



Los corpus de aprendices de español LE/L2: panorama actual y perspectivas futuras

Guillermo Rojo, Ignacio Palacios, María Sampedro Mella & Aurélie Marsily

To cite this article: Guillermo Rojo, Ignacio Palacios, María Sampedro Mella & Aurélie Marsily (2022) Los corpus de aprendices de español LE/L2: panorama actual y perspectivas futuras, Journal of Spanish Language Teaching, 9:2, 174-189, DOI: [10.1080/23247797.2022.2157085](https://doi.org/10.1080/23247797.2022.2157085)

To link to this article: <https://doi.org/10.1080/23247797.2022.2157085>



Published online: 11 Jan 2023.



Submit your article to this journal [↗](#)



Article views: 242



View related articles [↗](#)



View Crossmark data [↗](#)



Los corpus de aprendices de español LE/L2: panorama actual y perspectivas futuras

L2 Spanish learner corpora: current overview and future perspectives

Guillermo Rojo ^a, Ignacio Palacios^a, María Sampedro Mella^{a,*} and Aurélie Marsily^b

^aFacultad de Filología, Universidade de Santiago de Compostela, Santiago de Compostela, España; ^bLanguage and Communication Institute, Université catholique de Louvain, Louvain-la-Neuve, Bélgica

RESUMEN

Este capítulo tiene un objetivo doble. En primer lugar, tras una breve introducción delimitadora del tema, trataremos de describir las características de los corpus de aprendices (CA) de español como lengua extranjera o segunda (LE/L2) existentes en la actualidad y facilitar a los lectores todos los datos pertinentes: denominación, responsables, direcciones web, tamaño, naturaleza de las muestras (carácter escrito u oral), tipos de codificación incorporada, fecha de construcción, L1 de los participantes y condiciones actuales de uso. Toda esa información, detallada hasta donde sea posible, será finalmente resumida en un cuadro en el que se podrán apreciar con facilidad todas sus cualidades y también sus carencias. En segundo término, describiremos las aplicaciones más importantes que se pueden obtener de los CA para la mejora del proceso de enseñanza/aprendizaje de ELE, con una enumeración final de las líneas fundamentales por las que debería discurrir esta zona de trabajo en los próximos años.

ABSTRACT

The purpose of this study is twofold. Firstly, after a brief introduction narrowing down the area of study, we intend to describe the characteristics of the currently existing L2 Spanish learner corpora and provide the readers with the main data: denomination, team responsible for the corpus, websites, size, type of samples (written, spoken), type of data coding, construction date, participants' L1 and current conditions of use. All this information will be given as detailed as possible and will be summarised in a table where all the strengths and weaknesses of these corpora will be presented. We will secondly deal with the main applications of these learner corpora for the improvement of L2 Spanish teaching and learning processes together with some final notes containing a number of suggestions for future work.

ARTICLE HISTORY

Received 15 June 2022
Accepted 15 November 2022

PALABRAS CLAVE

Corpus textuales; corpus de aprendices; codificación; español como LE/L2; lingüística aplicada a la enseñanza de lenguas; interlengua

KEY WORDS

Text corpora; learner corpus; encoding; Spanish Language Teaching (SLT); linguistic studies applied to the teaching of languages; interlanguage

1. Introducción

Aunque no es un rasgo que suela aparecer destacado, lo cierto es que una parte importante de la evolución de la lingüística de corpus (LC) está determinada por su aplicación a la enseñanza y el

CONTACT Guillermo Rojo  guillermo.rojo@usc.es

*La participación de María Sampedro Mella en este trabajo se inscribe en el marco del proyecto "Explotación de corpus generales y de aprendices para la investigación y la enseñanza del discurso en el español como lengua extranjera (ELE)", financiado por la Xunta de Galicia (5040-I3JC-64100)

© 2023 Informa UK Limited, trading as Taylor & Francis Group

aprendizaje de lenguas en general y de las segundas lenguas y lenguas extranjeras en particular. En efecto, tras la aparición del Brown Corpus en 1964, esta nueva aproximación sufre unos años de marginación en los Estados Unidos, pero recibe, en cambio, una atención notable en Europa. Geoffrey Leech puso en marcha muy pronto la preparación de su contrapartida europea, el corpus de Lancaster-Oslo / Bergen (LOB), seguido por el Survey of Spoken English (SSE), desarrollado desde 1975 por Jan Svartvik en la Universidad de Lund, y, ya centrado en su aplicación a la enseñanza del inglés como L2, el corpus inicial del proyecto COBUILD, dirigido por John Sinclair. Svartvik (2007, 15) ha señalado explícitamente dos factores que, en su opinión, explican el éxito de esta orientación tanto en el Reino Unido como en los países del norte de Europa. De un lado, el hecho de que “in the long Scandinavian philological tradition in English studies, the text was central”, aspecto que se opone claramente a los postulados racionalistas y al uso de la introspección como herramienta fundamental de trabajo. En segundo lugar, sus posibilidades de aplicación al aprendizaje y enseñanza del inglés en todas sus dimensiones,¹ dado que

[t]o a non-native speaker of the language, the armchair approach of introspection is effectively ruled out. This may help to explain why certain parts of the world outside the English-speaking countries, such as northern Europe, were early strongholds of corpus linguistics.

En los primeros años, el trabajo consistió fundamentalmente en aprovechar las nuevas posibilidades que presentan los corpus y la informatización para obtener datos que permitan la confección de materiales más próximos a la lengua real, como sucede en el ya mencionado COBUILD, o bien en la utilización directa de los corpus en el proceso de enseñanza-aprendizaje de lenguas en la línea conocida como *data-driven learning* (cf. Leech 1997, 3).

En todas estas orientaciones, los corpus utilizados están formados por textos producidos por hablantes nativos. En 1990, en una línea vinculada al International Corpus of English (ICE),² Sylviane Granger lanza el proyecto denominado International Corpus of Learner English (ICLE). Tal como lo plantea desde el principio, se trata de aprovechar las nuevas posibilidades brindadas por la LC para superar los dos inconvenientes principales de las producciones de aprendices reunidas hasta ese momento: su escaso volumen (y, habitualmente, limitación a un tipo único de aprendices) y la ausencia de informatización, con la consiguiente dificultad de consulta para quienes no formaran parte del grupo recolector (Granger 1993). La importancia de los CA no ha dejado de crecer desde ese momento y se consolida con la creación de la *Learner Corpus Association* en 2013.

Los CA de español como lengua extranjera muestran un proceso parecido: surgieron con cierto retraso con respecto a los corpus en general y también a los CA de otras lenguas, pero no tardaron en alcanzar un grado de desarrollo aceptable. De todos modos, como se verá con detalle en los apartados siguientes, todavía en la actualidad se observa en este terreno la abundancia de CA de tamaños reducidos, centrados en un único tipo de aprendices y, en muchos casos, inaccesibles para quienes no han participado en su diseño y desarrollo.

2. Estado de la cuestión

2.1. Precisiones terminológicas y criterios de clasificación de los CA

De acuerdo con la definición más extendida, los CA son “electronic collections of natural or near-natural data produced by foreign or second language (L2) learners and assembled according to explicit design criteria” (Granger, Gilquin y Meunier 2015, 1). Esta definición integra, como es de esperar, los rasgos característicos de los corpus textuales y los aspectos que diferencian a los CA de todos los demás. Entre los primeros conviene destacar su naturaleza electrónica y el hecho de que han sido concebidos y contruidos con un determinado diseño, gracias al cual resultan representativos de la lengua o variedad lingüística correspondiente y pueden ser utilizados para su estudio científico (cf. Rojo 2021, 1-3). En la parte diferencial, el rasgo distintivo básico es, por supuesto, que están compuestos por producciones de aprendices de una o varias L2. Interesa