Centre for Open Educational Resources and Language Learning*]. Recuperado de <http://sites.la.utexas.edu/voices/barbara-bullock-and-almeida-jacqueline-toribio/> [visita 30.10.2016].

Bustos Gisbert, J. M. & Sánchez Iglesias, J. J. (2011). Espalex: un corpus para el estudio de la adquisición del español como lengua extranjera. En Hernández González, C., Carrasco Santana, A. & Álvarez Ramos, E. (Coord.), *La Red y sus aplicaciones en la enseñanza-aprendizaje del español como lengua extranjera* (pp. 149-159). Valladolid: ASELE.

Buyse, K. & González Melón, E. (2013). El corpus de aprendices Aprescrilov y su utilidad para la didáctica de ELE en la Bélgica multilingüe. En Blecua, B., Borell, S., Crous, B. & Sierra, F. (Eds.), *Plurilingüismo y enseñanza de lenguas* (pp. 247-252). Girona: ASELE.

Buyse, K., Fernández Pereda, L. & Verveckken, K. (2015). The reference to L1 and L2 in SFL: proposals based on the Aprescrilov learner corpus. *Procedia: Social and Behavioral Sciences*, 173, 274-278.

Buyse, K., Fernández Pereda, L. & Verveckken, K. (2016). The Aprescrilov corpus, or broadening the horizon of Spanish language learning in Flanders. En Alonso Ramos, M. (Ed.), *Spanish learner corpus research: current trends and future perspectives* (pp. 143-168). Amsterdam: John Benjamins.

Campillos Llanos, L. (2012). *La expresión oral en español lengua extranjera: interlengua y análisis de errores basado en corpus*. Tesis doctoral UAM.

Campillos Llanos, L. (2014). A Spanish Oral Learner Corpus for Computer-Aided Error Analysis. *Corpora* 9(2), 207-238.

Cestero Mancera, A. M. & Penadés, I. (2009). *Corpus de textos escritos para el análisis de errores de aprendices de E/LE (CORANE)*. CD-ROM. Universidad de Alcalá.

Cestero Mancera, A. M. *et al.* (2001). Corpus para el análisis de errores de aprendices de E/LE (CORANE). En Gimeno Sanz, A. M. (Ed.), *Tecnologías de la información y de las comunicaciones en la enseñanza de ELE: Actas del XII Congreso Internacional de ASELE* (pp.527-534). Valencia: Universidad Politécnica de Valencia.

Conrad, S. (2005). Corpus Linguistics and L2 Teaching. En Hinkel, E. (Ed.), *Handbook of research in Second Language Teaching and Learning* (247-256). Mahwah, NJ: L. Erlbaum Associates.

Díaz, L. (2007). *Interlengua Española: Estudio de Casos*. Barcelona: Printulibro Intergrup.

García Salido, M. (2014). O uso de construcións con verbo soporte en aprendices de español como lingua estranxeira e en falantes nativos. *Cadernos de Fraseoloxía galega*, 16, 181-198.

Granger, S. (2002). A Bird's-eye view of learner corpus research. En Granger, S., Hung, J. & Petch-Tyson, S. (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3–33). Amsterdam: John Benjamins.

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. En Aijmer, K. (Ed.), *Corpora and Language Teaching* (pp.13-32). Amsterdam & Philadelphia: Benjamins.

Granger, S., Gilquin, G. & Meunier, F. (2015). Introduction: Learner corpus research —past, present and future. En Granger, S., Gilquin, G. & Meunier, F. (Eds.), *The Cambridge Handbook of Learner Corpus Research* (pp. 1-5). Cambridge: Cambridge University Press.

Instituto Cervantes (2014). *Corpus de aprendices de español como lengua extranjera (CAES)*. Aplicación y guía de consulta en <http://galvan.usc.es/caes> [visita 30.10.2016].

Koike, D. (2007). *Spanish Learner Corpus and Exercises*. Austin: Universidad de Texas.

Lafford, B. & Salaberry, M. R. (Eds.). (2003). *Spanish Second Language Acquisition: State of the Science*. Washington: Georgetown University Press.

Lozano, C. & Mendikoetxea, A. (2013). Learner Corpora and SLA: The Design and Collection of CEDEL2. En Díaz-Negrillo, A., Ballier, N. & Thompson, P. (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data* (pp. 65–100). Amsterdam: John Benjamins.

Lozano, C. (2009). CEDEL2: Corpus Escrito del Español L2. En Bretones Callejas, C. M. *et al.* (Eds.), *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente* (pp. 197-212). Almería: Universidad de Almería.

Lozano, C. (2014). Word Order in Second Language Spanish. En Geeslin, K. L. (Ed.), *The Handbook of Spanish Second Language Acquisition* (pp. 287-310). Oxford: Wiley-Blackwell.

Lozano, C. (2015). Learner corpora as a research tool for the investigation of lexical competence in L2 Spanish. *Journal of Spanish Language Teaching*, 2 (2), 180-193.

Lu, H. (2012). An Annotated Taiwanese Learners' Corpus of Spanish, CATE. *Corpus Linguistics and Linguistic Theory*, 6(2), 297-300.

Mark, K. L. (1998). The Significance of Learner Corpus Data in Relation to the problems of Language Teaching. *Bulletin of General Education*, 312, 77-90.

Mendikoetxea, A. (2014). Corpus-Based research in Second Language Spanish. En Geeslin, K. L. (Ed.), *The Handbook of Spanish Second Language Acquisition* (pp. 11-29). Oxford: Wiley-Blackwell.

Mitchell, R., Domínguez, L., Arche, M., Myles, F. & Marsden, E. (2008). SPLLOC: A New Corpus for Spanish Second Language Acquisition

Research. En Roberts, L., Myles, F. & David, A. (Eds.), *EUROSLA Yearbook* 8 (pp. 287-304). Amsterdam: John Benjamins.

Myles, F. (2005). Interlanguage Corpora and Second Language Acquisition Research. *Second Language Research* 21 (4), 373–391.

Orol González, A. & Alonso Ramos, M. (2013). A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish. *Procedia-Social and Behavioral Sciences*, 95, 563-570.

Palacios Martínez, I. M. (2016). Applications of learner corpora to Language Teaching. The CAES — Corpus de aprendices de español — project as a case in point. Comunicación presentada en el 34 Congreso Internacional AESLA, Alicante (España) 14-16 de abril de 2016.

Parodi, G. (2010). 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. (2015). Reseña del Corpus de aprendices de español (CAES). *Journal of Spanish Language Teaching*, 2(2), 194-200.

Rojo, G. & Palacios Martínez, I. M. (2016). Learner Spanish on computer: The CAES ‘Corpus de Aprendices de Español’ project. En Alonso Ramos, M. (Ed.). *Spanish learner corpus research: current trends and future perspectives* (pp. 55-87). Amsterdam: John Benjamins.

Rojo, G. (2016). Corpus textuales del español. En Gutiérrez-Rexach, J. (Ed.), *Enciclopedia Lingüística Hispánica* (pp. 285-296). Oxon: Routledge.

Seidlhofer, B. (2002). Pedagogy and local learner corpora. Working with learning-driven data. En Granger, S., Hung, J. & Petch-Tyson, S. (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 213–234). Amsterdam: John Benjamins.

Sinclair, J. (1996). *EAGLES: Preliminary recommendations on Corpus Typology*. Recuperado de <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html> / [30.10.2016].

Tracy-Ventura, N., Mitchell, R. & McManus, K. (2016). The LANGSNAP Longitudinal Learner Corpus: Design and Use. En Alonso Ramos, M. (Ed.). *Spanish learner corpus research: current trends and future perspectives* (pp. 117-142). Amsterdam: John Benjamins.

