LOS CORPUS DE APRENDICES COMO RECURSO DIDÁCTICO EN EL AULA DE ESPAÑOL/LE PARA ITALÓFONOS

SONIA BAILINI
Università Cattolica del Sacro Cuore, Milano

Introducción

En los últimos años, en el ámbito de la investigación en lingüística de corpus aplicada a la enseñanza de lenguas extranjeras, se ha destacado la importancia de los corpus de aprendices (*learner corpora*) como una herramienta aprovechable también en el aula (Granger, 2012; Granger *et al*., 2015; Boulton y Pérez-Paredes, 2014; Mendikoetxea, 2014; Alonso Ramos, 2016; Boulton, 2017; Abad Castelló, 2019). Sin embargo, los *learner corpora* de ELE disponibles en línea no suelen incluir muestras de la interlengua de aprendientes italianos y, cuando las hay, su número es tan exiguo que no llega a ser representativo. El Corpus del Español de los Italianos (CORESPI) (Bailini, 2018b) trata de colmar este vacío, ya que reúne una cantidad de muestras representativa de la interlengua del español de los italianos, es una herramienta de acceso libre y es consultable gracias a una interfaz flexible y fácil de usar[1].

Este trabajo está estructurado en cuatro partes: en la primera se analiza el potencial de los *learner corpora*, bien como bases de datos para el profesor, bien como recursos que el aprendiente puede consultar para profundizar en el conocimiento de la lengua que está aprendiendo. En la segunda y tercera parte, tras una breve síntesis de los corpus de aprendices de español existentes, se presentan las características de CORESPI y se ilustran las funciones de su interfaz de consulta (CATMA). Por último, en la cuarta parte se incluyen algunas propuestas de aprovechamiento de CORESPI como herramienta didáctica. Desde esta perspectiva, la interlengua representa el punto de partida para estimular una reflexión lingüística contrastiva español-italiano que ayude al aprendiente a adquirir un conocimiento más profundo y consciente de las convergencias y divergencias entre su lengua materna y la lengua afín que está aprendiendo.

1. El potencial de los *learner corpora* para el aprendizaje de una lengua extranjera

Como sabemos, los corpus son conjuntos extensos de textos representativos de cierta lengua o variedad lingüística, recogidos según criterios preestablecidos de acuerdo con sus finalidades, almacenados en formato electrónico y consultables a través de aplicaciones informáticas específicamente pensadas para este fin (Sinclair, 2005; Cruz Piñol, 2012 y 2021; Rojo Sánchez, 2021). Los corpus de aprendices o *learner corpora* son recopilaciones de textos —orales o escritos— producidos por aprendientes de una lengua diferente de su L1. Granger los define como

> electronic collections of authentic Foreign Language/Second Language textual data assembled according to explicit designed criteria for a particular Second Language Acquisition/Foreign Language Teaching purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance. (Granger, 2002, p. 7)

Los corpus de aprendices reúnen formas correctas e incorrectas del idioma que el informante está aprendiendo. Su interés se debe a que, al contener una gran cantidad de datos procedentes de un número elevado de aprendientes, son una interlengua colectiva. Esto es lo que les confiere representatividad y posibilita compaginar el análisis cualitativo con el cuantitativo, aumentando el rigor metodológico y la solidez de la investigación empírica en lingüística adquisicional (Sánchez Rufat, 2015). En estos corpus la producción en la LE se puede analizar en su conjunto como discurso contextualizado del aprendiente, en el que no solo se observan aspectos relacionados con la morfología y la sintaxis, sino también con el

[1] Se puede acceder a CORESPI desde la página https://corespiycorite.altervista.org/ [consultado: 30/05/2023].

léxico, la fraseología, la pragmática y las convenciones textuales. Por esta razón los *learner corpora* se han convertido en una herramienta valiosa para el análisis de la interlengua y las secuencias de adquisición. Sin embargo, su empleo para la enseñanza-aprendizaje de la LE sigue siendo un terreno relativamente poco explorado (Martín Sánchez *et al.*, 2017; Abad Castelló, 2019; Palacios Martínez *et al.*, 2019; Ballestero de Celis y Sampedro Mella, 2021; Rojo Sánchez y Palacios Martínez, 2022). Más Álvarez y Gil Martínez subrayan la necesidad de trasvasar de manera fructífera los hallazgos de la investigación en adquisición de segundas lenguas a la enseñanza y definen los corpus de aprendices 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" (Más Álvarez y Gil Martínez, 2018, pp. 52-53).

Si bien los corpus de aprendices nacen como bases de datos para la investigación en lingüística computacional, también pueden aprovecharse como recurso didáctico en el aula de lenguas extranjeras, siempre y cuando el profesor esté familiarizado con las aplicaciones informáticas que posibilitan la búsqueda y extracción de los datos. Entre las opciones que suelen ofrecer las interfaces de consulta de los corpus, las más útiles son la búsqueda de concordancias (*key word in context*, KWIC), que permite observar el contexto de uso de una palabra o estructura, y las listas de la frecuencia con que aparecen los vocablos que conforman el corpus. Además, estas aplicaciones suelen contar con una serie de filtros que permiten circunscribir la consulta, como por ejemplo visualizar la colocación exacta de la palabra que se esté buscando con respecto a otra y la inclusión o exclusión de palabras parecidas (Granger, 2009; Cruz Piñol, 2012 y 2021; Granger, Gilquin y Meunier, 2015).

Es evidente que, para entender a fondo qué puede ofrecer un corpus y cómo se puede aprovechar a nivel didáctico, se precisa un buen dominio de las interfaces de consulta (Lenko-Szymanka, 2014; Vyatkina y Boulton, 2017), lo cual puede conllevar una significativa inversión de tiempo por parte del docente. Sin embargo, cuando se trabaja con corpus de aprendices el factor tiempo es relativo: en primer lugar, por la naturaleza y unicidad de su contenido, ya que rastrear muestras representativas de una determinada interlengua sin acudir a una herramienta informática sería mucho más complicado y requeriría sin duda una inversión de tiempo aún mayor y, en segundo lugar, porque, al ser los *learner corpora* de tamaño más reducido que los de una lengua natural, su consulta es más ágil. Adquirir la competencia técnica necesaria para utilizar con provecho las interfaces de consulta, por lo tanto, es una inversión de tiempo que merece la pena porque el empleo de los *learner corpora* en la enseñanza-aprendizaje de una lengua extranjera estimula la reflexión metalingüística contrastiva entre la L1 y la L2 a partir de la interlengua, que se convierte así en una herramienta de aprendizaje.

A nivel didáctico los *learner corpora*, al igual que otros corpus de lengua, se pueden utilizar o bien como bases de datos para el docente o bien como herramientas para la consulta directa por parte del aprendiente. En el primer caso, se pueden utilizar para extraer muestras a partir de las cuales elaborar materiales didácticos útiles para trabajar sobre aspectos lingüísticos que presenten alguna dificultad de aprendizaje o para llamar la atención sobre errores idiosincrásicos propios de cierto grupo de aprendientes (p. ej., italófonos que aprenden el español). En el segundo caso, los *learner corpora* se aprovechan como una herramienta de consulta a la que los aprendientes acuden para realizar tareas de reflexión sobre la lengua utilizando procedimientos inductivos de descubrimiento, análisis y síntesis, como sugiere el enfoque del *data-driven learning* (DDL) (Johns, 1991; Flowerdew, 2015; Tribble, 2015; Boulton, 2012 y 2021; Boulton y Vyatkina, 2021). Es así como el aprendiente se convierte en analista de fenómenos lingüísticos puntuales, profundizando en el conocimiento de la LE a través del análisis de los datos disponibles en el corpus y extrayendo sus propias conclusiones siguiendo un proceso que va de lo particular a lo general, es decir, del caso concreto a la norma que guía el uso de la lengua (Boulton, 2009).

De hecho, si por un lado es cierto que los corpus de referencia de una lengua concreta (p. ej., el CORPES XXI para el español) promueven en el aprendiente la observación de usos correctos de esa lengua en determinadas circunstancias, por otro no cabe duda de que los corpus de aprendices fomentan la reflexión contrastiva L1-LE/L2 por parte de un grupo concreto de aprendientes. Además, es importante tener en cuenta que, cuando estos corpus se explotan como recurso didáctico, los alumnos no están analizando sus propias producciones en la LE, sino datos representativos de la interlengua típica de aprendientes con su misma L1. Esto reduce de manera significativa el filtro afectivo y la sensación de

sentirse evaluados o juzgados, con todo lo que esto conlleva como amenaza a la imagen pública del alumno y a su autoestima.

2. La interlengua de los italianos en los corpus de aprendices de ELE

En los últimos años contamos con cada vez más corpus de aprendices de español/LE, como confirman los estudios más recientes (Alonso Ramos, 2016; Más Álvarez y Gil Martínez, 2018; Núñez Nogueroles, 2019; Rojo Sánchez y Palacio Martínez, 2022) y la base de datos *Learner Corpora Around the World* de la Universidad Católica de Lovaina[2]. Sin embargo, la mayoría de ellos recopilan la interlengua del español de aprendientes cuyas lenguas maternas son el inglés[3] u otras lenguas[4] y cuando incluyen producciones en ELE de aprendientes cuya L1 es el italiano, o bien la cantidad de textos es tan reducida como para no ser representativa, o bien estos no están disponibles en línea, como demuestra la breve reseña que se presenta a continuación.

El Corpus de Aprendices del Español (CAES, del Instituto Cervantes y la Universidad de Santiago de Compostela) está compuesto por 3878 pruebas escritas producidas por 1423 aprendices (niveles A1-C1) con distintas lenguas maternas. Los textos de aprendientes italianos son 87 (39 190 palabras) repartidos como sigue: 31 de A1, 26 de A2, 17 de B1, 12 de B2 y 1 de C1 y hay solo dos o tres textos por cada informante[5]. El Corpus de Aprendices de Español como Lengua Extranjera (CAELE) (Ferreira Cabrera *et al.* [2022])[6] cuenta con 1217 textos escritos por 201 aprendientes de ELE (niveles A2 y B1) con diferentes L1; sin embargo, los aprendientes italianos son 3 por un total de 22 textos. El Corpus Oral de Español como Lengua Extranjera (CORELE) (Campillos Llanos, 2012)[7] está compuesto por casi 14 horas de entrevistas a 40 hablantes de niveles A2 y B1 de 13 L1 diferentes, pero los aprendientes italianos son tan solo 4 por un total de 1 hora y 13 minutos de grabación. El Corpus Oral de Interlengua Español Italiano (CORINEI) (González Royo *et al.*, 2016) contiene 86 conversaciones en teletándem entre hablantes nativos y no nativos de español e italiano cada uno de los cuales está aprendiendo la lengua del otro. Es el único corpus de interlengua oral de aprendientes italianos de español LE existente y, además, sigue la misma metodología de recogida de datos que se ha utilizado con CORESPI, pero, lamentablemente, no está disponible en línea. El Corpus de Textos Escritos para el Análisis de Errores de Aprendices de E/LE (CORANE) (Cestero Mancera y Penadés Martínez, 2009) es uno de los primeros corpus de aprendices de ELE que se recopilaron y esto explica por qué solo está disponible en CD-ROM. Contiene 1091 composiciones escritas, procedentes de 321 estudiantes de cuatro niveles de competencia (elemental, intermedio, avanzado y superior) con 24 lenguas maternas. Sin embargo, los informantes italianos son tan solo 18 por un total de 29 textos. Por último, el Corpus ESPALEX (Bustos Gisbert y Sánchez Iglesias, 2011) contiene los textos de los exámenes escritos del DELE de nivel B2 que se realizaron en la Universidad de Salamanca. Pese a su interés por la extensión (12 576 textos) y la variedad de las lenguas maternas de los informantes (41), entre las cuales el italiano está muy bien representado con 1834 textos, este corpus no ha llegado a informatizarse.

La escasa presencia y disponibilidad de textos procedentes de aprendientes italianos de español LE justifica la recopilación e informatización del Corpus del Español de los Italianos (CORESPI) (Bailini, 2018a y 2018b; Bailini y Frigerio, 2018 y 2019) y su corpus gemelo CORITE (Corpus del Italiano de los Españoles). Ambos nacen con el objetivo de poner al alcance de la comunidad científica una herramienta

[2] Learner Corpora Around the World, https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html [consulta: 24/11/2022].

[3] Entre estos se señalan: CEDEL2, Corpus Escrito del Español como L2; SPLLOC, Spanish Learner Language Oral Corpus; SpinTX, Corpus vídeo del español hablado en Texas; SPT, Spanish Corpus Proficiency Level Training y LANGSNAP, Language and Social Networks Abroad Project.

[4] Entre estos cabe mencionar APRESCRILOV (Aprender a Escribir en Lovaina), que recoge muestras escritas de aprendientes de ELE (niveles de A1 a C1) cuya lengua materna es el holandés; CATE (Corpus de Aprendices Taiwaneses de Español), que contiene ensayos de estudiantes taiwaneses de ELE cuya L1 es el chino y el DIAZ Corpus, que contiene muestras de habla de estudiantes de ELE cuyas L1 son el alemán, el chino, el coreano, el islandés y el sueco.

[5] CAES, https://galvan.usc.es/caes/search [consulta: 31/05/2023].

[6] En el momento en que se escribe este trabajo (noviembre de 2022), el CAELE no está disponible en acceso libre.

[7] Se puede acceder a CORELE en http://www.lllf.uam.es/ESP/CORELE.html [consulta: 28/11/2022].

de consulta de acceso abierto para el análisis de la interlengua de aprendientes cuyas lenguas maternas son afines. Aunque se trata de corpus de tamaño reducido, son suficientemente representativos de ambas interlenguas, ya que contienen, respectivamente, 124 648 (CORESPI) y 103 147 (CORITE) palabras. En este trabajo nos centramos tan solo en CORESPI. Los datos se recopilaron utilizando el método e-tándem por correo electrónico, a través del cual 45 parejas de hablantes nativos y no nativos (HN/HNN) aprendientes italianos de español/LE y españoles de italiano/LE se intercambiaron sus producciones escritas por un período de 7 meses. Los informantes estaban aprendiendo sus respectivas lenguas extranjeras en Italia (Università Cattolica del Sacro Cuore, campus de Brescia) y España (Universidad de Salamanca y Escuela Oficial de Idiomas de Terrassa, Barcelona) y sus niveles de competencia iban del A1 al B2 del *MCER*. Para todos, el español o el italiano representaban la L3.

CORESPI y CORITE contienen producciones escritas tanto semiguiadas a partir de temas propuestos por el profesor, como libres, es decir, sobre temas que decidían los propios alumnos. Durante la tarea de escritura podían utilizar cualquier tipo de recurso (diccionarios, manuales, páginas web, etc.), no tenían límites de extensión ni de tiempo y sabían que sus trabajos no iban a ser evaluados. Estos criterios respondían a la necesidad de recopilar textos cuyo contexto de elaboración no supusiera ningún tipo de presión para el alumno. Como iban dirigidas a un interlocutor real (el compañero tándem), las producciones escritas que forman parte de estos corpus se pueden incluir dentro del macrogénero "carta informal", aunque se pueden identificar en ellas al menos cinco tipologías textuales (narración, descripción, argumentación, exposición e instrucción) y catorce áreas temáticas (información personal, familia, gustos y aficiones, actividades cotidianas, geografía, música, cocina, viajes, cine, recuerdos, vivienda, lecturas, fiestas y tradiciones, temas de actualidad). Estas características permiten extraer textos que contengan léxico o estructuras lingüísticas recurrentes relacionados con algún tema concreto: por ejemplo, si se quieren observar las dificultades de uso del imperfecto o su evolución en las respectivas interlenguas y en diferentes niveles de competencia, será más fácil encontrar muestras seleccionando textos que hablen de recuerdos.

La sistematización de los datos se realizó en tres etapas:

1. En un primer momento se reunieron, bajo el nombre *Spanish Corpus of Italian Learners* (SCIL), solo las producciones escritas de los aprendientes italianos de ELE por un total de 457 textos (124 186 palabras) (Bailini, 2013) y se hizo posible la consulta con la aplicación WordSmith Tools.
2. En un momento sucesivo se reunieron las producciones escritas de aproximadamente el 50 % de los informantes y se comprobó que los textos de los dos corpus gemelos (CORESPI y CORITE) fuesen efectivamente homogéneos por extensión, géneros textuales y temas tratados. Se reunieron en esta fase la primera parte de CORESPI (293 textos, 82 203 palabras) y de CORITE (257 textos, 70 307 palabras), correspondientes a los informantes de cuya interlengua era posible realizar un análisis longitudinal, es decir, aquellos que hubiesen escrito al menos 5 textos en un período de 7 meses[8]. En esta segunda etapa se hizo una experimentación piloto con la interfaz de consulta CATMA (*Computer Assisted Textual Markup and Analysis*) (Bailini, 2018a).
3. En la tercera etapa se añadieron los textos restantes hasta sumar 474 textos (124 648 palabras) en CORESPI y 385 (103 147 palabras) en CORITE, por un total de 45 parejas de informantes. Ambos corpus se cargaron en la plataforma CATMA y desde ese momento están disponibles en acceso abierto en la página web http://corespiycorite.altervista.org desde la cual solo hay que cumplimentar un formulario para solicitar el acceso a los corpus a través de CATMA (Bailini, 2018b; Bailini y Frigerio, 2018 y 2019).

3. CATMA, la interfaz de consulta de CORESPI

CATMA, acrónimo de *Computer Assisted Textual Markup and Analysis* (www.catma.de), es una interfaz de consulta de acceso libre desarrollada por la Universidad de Hamburgo. Es una herramienta flexible que permite consultas avanzadas utilizando simultáneamente criterios distintos; además, los resultados de las

[8] El 79 % de los informantes (34 de CORESPI y 30 de CORITE) produjo entre 7 y 16 textos, mientras que el 14 % y el 16 % respectivamente produjo entre 5 y 6 textos.

búsquedas se caracterizan por un alto grado de legibilidad y es posible compartir los corpus y su etiquetado entre varios usuarios. Los textos de CORESPI se han anotado por nivel de competencia en la lengua extranjera, género, edad, otros idiomas conocidos y palabras clave relacionadas con los temas tratados en cada texto. No están anotados desde el punto de vista morfológico, sintáctico, léxico y ortográfico.

A continuación, se ilustran algunas tipologías de consulta que pueden resultar útiles para el uso de CORESPI como base de datos para el docente. Por ejemplo, es posible visualizar la lista completa de las palabras que constituyen el corpus y la frecuencia con que aparece cada una de ellas. Las listas de palabras se pueden poner en orden alfabético o de frecuencia. En los corpus de aprendices esta función es útil porque permite visualizar también los ítems incorrectos, así que es posible identificar los errores más frecuentes en la ortografía. También es posible buscar ítems parecidos al que se está buscando o los que empiecen, contengan o terminen con una forma determinada: esta función permite, por ejemplo, analizar los errores de morfología verbal tan solo poniendo la raíz del verbo como criterio de búsqueda. Asimismo, se pueden realizar consultas complejas a través de la búsqueda de ítems que aparezcan distanciados uno de otro por una o más palabras: esta función permite extraer combinaciones de palabras, ver su frecuencia de aparición y detectar errores idiosincrásicos recurrentes.

CATMA, como cualquier interfaz para el análisis de corpus, permite la búsqueda de concordancias (*Key Word In Context*, KWIC), es decir, la visualización del contexto —parcial o total— en que aparece cierta palabra. También se pueden realizar consultas sencillas o avanzadas a partir de los criterios de etiquetado de los textos. Por ejemplo, se pueden visualizar todos los contextos en que aparece *hay* solo en las producciones de los informantes de un determinado nivel o se pueden combinar criterios diferentes (p. ej., *hay* en los informantes de nivel B1 en los textos que hablen de *vivienda*).



Document	Left Context	Keyword	Right Context	Start Point	End Point
CORESPI	se quiere, en efecto	hay	todo lo necesario para vivir	675849	675852
CORESPI	. Dentro en la casa	hay	tres habitaciones, la cocina	305873	305876
CORESPI	para vivir bien. No	hay	muchos electrodomésticos, sólo los	675891	675894
CORESPI	un poco de espacio,	hay	otro baño, una lavandería	305965	305968
CORESPI	sin demasiada dificultad, además	hay	dos televisiones, un ordenador	675999	676002
CORESPI	únicas "reglas" que	hay	que respetar son las siguientes	676074	676077
CORESPI	: en la planta baja	hay	la cocina, el comedor	675316	675319
CORESPI	que en la primera planta	hay	cuatro dormitorios, dos cuartos	675430	675433
CORESPI	mucho verde, hoy desafortunadam...	hay	muchísimas nuevas construcciones y p...	305528	305531
CORESPI	espectacular ;) Al final	hay	la sala donde estudio yo	229557	229560
CORESPI	; ademas de esto también	hay	una chimenea, una mesa	257070	257073
CORESPI	madre, de hecho no	hay	una parte que no me	675688	675691
CORESPI	de la naturaleza, porque	hay	muchas plantas, árboles y	305817	305820
CORESPI	y el desván. Fuera	hay	un pórtico muy grande con	675080	675083

Imagen 1 - CORESPI, consulta en CATMA con función KWIC: *hay* en textos de nivel B1 que hablan de *vivienda*.

Las consultas se pueden realizar utilizando tanto el lenguaje específico de CATMA para las búsquedas como, más sencillamente, la herramienta *query builder*, que guía paso a paso al usuario en la construcción de una consulta según varios niveles de complejidad. En la página web de presentación de CORESPI, también hay una guía de consulta de CATMA con varios ejemplos de búsqueda.

CATMA también tiene una función denominada *double tree*, que permite visualizar un árbol doble en cuyo centro aparece el ítem que se ha buscado.

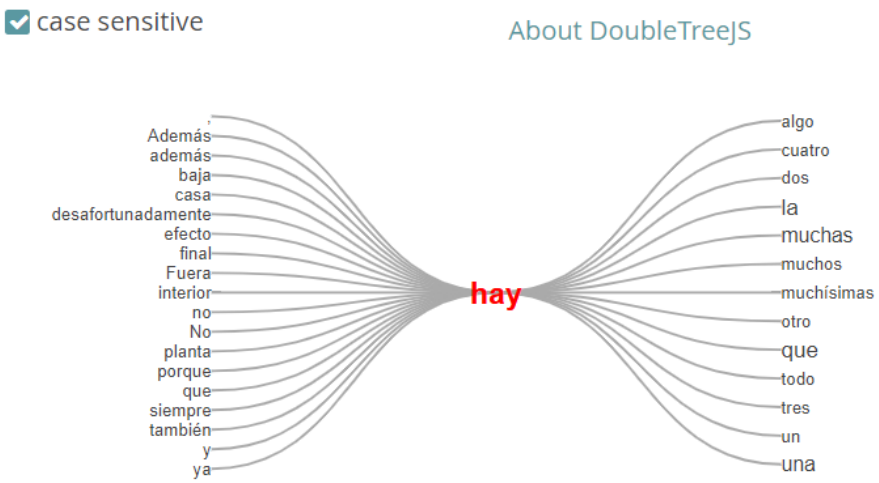

Imagen 2 - CORESPI, consulta en CATMA con la función *double tree*.

Cada una de las "hojas" del árbol está formada por las palabras que la siguen o preceden. Cuanto más grande es el tamaño de la palabra, mayor es el número de veces que aparece en esa posición. Por ejemplo, en la Imagen 2 se puede ver una parte del árbol que se ha generado buscando *hay* en CORESPI. Si se clica en una de las "hojas", aparecen las palabras que, en el corpus, siguen o preceden a la que está en el centro. Clicando en la palabra *muchas* (Imagen 3), aparecen, en rojo, todas las palabras que la siguen cuando está después de *hay*.

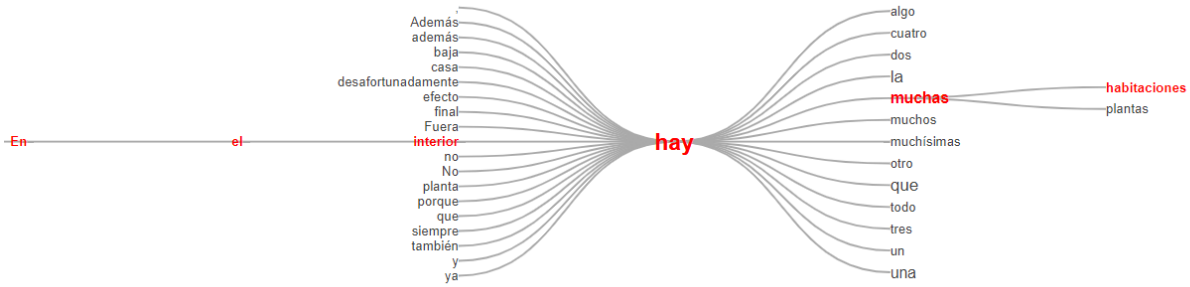

Imagen 3 - CORESPI, consulta en CATMA: secuencias de palabras con la función *double tree*.

4. Propuestas de explotación didáctica de CORESPI en el aula de español para italófonos

En este apartado se presentan tres ejemplos de explotación didáctica de CORESPI en el aula de español para italófonos. Las actividades que se proponen están destinadas a aprendientes de tres niveles de competencia (A2, B1 y B2) y se centran en tres dimensiones de la lengua: la ortografía, el léxico y la sintaxis.

En la primera de ellas, pensada para aprendientes de nivel A2, CORESPI se utiliza como fuente de recursos a la que el docente acude para extraer datos que luego utilizará para elaborar actividades de reflexión sobre dificultades concretas en la acentuación. La observación de la lista de palabras de todo el corpus permite identificar los errores de acentuación típicos de los italianos. Después, utilizando la función KWIC, se pueden extraer ejemplos de usos incorrectos y elaborar ejercicios en los que se les pide a los alumnos, que identifiquen los errores y los corrijan.

En las muestras de interlengua que se han seleccionado para esta actividad, además de errores de acentuación comunes a todos los aprendientes de ELE atribuibles a la no aplicación de las normas correspondientes, se observa que los italianos suelen olvidar poner tilde en aquellas palabras que en español presentan una ortografía o una pronunciación muy parecidas a su L1. Se trata de un error de

transfer negativo, atribuible al hecho de que en italiano solo se acentúan gráficamente las palabras agudas que terminan en vocal acentuada (*caffè, città, tabù*...): de ahí que este grupo de aprendientes simplemente prescinda de la norma ortográfica del español. Otro error típico de los italianos está relacionado con la simplificación de la norma de acentuación del español y consiste en poner tilde en la vocal tónica de todas las palabras sin distinguir entre agudas, llanas o esdrújulas. Por último, tampoco suelen poner tilde en las oraciones interrogativas directas e indirectas o, si la ponen, cometen errores de generalización. El objetivo de esta actividad, por lo tanto, es concienciar al alumno sobre la importancia de las normas de acentuación para una ortografía correcta. Cabe señalar que cuando se utiliza un *learner corpus* con fines didácticos, para evitar confundir a los alumnos, es buena norma corregir los errores presentes en las muestras y dejar solo aquellos sobre los cuales se desea llamar la atención. Este es el criterio que se ha seguido en la elaboración de las actividades 1 y 2.

Tabla 1 - Actividad sobre errores de acentuación. (Fuente: CORESPI)

Observa estas frases, identifica los errores de acentuación y trata de corregirlos.

1. Ahora te explico bién por que en febrero no puedo ir a Barcelona.
2. ¿Como fue tu exámen de italiano?
3. Estudio español y me gusta mucho, ademas veo muchas telenovelas colombianas.
4. Juntar los huevos con el azucar y mezclar hasta que parezca una crema.
5. Me alegro de que seas tía, desde tu mail he podido entender cuanto te haya hecho felíz.
6. En mi curso se estudia tambien economia y me gusta mucho.
7. Tengo un gato que se llama Matisse y es muy simpatico.
8. Lo que se queda mas vividamente en mi mente son las imagenes de mis abuelos.
9. Esta semana tengo muchas cosa qué hacer, nos escribimos la proxima semana con mas calma cuándo haya terminado mis examenes.
10. ¿Vienes a Italia? ¿de verdad? ¿cuando? ¡Ahora no! ¡Hace frio! Si quieres mí opinión, para mí Nápoles es estupenda...

Ahora reflexiona: ¿A qué se deben estos errores? _____
Apunta aquí algunas estrategias para evitarlos _____

La segunda actividad propone una reflexión contrastiva sobre el uso del marcador discursivo *finalmente* en español e italiano y está pensada para aprendientes de nivel B1. Los alumnos trabajan en parejas o en grupos pequeños con un listado de usos de *finalmente* que el profesor ha extraído de CORESPI. Este marcador suele plantear dificultades para este grupo de aprendientes debido a su falsa afinidad con el italiano *finalmente*, cuyo equivalente español es *por fin*. Se les pide a los alumnos que observen las muestras de interlengua recopiladas (Tabla 2) y que indiquen cuáles son correctas y cuáles no. Además, se les ofrece un listado de usos correctos de *finalmente* sacados del CORPES XXI para que, comparando su interlengua con ejemplos auténticos de uso en español, traten de entender en qué contextos se usa correctamente este marcador. Después de esta primera fase de observación y análisis entre pares, se hace una puesta en común para comprobar que hayan identificado las oraciones incorrectas, se les estimula a pensar con qué marcadores se podrían corregir los fragmentos incorrectos y se van guiando las correcciones explicando las diferencias y los matices de *finalmente*, *por fin* y *al final* en español y en italiano.

Tabla 2 - Actividad de reflexión contrastiva sobre *finalmente*. (Fuentes: CORESPI y CORPES XXI)

*Observa cómo algunos compañeros italianos aprendientes de español utilizan el marcador "*finalmente*". Identifica los usos correctos e incorrectos y trata de explicar a qué se deben sus errores.*

1. Durante estas últimas semanas **finalmente** ha llegado la primavera.
2. Hoy hace un poco de calor, **finalmente**...
3. Después con la ayuda de algunas personas mi padre logró sacar el coche del foso y **finalmente** fuimos al hospital a saludar a mi hermanito y mi madre.
4. La casa había pertenecido a una rica familia, los Fleischmann, que durante muchos años no tuvo hijos y **finalmente** tuvo uno llamado Jacob, que murió ahogado en extrañas circunstancias.
5. Quitado el quiste, volví a mi habitación, donde me esperaban mi madre y Lorenzo... **finalmente** podía llorar y desahogar toda la tensión acumulada! :)
6. Todavía no sé lo que voy a hacer durante la Semana Santa. Inicialmente mi novio y yo pensábamos ir a Nápoles, pero creo que **finalmente** estaremos aquí.
7. Mr Knightley piensa que él la corresponde: Emma es celosa de su amiga y **finalmente** comprende que ella está enamorada de su cuñado.
8. Durante estas vacaciones de Pascua no he hecho nada especial. Quería ir a Liguria, donde vive la familia de mi madre, pero, **finalmente**, he decidido quedarme en mi ciudad porque tenía que hacer muchas cosas.

Si puede resultarte útil como término de comparación, aquí tienes algunos ejemplos de uso correcto de "finalmente" en español, procedentes del CORPES XXI.

1. En un recipiente mediano, cubre la base con las láminas de patata. Seguidamente, coloca una capa de berenjena, continúa con una capa de la mezcla de soja, verdura y tomate. Después otra de berenjena, otra de mezcla de soja, y **finalmente** la bechamel. Puedes acabar con queso rallado por encima y orégano. Hornea a 170ºC durante 40 minutos.
2. 22 de enero de 2020: la OMS se reúne de urgencia en Ginebra para analizar el riesgo del Coronavirus. Se debate entre declarar el estado de emergencia internacional y **finalmente** deciden esperar.
3. Después subíamos para visitar Herculano, Pompeya, Positano y Nápoles, llegábamos a Roma para después subir a Florencia, Venecia y **finalmente** Milán, desde donde retornaríamos a la Argentina.
4. Primero habló Simon, luego yo y **finalmente** Josep, que, en definitiva, era el auténtico protagonista porque era quien había tomado una decisión que cambiaría su vida.
5. Dos días antes, el Barcelona había remontado milagrosamente al Paris Saint-Germain y al acabar la pachanga le pregunté a Daniel si había podido ver el partido. Como me dijo que no, le ofrecí ver un vídeo de los goles que me había mandado un colega por WhatsApp. Daniel dudó, pero **finalmente** cogió el teléfono y se puso a ver los goles.

La tercera propuesta, pensada para aprendientes de nivel B2, se enmarca en el enfoque pedagógico del *data-driven learning* y consiste en una actividad de análisis contrastivo entre el español y el italiano sobre las subordinadas temporales con *cuando*. Los alumnos, trabajando en parejas en el aula, consultan directamente CORESPI. Se les guía en la elaboración de la consulta en CATMA y se les pide que busquen la palabra *cuando* en las producciones de nivel B2. Después, utilizando la función *double tree*, se los invita a observar los modos y tiempos verbales que suelen acompañar *cuando*. Clicando en los verbos, aparece, en rojo, el fragmento de texto. Los alumnos observan las oraciones e indican cuáles son correctas y cuáles no, tratando de explicar qué significado expresan de acuerdo con el modo verbal que se emplee. Si necesitan entender mejor el contexto, también pueden utilizar la función *Key Word In Context* y visualizar

el texto completo en que aparece *cuando* (Imagen 4). Por último, en la puesta en común, se los invita a comentar las diferencias entre el español y el italiano en las subordinadas temporales con *cuando*.

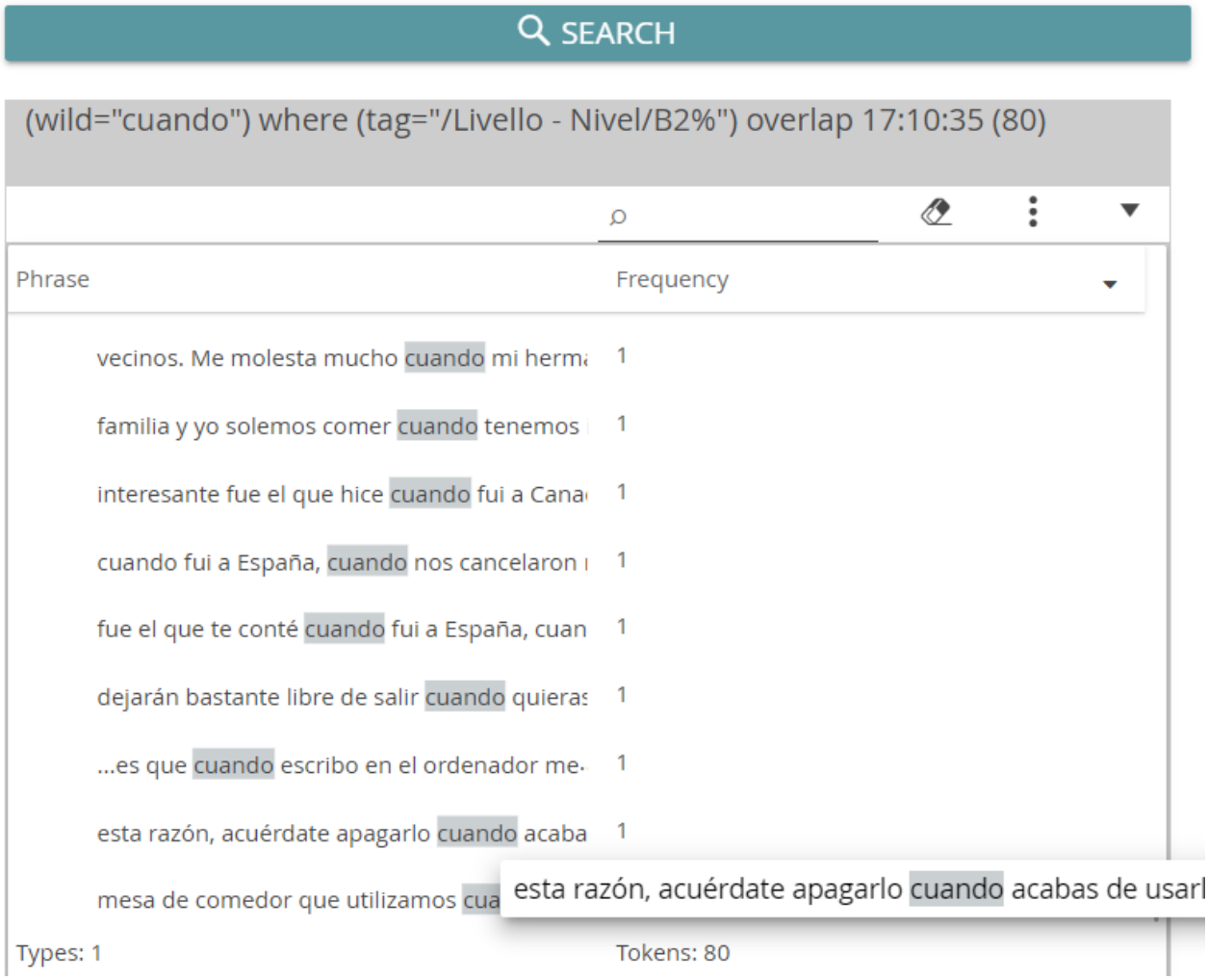

Imagen 4 - Ejemplos de visualización en contexto, actividad 3.

Conclusiones

En este trabajo se ha analizado el potencial de los *learner corpora* para la enseñanza-aprendizaje de una lengua extranjera presentando algunas propuestas de explotación didáctica en el aula de español para italianos a partir de CORESPI. Los corpus de aprendices permiten realizar actividades que estimulan la concienciación lingüística utilizando la interlengua como herramienta de aprendizaje. Su uso en el aula de LE posibilita la elaboración de tareas que, llamando la atención sobre la forma de la interlengua, por un lado, contribuyen a que el aprendiente tome conciencia de errores idiosincrásicos atribuibles a la interferencia de su L1 y, por otro, le estimulan a ampliar su conocimiento de las divergencias y convergencias entre su L1 y la LE. Los *learner corpora* son, por lo tanto, una herramienta más al alcance de aprendientes y docentes y su uso en el aula de LE no solo promueve un aprendizaje consciente, sino que contribuye al desarrollo de competencias instrumentales, como por ejemplo el dominio de las interfaces de consulta, y de la autonomía en la gestión del propio proceso de aprendizaje.

Referencias bibliográficas

Abad Castelló, M. (2019). Uso de corpus lingüísticos por y para profesores de español lengua extranjera. *redELE revista electrónica de didáctica del español lengua extranjera*, 31, pp. 48-167.

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

Bailini, S. (2013). SCIL: a Spanish Corpus of Italian Learners. En C. Vargas Sierra (Ed.), *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 (CILC 2013)*, Elsevier, 95, pp. 542-549. http://www.sciencedirect.com/science/journal/18770428/95.

Bailini, S. (2018a). CORESPI y CORITE: criterios de construcción, características y potencialidades. En S. Chiapello, T. Martín Sánchez, C. González Royo y N. Puigdeval Balafuy (Eds.), *Telecolaboración y corpus para el estudio de la lengua y la cultura*. Alicante: Publicaciones de la Universidad de Alicante, pp. 115-128.

Bailini, S. (2018b). *CORESPI y CORITE*. [Corpus en línea]. https://corespiycorite.altervista.org.

Bailini, S. y Frigerio, A. (2018). CORESPI e CORITE, due nuovi strumenti per l'analisi dell'interlingua di lingue affini. *CHIMERA: Romance Corpora and Linguistic Studies*, 5 (2), pp. 313-319. https://revistas.uam.es/chimera.

Bailini, S. y Frigerio, A. (2019). CORESPI y CORITE: criterios de construcción e implementación en línea de dos corpus de interlengua paralelos. *e-AESLA Revista digital*, 5, pp. 303-312. https://cvc.cervantes.es/lengua/eaesla/eaesla_05.htm.

Ballestero de Celis, C. y Sampedro Mella, M. (Eds.) (2021). *Aportes del CAES a la enseñanza del español como lengua extranjera. Estudios aplicados de gramática y pragmática*. Santiago de Compostela: Universidad de Santiago de Compostela.

Boulton, A. (2009). Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35 (1), 81-106. Recuperado de: https://hal.archives-ouvertes.fr/hal-00326990v2.

Boulton, A. (2012). Data-Driven Learning: The Perpetual Enigma. En S. Gozdz-Roszkowski (Ed.), *Explorations across Languages and Corpora. PALC 2009*. Berna: Peter Lang, pp. 563-580.

Boulton, A. (2017). Research timeline: Corpora in language teaching and learning. *Language Teaching*, 50 (4), pp. 483-506.

Boulton, A. (2021). Research in data-driven learning. En P. Pérez-Paredes y G. Mark (Eds.), *Beyond the concordance: Corpora in language education*. Ámsterdam: John Benjamins, pp. 9-34.

Boulton, A. y Pérez-Paredes, P. (2014). Researching new uses of corpora for language teaching and learning. *ReCALL*, 26 (2), pp. 121-127. https://doi.org/10.1017/S0958344014000068.

Boulton, A. y Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions. *Language Learning & Technology*, 25 (3), pp. 66-89.

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

Campillos Llanos, L. (2012). Designing a search interface for a Spanish learner spoken corpus: the end-user's evaluation. En N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk y S. Piperidis (Eds.), *Proceedings of the 8th Language Resources and Evaluation Conference*. Estambul: European Language Resources Association. https://aclanthology.org/volumes/L12-1/.

Cestero Mancera, A. M. (2009). *Corpus de textos escritos para el análisis de errores de aprendices de E/LE (CORANE)*. [Volumen y CD-ROM]. Alcalá de Henares: Servicio de publicaciones de la Universidad de Alcalá.

Cruz Piñol, M. (2012). *Lingüística de corpus y enseñanza del español como segunda lengua*. Madrid: Arco Libros.

Cruz Piñol, M. (2021). *e-Research y español LE/L2 Investigar en la era digital*. Londres/Nueva York: Routledge.

Ferreira, A., Elejalde Gómez, J. A. y Blanco San Martín, L. P. (2022). Diseño e implementación del Corpus de Aprendientes de Español como Lengua Extranjera (CAELE). *Círculo de Lingüística Aplicada a la Comunicación*, 90, pp. 137-152.

Flowerdew, L. (2015). Data-driven learning and language learning theories. Whither the twain shall meet. En A. Leńko-Szymańska y A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Ámsterdam: John Benjamins, pp. 15-36.

González Royo, C., Chiapello, S., Martín Sánchez, T., Mura, G. A., Pascual Escagedo, C., Puigdevall Bafaluy, N. y Regagliolo, A. (2016). CORINÉI (Corpus Oral de Interlengua Bilingüe Español-Italiano): Elaboración y aplicación a la enseñanza/aprendizaje de la interacción con la TICs. En J. D. Álvarez Teruel, S. Grau Company y M. T. Tortos Ybáñez (Eds.), *Innovaciones metodológicas en docencia universitaria: resultados de investigación*. Alicante: Universidad de Alicante, pp. 1951-1958.

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

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

Granger, S. (2012). How to Use Foreign and Second Language Learner Corpora. En A. Mackey y S. Gass (Eds.). *Research Methods in Second Language Acquisition: A Practical Guide*. Chichester: Wiley-Blackwell, pp. 5-29.

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

Johns, T. (1991). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. En T. Johns y P. King (Eds.), *Classroom Concordancing. English Language Research Journal*, 4, pp. 27-45.

Lenko-Szymanska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26, pp. 260-278. https://doi.org/10.1017/S095834401400010X.

Martín Sánchez, M. T., Pascual Escagedo, C. y Rodríguez, M. P. (2017). Creación de material didáctico para el nivel A2 de ELE, a partir de conversaciones procedentes del Corpus Corinéi (Corpus oral de interlengua español/italiano). En R. Roig-Vila, (Ed.), *Investigación en docencia universitaria. Diseñando el futuro a partir de la innovación educativa*. Barcelona: Octaedro, pp. 969-979.

Mas Álvarez, I. y Gil Martínez, A. (2018). Los corpus de aprendices: un terreno en expansión para la enseñanza del español. En M. Ellison, P. Nicolás y S. Valente Rodrigues (Eds.), *As línguas estrangeiras no ensino superior: propostas didáticas e casos em estudo* Oporto: Universidade do Porto, pp. 35-55.

MCER = Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: MECD y Anaya. http://cvc.cervantes.es/Ensenanza/Biblioteca_Ele/marco/cvc_mer.pdf.

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

Núñez Nogueroles, E. (2019). Pasado, presente y futuro de los corpus de aprendices de ELE. Una revisión bibliográfica. *ReiDoCrea*, Monográfico 2019, pp. 170-190. https://doi.org/10.30827/Digibug.58473.

Palacios Martínez, I. M., Barcala Rodríguez, F. M. y Rojo Sánchez, G. (2019). El "Corpus de Aprendices del Español" (CAES) y sus aplicaciones para la enseñanza/aprendizaje del español como lengua extranjera. En M. Blanco Domínguez, H. Olbertz y V. Vázquez Rozas (Eds.), *Corpus y construcciones: perspectivas hispánicas*. Santiago de Compostela: Servicio de Publicaciones Universidad Santiago de Compostela, pp. 273-302.

Rojo Sánchez, G. (2021). *Introducción a la lingüística de corpus en español*. Londres/Nueva York: Routledge.

Rojo Sánchez, G. y Palacios Martínez, I. M. (2022). Los corpus de aprendientes de español como L2. En G. Parodi, P. Cantos Gómez y C. Howe (Eds.), *Lingüística de corpus en español*. Londres/Nueva York: Routledge, pp. 74-88.

Sánchez Rufat, A. (2015). La investigación de corpus de aprendientes y el desarrollo de los estudios de la interlengua del español. *Language Design*, 17, pp. 57-84.

Sinclair, J. M. (2005). How to build a corpus. En M. Wynne, (Ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 79-83.

Tribble, C. (2015). Teaching and language corpora: Perspectives from a pedagogical journey. En A. Leńko-Szymańska y A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, pp. 37-62.

Vyatkina, N. y Boulton, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, 21 (3), pp. 1-8. http://llt.msu.edu/issues/october2017/commentary.pdf.