

Corpus paralelos y enseñanza del español LE/L2 a sinohablantes. Estado de la cuestión y propuestas didácticas*

WEIQIAN WEI
Universitat de Barcelona

MAR CRUZ PIÑOL
Universitat de Barcelona

KRIS BUYSE
KU Leuven / Universidad Nebrija

1. Presentación

En este capítulo pretendemos llamar la atención sobre las aplicaciones de un tipo especial de corpus multilingües, los corpus paralelos, en el contexto específico de la enseñanza del Español como Lengua Extranjera o Segunda (en adelante, ELE) a estudiantes sinohablantes. Para ello, en el apartado 2 revisamos el estado de la cuestión en relación con los corpus paralelos y sus usos prácticos en el marco del ELE. En el apartado 3 ofrecemos una selección argumentada de corpus paralelos chino-español disponibles en línea y los analizamos a partir de una serie de criterios. Finalmente, en el apartado 4 proponemos dos actividades y otros posibles usos de los corpus paralelos para enseñar ELE a los estudiantes sinohablantes.¹

2. Los corpus paralelos y sus aplicaciones didácticas

2.1. ¿Qué es un corpus paralelo?

Los corpus que recogen muestras en más de una lengua pueden ser bilingües (si incluyen textos en dos lenguas) o multilingües (si contienen

* Este trabajo tiene su origen y se basa en Wei (2020).

1. En el momento de redacción de estas páginas, todavía no se han llevado a la práctica las propuestas didácticas que aquí se presentan. En el estudio experimental que está desarrollando una de las autoras (Weiqian Wei) en el marco de su tesis doctoral, se pondrán en práctica estas y otras propuestas.

más lenguas). En ambos casos, los textos que conforman el corpus se tienen que seleccionar según idénticos criterios para todas las lenguas representadas (Hallebeek 1999). Si además el corpus contiene la misma colección de textos en dos o más idiomas (uno de ellos es el idioma original y los otros, la traducción), estaremos hablando de corpus paralelos (McEnery y Hardie 2012). Los corpus paralelos pueden clasificarse a partir de dos criterios: según los idiomas incluidos y según la dirección de traducción (McEnery y Xiao 2007). Los corpus paralelos juegan un papel destacable especialmente para las investigaciones en traducción, los estudios contrastivos, la lingüística computacional y la enseñanza de lenguas extranjeras. En comparación con los corpus monolingües, los corpus paralelos aportan ventajas como mostrar la correspondencia y el contraste entre dos lenguas (Barlow 2000; Kenning 2010). Según los idiomas que incluya, un corpus paralelo puede ser bilingüe (los que constan de textos en dos idiomas) o multilingüe (los que constan de textos en más idiomas). Según la dirección de traducción, se pueden clasificar en unidireccionales, si la dirección de la traducción es unívoca (por ejemplo, textos españoles traducidos al chino); bidireccionales, cuando la traducción se efectúa en los dos sentidos (por ejemplo, textos españoles traducidos al chino y viceversa), o multidireccionales, cuando un mismo original es traducido a varios idiomas.

En este contexto conviene llamar la atención hacia un tipo de corpus muy especial que se ha llegado a denominar «corpuscionario». Puesto que las fronteras entre el corpus y el diccionario vienen desapareciendo en los últimos tiempos, el neologismo «corpuscionario» (acuñado por Alonso Ramos 2009) se refiere tanto a los diccionarios enriquecidos con corpus (generalmente de acceso libre y con información textual) como a los corpus complementados con un diccionario. Según Alonso Ramos (2009: 1195), un ejemplo de estos tipos híbridos podría ser la «reutilización de un corpus paralelo aunque complementado con otros repertorios lexicográficos». Así, las plataformas *Linguee*,² *Glosbe* y *Re-*

2. Un diccionario en línea que recopila las traducciones realizadas por humanos, alineadas con su texto de origen, y tomadas de sitios web multilingües (Alonso Jiménez 2013). Cabe mencionar que, a pesar de que en *Linguee*, un «corpuscionario»

verso muestran dicha interacción entre corpus y diccionario, es decir, son diccionarios bi-/multilingües de libre acceso en línea que están basados en (o complementados con) corpus paralelos alineados.

Con el objetivo de conocer mejor los corpus paralelos y compararlos entre sí, a continuación, vamos a definir tres conceptos clave en la Lingüística de corpus (en adelante, LC): las concordancias paralelas, los listados de frecuencias y la lematización. En primer lugar, a diferencia de las concordancias *KWIC*,³ las concordancias paralelas⁴ (Figura 1) se refieren a las listas de ocurrencias de una palabra o una combinación de palabras junto con sus contextos y alineadas con sus traducciones en una o más lenguas de forma paralela (Frankenberg-Garcia 2005). Las concordancias que ofrecen los corpus son una herramienta de suma importancia y muestran múltiples utilidades que pueden ser de gran ayuda para todas las personas interesadas en la didáctica de lenguas (Alonso Pérez-Ávila 2006).

The screenshot shows the CPEIC interface. On the left, there are filters for 'romano', 'Español', and 'Inglés'. The main area displays search results for 'ejemplo' in Spanish. It shows 7 articles, 3 orations, and 7 examples. Two examples are highlighted: 'Resultado #1 Oración #145' and 'Resultado #2 Oración #786'. Each example shows the word in Spanish, its English translation, and its Chinese translation, along with the context in each language.

Figura 1. Concordancias paralelas en CPEIC
(Corpus Paralelo de Español, Inglés y Chino).⁵

que tiene mucha popularidad, se incluye el chino, solo se puede combinar con el inglés para formar par de idiomas. Es decir, en nuestro caso, no nos proporciona las concordancias paralelas chino-español.

3. Sobre las concordancias *KWIC*, es decir, *Key Words in Context* ('palabras clave en contexto'), véase Pérez Hernández (2002).

4. También denominadas concordancias bilingües.

5. *CPEIC* es un corpus paralelo trilingüe elaborado en la Universidad Nacional de Cheng Kung (NCKU) desde el año 2007 con el propósito de facilitar el análisis contrastivo entre las tres lenguas.

En segundo lugar, disponer de la información sobre las palabras más frecuentes de una lengua puede constituir una gran ventaja a la hora de elaborar materiales para la enseñanza del ELE (Cruz Piñol 2012). Asimismo, las listas de frecuencias pueden ayudar a los alumnos a elegir las estructuras lingüísticas más habituales en el uso real (Contreras Izquierdo y Martos Eliche 2020). En el caso específico que nos ocupa, si los diccionarios tienen en cuenta la frecuencia (Gargallo Gil 1992; Jiménez Ríos 2001; *DLE* 2020), el estudiante no nativo podrá identificar las traducciones más frecuentes para el contexto en cuestión.

En tercer lugar, en un corpus es fundamental la lematización, que consiste en vincular cada forma con su lema. La forma se refiere a la palabra tal como aparece en su contexto de uso, mientras que el lema es la entrada que, por convención, engloba todas las formas flexionadas de esa palabra. Si tomamos como ejemplo el paradigma verbal, la forma se refiere a cada una de las posibilidades de flexión (*hablo, hablaste, hablen...*) y el lema será el verbo en infinitivo (*hablar*). La ausencia de lematización muestra varias limitaciones al usar los corpus, especialmente cuando se trata de lenguas con mucha flexión (como, en nuestro caso, el español), porque habría que ejecutar la búsqueda de todas las formas flexionadas posibles para poder cubrir todas las ocurrencias de un lema.

2.2. Diseño y elaboración de los corpus paralelos

Conocer cómo se construye un corpus paralelo nos sirve de gran ayuda para entender mejor sus características, su naturaleza y su utilidad. Según Doval (2017), en el proceso de elaboración de un corpus paralelo, destacan tres fases principales: la compilación de textos y su preprocesamiento, la alineación de los textos y la anotación documental y lingüística. Asimismo, durante el proceso de la compilación de textos, de acuerdo con Doval (2017), hay que tener en cuenta la importancia de la cantidad y la calidad de los textos. Además, cabe subrayar la importancia del equilibrio de los datos en cuanto a la dirección de traducción.

Por lo que respecta al proceso de alineación, es una tarea crucial e imprescindible en la construcción y elaboración de un corpus paralelo (Castillo Rodríguez 2011). En palabras de Castillo Rodríguez (2009: 7), la alineación es «la reestructuración de los textos de tal forma que se pueda establecer una correspondencia entre los párrafos, oraciones y/o palabras de los textos involucrados». La alineación consta de dos tareas, una previa de segmentación de los textos y la posterior, la vinculación de esos segmentos a partir de la correspondencia entre las dos lenguas para formar bisegmentos, es decir, pares alineados de segmentos (Doval 2017).

En cuanto a los niveles de la segmentación, las unidades de alineación pueden ser párrafos, frases o palabras. Entre ellos, lo más habitual actualmente es la alineación a nivel oracional. Con respecto a la anotación, siguiendo a McEnery y Hardie (2012), un corpus contiene básicamente los siguientes tres tipos de información añadida para la investigación de los datos en el corpus: los metadatos, el marcado textual y la anotación lingüística (Kübler y Zinsmeister 2015). Más concretamente, los metadatos se refieren a la información externa relacionada con la obra original, la traducción y otros datos bibliográficos. Los metadatos nos permiten, posteriormente, recuperar información relevante del corpus. El marcado textual, en cambio, proporciona la división interna de los textos tales como capítulos o partes a fin de localizar una cadena de texto dentro de una obra. Por último, la anotación lingüística está relacionada con el análisis propiamente lingüístico de las palabras que conforman el corpus (lema, categoría gramatical o función sintáctica, principalmente).

2.3. La web como corpus paralelo

En la mayoría de las ocasiones, los corpus paralelos tienen un tamaño relativamente limitado y es habitual que las concordancias paralelas mostradas sean insuficientes para extraer conclusiones fiables. Ante este problema, una solución es utilizar también la web como corpus.

En este sentido, cabe adelantar que, de los cuatro⁶ corpus que hemos seleccionado para nuestra propuesta, tres se compilan a partir de los datos de la web: *Glosbe*, *Eshelper* y *ReversoContext*. Los corpus de este tipo son fáciles de manejar y contienen gran cantidad de datos bi-/multilingües, pero también presentan algunos problemas a la hora de utilizarlos en el aula. Uno de estos consiste en «la baja calidad de los contextos encontrados» (Gelbukh *et al.* 2002: s/p), como veremos en el apartado 3.4. Por ejemplo, los textos pueden ser producidos por autores poco cultos, la alineación automática sin revisión puede causar incidencias, los textos pueden estar mal segmentados, etc.

2.4. Aplicaciones de los corpus paralelos en la enseñanza de lenguas

Los corpus paralelos se están empleando cada vez más en el ámbito de la enseñanza de lenguas, un terreno en el que este tipo de corpus presenta varias ventajas. Como sostiene Barlow (2000), utilizar los textos paralelos permite a los aprendices de lenguas investigar directamente la correspondencia entre las palabras o estructuras en dos lenguas. Por su parte, Doval (2018) ilustra muy bien las posibilidades que presenta el corpus paralelo creado por la misma autora⁷ en el aula de ELE: dicho corpus pretende ser un recurso útil para los estudiantes avanzados de español o de alemán, para que dispongan de diversos contextos de uso y ejemplos de traducción.

De acuerdo con Kenning (2010), los corpus paralelos permiten a los estudiantes acceder a la evidencia contrastiva que destaca las similitudes y divergencias entre las diferentes lenguas. Además, el acceso a la traducción, tanto la completa como la parcial, posibilita que los estudiantes estén expuestos a las ocurrencias que contienen vocabulario o frases que no conocen. De ahí que se favorezca el *Aprendizaje Basado*

6. Datos correspondientes a abril de 2021. Hay que decir que previamente contemplamos también el corpus *TAUS Data Cloud*, pero aquí lo hemos descartado porque en 2021 dejó de estar accesible de forma gratuita.

7. Se trata del corpus *PaGeS*, un corpus paralelo español-alemán, disponible en <https://www.corpuspages.eu/corpus/about/about?lang=en>.

en Datos (*Data-Driven Learning*)⁸ (Kenning 2010). A continuación, se verá que los estudios distinguen cuatro tipos de usos de los corpus paralelos en la enseñanza de lenguas:

a) Uso directo de las concordancias paralelas en el aula (p. ej. Barlow 2000; Olímpio de Oliveira Silva y Penadés 2016; Lu y Cheng 2016; Doval 2018):⁹ el uso directo de los corpus paralelos en el aula ha sido investigado por muchos autores, especialmente en lo concerniente a los corpus que incluyen el inglés y desde la década de 1990 (Zanettin 1994; Barlow 2000; St. John 2001, etc.). Utilizando los datos bilingües inglés-francés, Barlow (2000) sugiere que se puede sensibilizar a los estudiantes sobre la manera de expresar significados equivalentes, gracias a los abundantes contextos que proporcionan. Señala que, con la ayuda de la consulta en paralelo de los pronombres reflexivos, los alumnos pueden observar que las formas reflexivas en inglés no se traducen necesariamente por una forma reflexiva en francés, e incluso para algunas formas no existen equivalentes similares en otra lengua. Por su parte, St. John (2001) plantea la utilidad de un corpus paralelo inglés-alemán con alumnos de nivel inicial. Las tareas propuestas en este caso consisten en la comparación entre la lengua meta y la lengua de origen para extraer conclusiones de esta observación: a partir de la consulta de la palabra *also* en alemán, que puede ser una partícula o un adverbio dependiendo de la sintaxis y del contexto, y que es un *falso amigo* entre estas dos lenguas, los alumnos pueden observar la función, la posición y el uso como modificador de esta palabra e intentar derivar independientemente las reglas que rigen su uso. Otro ejemplo lo encontramos en Olímpio de Oliveira Silva y Penadés (2016), quienes presentan cómo utilizar los «corpuscionarios» de tipo *Linguee*, en la enseñanza-aprendizaje de las unidades fraseológicas en el aula de ELE. Afirman que el uso de las concordancias paralelas

8. Sobre *Data-Driven Learning* (DDL), puede verse Johns (2000).

9. Véase, también, Zanettin (1994); St. John (2001); Wang (2001); Frankenberg-García (2005); Cao *et. al* (2018).

portugués-inglés en el aula facilitará la observación de diferentes equivalencias en distintos contextos, la comprensión de los conceptos culturales, la reflexión sobre los errores interlingüísticos, etc. Concretamente sobre las aplicaciones de los corpus paralelos chino-español en la enseñanza del ELE a estudiantes sinohablantes, Lu y Cheng (2016: 133) apuntan que la utilización del corpus CPEIC (Figura 1) podrá «facilitar la manipulación de la lengua meta (español), a través de las experiencias y conocimientos de su primera lengua extranjera (inglés) y su lengua materna (chino)». En resumen, todos los estudios mencionados en este apartado nos han inspirado para valorar la utilidad de los corpus paralelos chino-español en el aula.

- b) Uso indirecto de los corpus paralelos para la elaboración de materiales didácticos (Chang 2004; Römer 2008; Doval 2018): aunque en la propuesta de nuestro trabajo no nos vamos a centrar en el uso indirecto de los corpus paralelos, conviene describirlo para tener una visión más global sobre las aplicaciones didácticas de los corpus. En su uso indirecto, los corpus constituyen una base para el diseño y la preparación de materiales didácticos, manuales de referencia o pruebas de nivel (Doval 2018). Cabe mencionar que este tipo de uso se considera muy ventajoso en el ámbito de la enseñanza de lenguas con fines específicos. Como indica Römer (2008), un corpus paralelo, sobre todo si contiene textos de especialidad, es adecuado para servir de base para la elaboración de materiales didácticos y manuales de referencia como diccionarios bilingües, puesto que en este tipo de corpus se muestra la equivalencia y correspondencia.
- c) Uso como herramienta de consulta autónoma para mejorar la expresión escrita (p. ej. Cruz Piñol 2012; Bluemel 2014; Buyse 2020):¹⁰ Frankenberg-Garcia (2004) señala que la consulta tipo *self-access* de los corpus paralelos puede ser de gran ayuda para los

10. Véase, también, Frankenberg-Garcia (2004); Buyse y Verlinde (2013); Buyse (2014); Cruz Piñol (2015); Buyse (2020).

alumnos durante las tareas de escritura en la lengua meta, puesto que, por un lado, los corpus paralelos tipo *Linguee* ofrecen tanto el significado de las palabras desconocidas para los alumnos como los contextos donde se usan las palabras adecuadamente, y, por otro, les permite observar en ocasiones las unidades léxicas que no tienen equivalente directo en otra lengua. Debido a que la consulta en este caso la inician los alumnos mismos según su propia necesidad, este tipo de utilización favorecerá significativamente su aprendizaje de la lengua meta (siempre y cuando los alumnos sepan extraer las conclusiones relevantes de la consulta, para lo cual es recomendable que el profesor forme a los alumnos en el uso experto de estas herramientas).

d) Uso como base de la investigación en lingüística contrastiva (p. ej. Lu y Cheng 2016; Lu y Lu 2016; Lu *et al.* 2017):¹¹ por lo que respecta al tema de este trabajo, en los últimos años, el estudio contrastivo entre el chino y el español a partir de corpus ha suscitado el interés de algunos profesores y estudiantes de español en China. Por ejemplo, en las tres publicaciones mencionadas se destaca la utilidad del corpus *CPEIC* para facilitar varios estudios contrastivos sobre los pronombres de complemento directo e indirecto, los verbos copulativos, los tiempos y las voces gramaticales, etc.

En el contexto específico que nos ocupa, de estos cuatro usos creemos que los que más pueden favorecer el aprendizaje del ELE son el (a) y el (c), es decir, el uso directo (llevar las concordancias paralelas al aula, con la intervención del docente) y el uso como herramienta de consulta (que los alumnos consulten los corpus de forma autónoma, sobre todo en las tareas de escritura). Por eso, en el próximo apartado vamos a mostrar los corpus paralelos chino-español existentes en línea y, tras analizarlos de acuerdo con unos criterios, seleccionaremos los cuatro que nos parecen más adecuados para el uso directo y para el uso como herramienta de consulta.

11. Véase, también, Luo (2017), McEnery y Xiao (2007).

3. Corpus paralelos de nativos chino-español útiles para la enseñanza del ELE¹²

3.1. Cuatro de diez: justificación de la selección

Tras consultar algunas selecciones previas¹³ y utilizar motores de búsqueda,¹⁴ hemos localizado diez corpus paralelos de nativos chino-español, que se pueden clasificar en los siguientes tres grupos:

- Grupo A: Los repositorios de traducciones humanas que presentan el texto original y la traducción de forma alineada, y los «corpuscionarios» en línea de tipo *Linguee*. Generalmente los textos proceden de sitios web bilingües o multilingües y han pasado por un procedimiento altamente automatizado (*MyMemory*, *Glosbe*, *Eshelper*, *ReversoContext*, *OPUS*).¹⁵
- Grupo B: Los corpus paralelos que solo permiten a los usuarios descargar los textos alineados, pero no cuentan con un sistema de consulta. Los corpus de este grupo no muestran la hibridación entre diccionario y corpus, sino que sirven como base de datos para los corpus de grupo A (*The Holy Bible* [Resnik *et al.* 1999], *Corpus Paralelo de las Naciones Unidas* [Ziemski *et al.* 2016]).
- Grupo C: Los corpus paralelos establecidos por un grupo de investigación y diseñados para una finalidad puntual de investigación lingüística (*PaChEs* [Doval *et al.* 2019], *CPEIC* [Lu 2016], *RST Spanish-Chinese Treebank* [Cao *et al.* 2018]).

Sin embargo, para la propuesta didáctica que mostraremos enseguida, no vamos a utilizar todos los corpus mencionados arriba (once en total) porque, por una parte, el uso de algunos corpus encierra mayor

12. Cabe aclarar que en nuestro trabajo nos centramos en los corpus (paralelos) de nativos y no nos ocupamos de los corpus (paralelos) de aprendices (McEnery y Xiao 2011).

13. Por ejemplo, la página personal de Chelo Vargas Sierra (<https://personal.ua.es/es/chelo-vargas/enlaces/corpus-paralelos-en-linea.html>), y la página de *Tagpacker* de Mar Cruz Piñol (<https://tagpacker.com/user/mar.cruz.pinol?t=corpus,paralelos%25Fy%25Fmultilingues>).

14. Por ejemplo, *Google*, *Baidu*, *Bing*, etc.

15. También *Taus Data Cloud* era de este tipo, pero lo descartamos porque ya no se puede consultar en abierto (véase la nota 6).

grado de dificultad y, por otra parte, los corpus como *PaChEs* todavía están en construcción o no están en línea hasta la fecha. Por eso, en nuestro trabajo hemos decidido utilizar los cuatro corpus paralelos chino-español fácilmente manejables en línea, es decir, *Glosbe*, *Eshelper*, *ReversoContext* y *CPEIC*. Partiremos de las propuestas de Buyse (2011, 2021) como referencia para comparar las posibilidades que estos corpus presentan en el aula de ELE. Además, compararemos los corpus teniendo en cuenta las características y conceptos clave en el proceso de la construcción de los corpus, como se ha visto en el apartado 2. Es decir, analizaremos los corpus a partir de las siguientes preguntas:

- 1) Idiomas: ¿cuántos idiomas contiene este corpus? En cuanto al par de idiomas chino-español, ¿qué variedades incluye?
- 2) Propósito: ¿con qué objetivo se ha construido este corpus?
- 3) Tamaño: ¿cuál es el tamaño de este corpus?
- 4) Selección de textos: ¿qué textos se han seleccionado?
- 5) Lematización: ¿es un corpus lematizado?
- 6) Palabra clave resaltada en ambas versiones de lengua: ¿la palabra que buscamos está resaltada tanto en el texto original como en la traducción?
- 7) POS-tagging (*Part of Speech*, es decir, categorías gramaticales): ¿hay etiquetación gramatical?
- 8) Definición: ¿se muestra la definición de las palabras que buscamos?
- 9) Combinación de palabras: ¿se pueden hacer búsquedas de combinaciones de palabras?
- 10) Búsqueda filtrada: ¿se pueden filtrar los resultados según el tema, el tipo de textos, etc.?
- 11) Alineación: ¿cuál es el nivel de alineación de este corpus?
- 12) Información anotada: ¿hay información de los metadatos y el marcado textual?
- 13) Dirección de traducción: ¿es un corpus unidireccional, bidireccional o multidireccional?
- 14) Particularidad: ¿tiene algunas características destacables?

3.2. Presentación breve de los cuatro corpus seleccionados

Glosbe es un «corpuscionario» multilingüe en línea creado a partir de memorias de traducción, provenientes de traducciones hechas por humanos. La interfaz permite ver no solo la traducción de la palabra, sino también cómo se comporta en la frase. Actualmente cuenta con 375.000 frases traducidas alineadas en el diccionario chino-español. Comparado con *TAUS Data Cloud*, *Glosbe* nos proporciona más información sintáctico-semántica, por ejemplo, la etiquetación de la categoría gramatical (POS), la definición, etc.

Eshelper es un «corpuscionario» diseñado especialmente para los aprendices sinohablantes de español. Recopila 450.000 bisegmentos alineados a partir de los datos de la web. Además, permite acceder al diccionario bilingüe de términos jurídicos, económicos y de negocio disponible en *Eshelper* (*XiBanYaYuFaLuWaiMaoJingMaoCiHui*). Puesto que es un recurso diseñado especialmente para los aprendices sinohablantes, las frases recopiladas parecen más cotidianas y fáciles de comprender.

ReversoContext es un «corpuscionario» que utiliza millones de textos bilingües, procesados por complejos algoritmos elaborados con inteligencia artificial. Incluye más de 100 pares de idiomas y miles de millones de traducciones extraídas de diálogos de películas, documentos oficiales, páginas web, periódicos. En comparación con los dos «corpusccionarios» mencionados arriba, una particularidad de *ReversoContext* consiste en que se permite ver el contexto más amplio de una traducción y verificar la fuente de un ejemplo.

El *Corpus Paralelo de Español, Inglés y Chino* (*CPEIC*) fue creado por un grupo de investigación de la National Cheng Kung University con el objetivo de proporcionar datos fiables para facilitar el análisis contrastivo de tres lenguas: español, inglés y chino, que son respectivamente la lengua meta (L3), la primera lengua extranjera (L2) y la lengua materna (L1) de los estudiantes sinohablantes. Se ha recogido una gran cantidad de materiales idiomáticos (p. ej. cuentos, textos de la Biblia, documentos de las Naciones Unidas, subtítulos de películas etc.), y se continúa mejorando la construcción del corpus. A diferen-

cia de los corpus mencionados más arriba, los textos en *CPEIC* son seleccionados por expertos.

3.3. *Tabla comparativa de los cuatro corpus*

Para tener una visión general, a continuación (Tabla 1), mostramos la comparación de estos cuatro corpus siguiendo los criterios mencionados en el apartado 3.1.¹⁶ Se puede observar que cada corpus tiene sus puntos a favor y en contra.¹⁷

Criterios	Corpus			
	<i>Glosbe</i>	<i>ReversoContext</i>	<i>Eshelper</i>	<i>CPEIC</i>
Multilingüe	+	+	-	+
Finalidad ¹⁷	+	+	+	+
Selección y revisión de los textos	-	-	-	+
Lematizado	+	-	+	-
Palabra clave destacada en dos versiones de lengua	-	+	-	+
Etiquetado gramatical (POS-tagging)	+	+	+	+
Etiquetado semántico (definiciones)	+	+	+	-
Combinaciones de palabras	+	+	+	-
Búsqueda filtrada por tema, tipo de texto, o categoría gramatical	-/-/-	-/-/+	-/-/+	+/+/+
Alineado por frase	+	+	+	+
Metadatos y marcado textual	Fuente	Fuente	Título	Título
Traducción multidireccional	+	+	+	+
Otras funciones	+	+	+	-

Tabla 1. Comparación de los corpus *Glosbe*, *ReversoContext*, *Eshelper* y *CPEIC*.

16. En la tabla vamos a excluir el criterio del tamaño porque, por un lado, la información sobre el tamaño del corpus chino-español no está registrada en la página de *ReversoContext* y, por otro lado, en los otros tres corpus los editores utilizan diferentes parámetros para describir el tamaño y somos conscientes de que estos parámetros no son comparables.

17. Esta tabla se ha diseñado desde la perspectiva de la utilidad de los corpus de aprendices para la enseñanza-aprendizaje de lenguas.

3.4. Ventajas y desventajas del uso de los corpus en el aula de ELE

Tras presentar las características de los cuatro corpus seleccionados, es preciso llamar la atención sobre algunas de sus limitaciones, dado que pueden ser relevantes a la hora de usarlos en el aula de ELE. En primer lugar, conviene tener en cuenta que, excepto uno (*CPEIC*), los corpus que analizamos (*Glosbe*, *ReversoContext* y *Eshelper*) toman la web como fuente; es decir, los textos incluidos en estos corpus no han sido seleccionados ni revisados manualmente por expertos. En este sentido, no hay que olvidar que, si bien los corpus basados en los datos extraídos de la web tienen mayor tamaño, la calidad de los textos puede ser peor (Gelbukh *et al.* 2002).

En segundo lugar, se observa que entre los cuatro corpus seleccionados solo *Glosbe* y *Eshelper* están lematizados, y el nivel de lematización de *Glosbe* es muy limitado. Sin duda alguna, la ausencia de lematización limitará su utilidad en el aprendizaje-enseñanza del ELE (Cruz Piñol 2012). Por ejemplo, en el caso del aprendizaje de una colocación que incluya un verbo, un corpus que permita la búsqueda por lemas posibilitará que los alumnos dispongan de más ejemplos para observar la combinación más habitual y extraer sus propias conclusiones.

En tercer lugar, la información de anotación (la anotación lingüística, los metadatos, el marcado textual) se encuentra muy limitada en estos cuatro corpus. Por una parte, la anotación lingüística limitada restringe el nivel de análisis al que se pueda someter un corpus. Por otra parte, la ausencia de los metadatos sobre la obra, el autor, la fecha, etc. impide la recuperación de información relevante por parte de los usuarios cuando tienen dudas.

En cuarto lugar, la posibilidad de comparar entre géneros, temas, países, etc. es baja o inexistente en la mayoría de estos corpus. Esto se debe, por una parte, a que los textos que se recogen en estos corpus muestran poca diversidad en cuanto al género, medio, tema, área geográfica, etc. Por otra parte, entre estos cuatro corpus, solo hay uno (*CPEIC*) que nos permite seleccionar el tema que deseamos buscar; es decir, la mayoría de los corpus no disponen de un sistema para realizar una búsqueda filtrada.¹⁸

18. Esta constatación nos lleva a poner de relieve que la interacción con los

Teniendo en cuenta todo lo expuesto, consideramos que el que más se acerca al «corpus paralelo ideal» para utilizar en el aula de español sería el de *Eshelper*, por ser una herramienta especialmente diseñada para los alumnos sinohablantes de español y tener una calidad de textos relativamente buena. Sin embargo, cabe destacar que cada corpus tiene sus pros y contras. Así que es recomendable utilizar estos corpus de forma combinada, con el objetivo de tener más ejemplos variados. Además, consideramos que la utilización de los corpus basados en los datos de la web (*Glosbe*, *ReversoContext* y *Eshelper*) por parte de los alumnos de nivel inicial requiere la instrucción o supervisión del docente, ya que las frases en estos corpus no están siempre bien redactadas y traducidas. Por supuesto, los cuatro corpus pueden ser de ayuda para los alumnos de nivel avanzado para hacer consultas autónomas (de tipo *self-access*) durante las tareas de escritura, puesto que los alumnos de este nivel ya son capaces de distinguir los pares de textos que no tienen correspondencia y escoger los ejemplos que les parecen útiles. Asimismo, el corpus que cuenta con un sistema para filtrar los resultados de búsqueda (*CPEIC*) puede ser una buena herramienta a la hora de comparar un fenómeno lingüístico entre diferentes tipos de textos o áreas temáticas.

4. Propuesta didáctica

4.1. Actividades para aprender locuciones

Una de las propuestas que lanzamos en este trabajo va orientada al aprendizaje de las locuciones con los corpus paralelos. Este tipo de unidades fraseológicas constituyen un claro obstáculo para los aprendices sinohablantes de español debido a su idiomática, fijación,

corpus debe avanzar hacia interfaces intuitivas adaptadas a los hábitos actuales de navegación en la red, lo cual incluiría opciones de búsqueda más visuales, como mapas interactivos (Barros 2021, Cruz Piñol 2021, Trigo 2021), en la línea de lo que se ha venido a denominar «humanidades espaciales» (Gregory y Geddes 2014; Sartor 2019). En este sentido, en los cuatro corpus seleccionados se echan en falta representaciones gráficas que aporten una dimensión espacial a la hora de visualizar los datos lingüísticos.

sentido cultural, etc. (Wu 2016). En este contexto, a nuestro juicio, los corpus paralelos chino-español pueden ser de gran utilidad para elaborar un amplio repertorio de ejercicios destinados al aprendizaje de locuciones por parte de los estudiantes sinohablantes. Se pueden utilizar para, por un lado, ayudar a los aprendices a inducir los significados idiomáticos y, por otro lado, para comprobar si el alumno ha captado sus significados y usos. Con la ayuda de los corpus paralelos, los estudiantes pueden aprender las locuciones a nivel textual, comunicativo y pragmático. A continuación, mostramos una propuesta de actividad¹⁹ (Figura 2).

Observa estos ejemplos y las traducciones, y busca en las oraciones en chino los equivalentes de las locuciones destacadas en español. Asocia, después, las locuciones con sus equivalentes en chino.

Se me abría la boca por haber dormido poco. 因为睡眠时间少，我老打哈欠。
Cuando ve los pasteles, se le hace la boca agua. 他看到蛋糕的时候，就馋的流口水。
La noticia corrió de boca en boca y ya lo saben todo. 那消息已口口相传，现在大家都已知此事
Te fuiste de la boca y ahora lo sabe todo. 是你说漏了嘴，现在他全知道了。

(Fuente: Esbelper y elaboración propia)

Irse de boca/ irse la boca a uno	说漏嘴
Abrirse la boca	打呵欠
Hacerse la boca agua a uno	口口相传
Andar/correr/ir de boca en boca	流口水

Figura 2. Actividad para aprender locuciones (a partir del corpus *Esbelper*).

19. Se pueden ver más propuestas en Wei (2020).

En esta actividad (duración: 10 minutos aproximadamente), los alumnos están expuestos a las concordancias paralelas extraídas del corpus *Eshelper*, que les ayudarán a establecer la correspondencia entre las dos lenguas. Esta correspondencia facilitará la comprensión y memorización de las locuciones para los alumnos sinohablantes, puesto que la mayoría de estas locuciones comparten el elemento *boca* (口/嘴) y las otras tampoco son difíciles de comprender con la ayuda del contexto y la traducción. Después, la actividad para asociar las locuciones con sus equivalentes obliga a los alumnos a reparar estas locuciones, lo cual favorecerá la retención de estas unidades léxicas.

4.2. Actividades para aprender marcadores del discurso

Otro ejemplo del uso de los corpus paralelos chino-español que proponemos en este trabajo sería para el desarrollo de la competencia discursiva. Más concretamente, vamos a centrarnos en el conector concesivo *aunque* (Figura 3). Desde nuestro punto de vista, los corpus paralelos podrían ser de gran utilidad en la enseñanza de los marcadores discursivos, puesto que en un diccionario convencional no se puede encontrar una «definición» de un marcador y una manera más eficaz para saber qué significa un marcador es verlo en su uso y contexto. Además, gracias a la observación de las concordancias paralelas, los aprendices pueden comparar las similitudes y diferencias del uso de los marcadores en dos lenguas, extraer su propia conclusión de modo inductivo y reflexionar sobre sus errores interlingüísticos causados por las interferencias de su lengua materna.

ACTIVIDAD 1

Aunque vive conmigo, no tenemos ninguna vinculación.

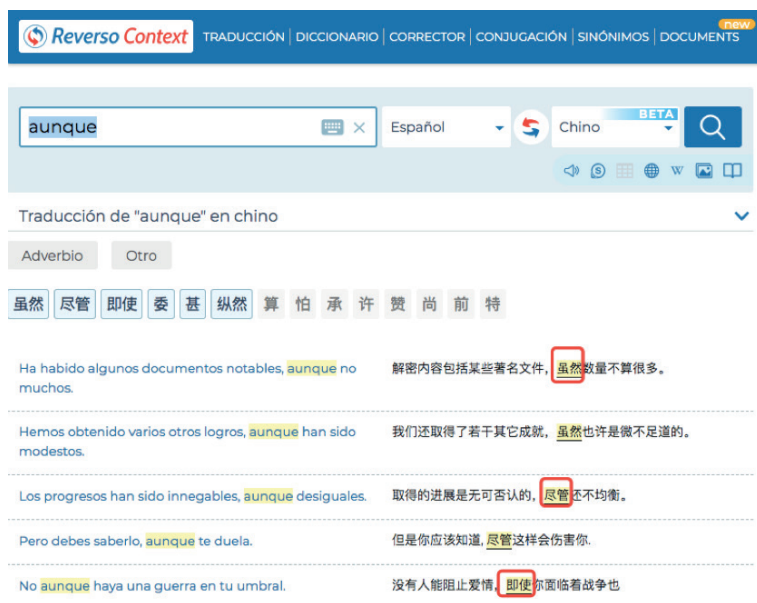
虽然跟我住在一起，**但是**我们没有任何交集。

(Fuente: *Eshelper* y elaboración propia)

1. En este ejemplo arriba, has visto que para expresar **虽然** en español, se usa el conector ***aunque***. Busca en Glosbe, ReversoContext, Eshelper y CPEIC el conector ***aunque*** y observa los resultados. ¿Has encontrado otra traducción de este conector además de **虽然**? y, ¿has observado alguna similitud o diferencia del uso de estos conectores en chino y en español?

Figura 3. Actividad para aprender el conector ***aunque*** (a partir del corpus *Eshelper*).

En esta actividad, empezamos con un ejemplo extraído de un corpus paralelo (en este caso, el corpus *Eshelper*). En el ejemplo, el conector ha sido traducido por **虽然**. Sin embargo, el conector ***aunque*** puede corresponder a muchos marcadores en chino en diferentes contextos (*p. ej.* 虽然, 即使, 尽管, 纵使, etc.). Después de este ejemplo, se pide que los alumnos hagan búsquedas en los cuatro corpus. A modo de ejemplo, si un alumno busca la palabra ***aunque*** en el corpus *ReversoContext*, va a obtener los resultados que se muestran en la Figura 4:



The screenshot shows the Reverso Context website interface. At the top, there is a navigation bar with the logo and links for 'TRADUCCIÓN', 'DICCIONARIO', 'CORRECTOR', 'CONJUGACIÓN', 'SINÓNIMOS', and 'DOCUMENTS'. Below this is a search bar with 'aunque' entered, and language dropdowns for 'Español' and 'Chino'. The search results are titled 'Traducción de "aunque" en chino' and show a list of Chinese equivalents: 虽然, 尽管, 即使, 委, 甚, 纵然, 算, 怕, 承, 许, 赞, 尚, 前, 特. Below this, there are five example sentences in Spanish with their Chinese translations. The Chinese words corresponding to 'aunque' in each example are highlighted with red boxes: 虽然, 虽然, 尽管, 尽管, 即使.

Spanish Example	Chinese Translation
Ha habido algunos documentos notables, aunque no muchos.	解密内容包括某些著名文件, 虽然数量不算很多。
Hemos obtenido varios otros logros, aunque han sido modestos.	我们还取得了若干其它成就, 虽然也许是微不足道的。
Los progresos han sido innegables, aunque desiguales.	取得的进展是无可否认的, 尽管还不均衡。
Pero debes saberlo, aunque te duela.	但是你应该知道, 尽管这样会伤害你。
No aunque haya una guerra en tu umbral.	没有人能阻止爱情, 即使你面临着战争也。

Figura 4. Búsqueda de la palabra *aunque* en *ReversoContext*.

Igual que en el estudio de Barlow (2000), tras buscar una determinada palabra y observar las concordancias paralelas repetidamente, los alumnos pueden llegar a ser conscientes de que una palabra puede corresponder a varias unidades léxicas en distintos contextos. Por ejemplo, el conector *aunque* puede corresponder a muchos marcadores en chino en diferentes contextos (p. ej. 虽然, 即使, 尽管, 纵使, etc.). Además, esta actividad, cuya duración es de unos 20 minutos, también tiene el objetivo de que los alumnos descubran las similitudes y diferencias entre el chino y el español en cuanto al uso de estos marcadores. Por ejemplo, a través del ejemplo que se ha mostrado más arriba (Figura 4: *Aunque* vive conmigo, no tenemos ninguna vinculación. 虽然跟我住在一起, 但是我们没有任何交集。), se ha visto que, por una parte, el fragmento español, igual que el chino, ha sido dividido en dos unidades y que las primeras unidades en

dos lenguas empiezan con un marcador del discurso que muestra una relación concesiva. Por otra parte, el uso de los marcadores discursivos ha mostrado cierto contraste en dos lenguas. De acuerdo con Cao *et al.* (2016), para expresar el mismo significado, es necesario incluir dos marcadores del discurso en este enunciado en chino: 虽然 (*aunque*) al principio de la primera unidad y 但是 (*pero*) al principio de la segunda unidad. No obstante, en español solo se usa un marcador (*aunque*) al principio de la primera unidad. Tras comparar repetidamente las concordancias paralelas en dos lenguas, esperamos que los alumnos investiguen este tipo de contraste ellos mismos y consigan evitar este tipo de error interlingüístico causado por la interferencia de su lengua materna en su producción. Así, en esta actividad, los alumnos se han convertido en investigadores, lo cual aumentará su autonomía de aprendizaje.

4.3. Otras posibilidades

En las dos propuestas presentadas más arriba, hemos explorado la utilización de los corpus paralelos en la enseñanza del ELE para el desarrollo de la competencia léxica y discursiva. Se ha comprobado que, en comparación con los corpus monolingües, los corpus paralelos tienen la ventaja de mostrar la correspondencia y el contraste entre dos lenguas, como se ha visto en el apartado 2.4. Por lo tanto, con la finalidad de utilizar esta herramienta especialmente para los alumnos sinohablantes de español, sería recomendable prestar especial atención a los análisis contrastivos entre el chino y el español, tomándolos como punto de partida a la hora de diseñar las actividades. Aparte de los dos ejemplos que hemos propuesto, los corpus paralelos también pueden aportar muchas posibilidades de la enseñanza del ELE a nivel sintáctico. Así, para aprender las estructuras de la voz pasiva en español, los alumnos podrían buscar la palabra «被» (uno de los marcadores de la voz pasiva en chino) con el objetivo de obtener ejemplos en voz pasiva en español y familiarizarse con las estructuras.

Sin duda, son variados los usos de los corpus paralelos en la enseñanza de lenguas. Y es que, gracias a la manera como presentan los datos auténticos, los alumnos tienen la oportunidad de explorar, aprender e investigar ellos mismos con mayor autonomía.

5. Recapitulación

En este capítulo se ha abordado la aplicación de los corpus paralelos chino-español en el contexto específico de la enseñanza del ELE para los estudiantes sinohablantes. Se ha visto que los estudios distinguen cuatro tipos de usos de los corpus paralelos en la lingüística aplicada: el uso directo de las concordancias paralelas en el aula, el uso indirecto de los corpus paralelos para la elaboración de materiales didácticos, el uso como herramienta de consulta para mejorar la expresión escrita y el uso como base para la investigación en lingüística contrastiva. A continuación, hemos presentado unos criterios para escoger los corpus más adecuados para utilizar en el aula, que nos han permitido seleccionar cuatro corpus, que recomendamos utilizar de forma combinada, teniendo en cuenta el nivel de los alumnos y el tipo de instrucción del docente. Por último, hemos planteado dos propuestas de actividades utilizando los corpus seleccionados para aprender locuciones y conectores, y hemos dejado abiertas otras posibilidades para el desarrollo de diferentes competencias. Esperamos que esta aportación pueda servir de ayuda tanto para los investigadores como para los profesores interesados en utilizar los corpus paralelos para la enseñanza del ELE a aprendices sinohablantes.

Referencias bibliográficas

- ALONSO JIMÉNEZ, Elisa (2013): «*Linguee* y las nuevas formas de traducir», *Skopos: revista internacional de traducción e interpretación*, vol. 2, pp. 5-28. En línea, <https://rio.upo.es/xmlui/bitstream/handle/10433/851/2013_Alonso_Linguee_Skopos.pdf?sequence=1> [Consulta: 11/11/2020].
- ALONSO PÉREZ-ÁVILA, Elena (2006): «El corpus lingüístico en la didáctica del léxico en el aula de E/LE», memoria de máster publicada en *RedELE*, vol. 11, pp.

- 1-83. En línea, <<https://www.educacionyfp.gob.es/dam/jcr:c9948929-a65a-4afc-8df2-9c104e1e6175/2010-bv-11-01alonso-pdf.pdf>> [Consulta: 11/11/2020].
- ALONSO RAMOS, Margarita (2009): «Hacia un nuevo recurso léxico: ¿fusión entre corpus y diccionario?», *A Survey of Corpus-based Research. Panorama de investigaciones basadas en corpus*, Murcia, AELINCO, pp. 1191-1207. En línea, <<http://www.dicesp.com/app/webroot/files/file/CILC%2009.pdf>> [Consulta: 11/11/2020].
- BARLOW, Michael (2000): «Parallel Texts in Language Teaching», en Simon Botley et al. (eds.), *Multilingual Corpora in Teaching and Research*, Amsterdam, Rodopi, pp. 106-115. En línea, <https://books.google.es/books?id=eBwAj1nyRu4C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false> [Consulta: 11/11/2020].
- BARROS, Benamí (2021): «Representar visualmente los resultados de la investigación sobre el español LE/L2», en Mar Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 228-247.
- BLUEMEL, Brody (2014): «Learning in Parallel: Using Parallel Corpora to Enhance Written Language Acquisition at the Beginning Level», *Dimension*, vol. 31, pp. 31-48. En línea, <<https://files.eric.ed.gov/fulltext/EJ1080263.pdf>> [Consulta: 11/11/2020].
- BUYSE, Kris (2011): «¿Qué corpus en línea utilizar para qué fines en la clase de ele?», en Javier de Santiago Guervós, Hanne Bongaerts, Jorge J. Sánchez Iglesias, Marta Seseña Gómez (eds.), *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE*, Salamanca, ASELE, pp. 277-288. En línea, <https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/21/21_0277.pdf> [Consulta: 11/11/2020].
- BUYSE, Kris (2014): «Una hoja de ruta para integrar las TIC en el desarrollo de la expresión escrita recursos y resultados», *Journal of Spanish Language Teaching*, vol. 1, n. 1, pp. 101-115.
- BUYSE, Kris (2020): «Qué tipo de corpus para qué tipo de texto: de la teoría a la práctica», en Javier de Santiago Guervós, Lourdes Díaz (eds.), *Lingüística textual y enseñanza de español LE/L2*, London/New York, Routledge, pp. 282-306.
- BUYSE, Kris (2021): «Corpus textuales de nativos para investigar sobre la enseñanza/aprendizaje del español LE/L2», en Mar Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 109-136.
- BUYSE, Kris; VERLINDE, Serge (2013): «Possible Effects of Free on Line Data Driven Lexicographic Instruments on Foreign Language Learning: The Case of *Linguee* and the *Interactive Language Toolbox*», *Procedia – Social and Behavioral Sciences*, vol. 95, pp. 507-512. En línea, <<https://doi.org/10.1016/j.sbspro.2013.10.675>> [Consulta: 11/11/2020].
- CAO, Shuyuan; DA CUNHA, Iria; IRUSKIETA, Mikel (2016): «A Corpus-based Approach for Spanish-Chinese Language Learning», *Proceedings of the 3rd Wor-*

- kshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, Osaka, The COLING 2016 Organizing Committee, pp. 97-106. En línea, <<https://www.aclweb.org/anthology/W16-4913/>> [Consulta: 11/11/2020].
- CAO, Shuyuan; DA CUNHA, Iria; IRUSKIETA, Mikel (2018): «The RST Spanish-Chinese Treebank», *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, Santa Fe, Association for Computational Linguistics, pp. 156-166. En línea, <<https://www.aclweb.org/anthology/W18-4917.pdf>> [Consulta: 11/11/2020].
- CASTILLO RODRÍGUEZ, Cristina (2009): «La elaboración de un corpus *ad hoc* paralelo bilingüe», *Tradumática*, n. 7. En línea, <http://www.lexytrad.es/assets/castillo2009a_0.pdf> [Consulta: 11/11/2021].
- CASTILLO RODRÍGUEZ, Cristina (2011): «La alineación de un corpus paralelo multilingüe: propuesta de fases para la didáctica de traducción especializada inversa», *Cadernos de Tradução*, vol. 1, n. 27, pp. 117-142. En línea, <<https://periodicos.ufsc.br/index.php/traducao/article/view/2175-7968.2011v1n27p117>> [Consulta: 11/11/2021].
- CHANG, Baobao (2004): «Chinese-English Parallel Corpus Construction and its Application», *Proceedings of the 18th Pacific Asia Conference on Language, Information and Computation*, Tokyo, Waseda University, Logico-Linguistic Society of Japan, pp. 283-290. En línea, <<http://aclweb.org/anthology/Y04-1030>> [Consulta: 11/11/2020].
- CONTRERAS IZQUIERDO, Narciso; MARTOS ELICHE, Fermín (2020): «Las secuencias formulaicas en la enseñanza del español como LE/L2. La frecuencia léxica como criterio de nivelación», *Porta Linguarum*, n. 33, pp. 111-127. En línea, <http://www.ugr.es/~portalin/articulos/PL_numero33/7_Narciso%20Contreras.pdf> [Consulta: 11/11/2021].
- CRUZ PIÑOL, Mar (2012): *Lingüística de corpus y enseñanza del español como 2/L*, Madrid, Arco/Libros [2.ª edición: 2017].
- CRUZ PIÑOL, Mar (2015): «Léxico y ELE: enseñanza/aprendizaje con tecnologías», *Journal of Spanish Language Teaching*, vol. 2, n. 2, pp. 165-179.
- CRUZ PIÑOL, Mar (2021): «e-Research: fundamentos metodológicos y aplicaciones a la investigación sobre el español LE/L2», en M. Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 1-39.
- DLE = REAL ACADEMIA ESPAÑOLA; ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA (2020): «Advertencias», *Diccionario de la Lengua Española*, versión electrónica 23.4. En línea, <<https://dle.rae.es/contenido/advertencias>> [Consulta: 11/11/2020].
- DOVAL, Irene (2017): «La construcción de un corpus paralelo bilingüe multifuncional», *Moenia*, vol. 27, pp. 125-141. En línea, <<https://revistas.usc.es/index.php/moenia/article/view/4071>> [Consulta: 11/11/2021].

- DOVAL, Irene (2018): «Corpus paralelos en la enseñanza de lenguas extranjeras: un ejemplo de aplicación basado en el corpus PaGeS», *CLINA: Revista Interdisciplinaria de Traducción, Interpretación y Comunicación Intercultural*, vol. 4, n. 2, pp. 65-82. En línea, <<https://doi.org/10.14201/clina2018426582>> [Consulta: 11/11/2021].
- DOVAL, Irene et al. (2019): *Protocolo para constituir el corpus chino-español del Proyecto PaChEs*, Santiago de Compostela / Barcelona [trabajo inédito].
- FRANKENBERG-GARCIA, Ana (2004): «Lost in Parallel Concordances», en Aston Guy et al. (eds.), *Corpora and language learners*, Amsterdam, John Benjamins, pp. 213-229. En línea, <https://www.researchgate.net/publication/253553681_Lost_in_parallel_concordances> [Consulta: 11/11/2020].
- FRANKENBERG-GARCIA, Ana (2005): «Pedagogical Uses of Monolingual and Parallel Concordances», *ELT Journal*, vol. 59, n. 3, pp. 189-198.
- GARGALO GIL, José Enrique (1992): «La ordenación de acepciones en algunos artículos lexicográficos del DRAE: cuestiones de norma y uso», en Manuel Ariza Viguera (ed.), *Actas del II Congreso Internacional de la Historia de la Lengua Española*, Madrid, Pabellón de España, t. I, pp. 1043-1054. En línea, <<http://data.cervantesvirtual.com/manifestation/230923>> [Consulta: 11/11/2020].
- GELBUKH, Alexander et al. (2002): «Corpus virtual, virtual: Un diccionario grande de contextos de palabras españolas compilado a través de Internet», en Julio Gonzalo, Anselmo Peñas, Antonio Ferrández (eds.), *Multilingual Information Access and Natural Language Processing. International Workshop at IBERAMIA-2002: VII Iberoamerican Conference on Artificial Intelligence*, Sevilla, Elsnet y Ritos, s/p. En línea, <<https://nlp.cic.ipn.mx/Publications/2002/MLIA-2002-Corpus.pdf>> [Consulta: 11/11/2020].
- GREGORY, Ian N.; GEDDES, Alistair (2014): *Toward Spatial Humanities. Historical GIS and Spatial History*, Bloomington, Indiana University Press.
- HALLEBEEK, Jos (1999): «El corpus paralelo», *Procesamiento del lenguaje natural*, n. 24, pp. 58-69. En línea, <<https://core.ac.uk/download/pdf/16355934.pdf>> [Consulta: 11/11/2020].
- JIMÉNEZ RÍOS, Enrique (2001): «Notas a propósito de la ordenación de las acepciones en el Diccionario de la Academia», *Anuario de Estudios Filológicos*, vol. 24, pp. 255-272. En línea, <<https://dialnet.unirioja.es/servlet/articulo?codigo=59018>> [Consulta: 11/11/2020].
- JOHNS, Tim (2000): *Tim Johns' data-driven learning Page*. En línea, <http://archive.ecml.at/projects/voll/our_resources/graz_2002?ddrivenlrning/whatisddl/resources/tim_ddl_learning_page.htm> [Consulta: 11/11/2020].
- KENNING, Marie-Madeleine (2010): «What Are Parallel and Comparable Corpora and How Can We Use Them», en Anne O'Keeffe, Michael McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*, London/New York, Routledge, pp. 487-500.

- KÜBLER, Sandra; ZINSMEISTER, Heike (2015): *Corpus Linguistics and Linguistically Annotated Corpora*, London, Bloomsbury.
- LU, Huichuan; CHENG, Anchung (2016): «Rasgos semánticos a partir de corpus y su aplicación en el aprendizaje de la estructura 'ser/estar+adjetivo' en español», *Círculo de Lingüística Aplicada a la Comunicación*, vol. 68, pp. 117-137. En línea, <<https://doi.org/10.5209/clac.54524>> [Consulta: 11/11/2020].
- LU, Huichuan et al. (2017): «Glocalized rhetoric practices on medical writing», *Círculo de lingüística aplicada a la comunicación*, vol. 69, pp. 217-236. En línea, <<https://revistas.ucm.es/index.php/CLAC/article/view/55320>> [Consulta: 11/11/2020].
- LU, Huichuan; LU, Wenxiang (2016): *The Acquisition of the Spanish Object Pronouns as a Third Language*, National Cheng Kung University, informe de resultados de investigación. En línea, <<http://ir.lib.ncku.edu.tw/bitstream/987654321/175889/1/1010203001-000001.pdf>> [Consulta: 11/11/2020].
- LUO, Xiaoqing (2017): 基于语料库的中文政治文本英西译本的语言特征差异研究, Beijing Foreign Studies University, Trabajo de Fin de Máster dirigido por Yin Jun.
- MCENERY, Tony; HARDIE, Andrew (2012): *Corpus Linguistics: Method, theory and practice*, Cambridge, Cambridge University Press. En línea, <<https://epdf.pub/corpus-linguistics-method-theory-and-practice2f8645efd2df79651f7a90fe3faf67255873.html>> [Consulta: 11/11/2020].
- MCENERY, Tony; XIAO, Richard (2007): «Parallel and Comparable Corpora: What Are They Up to?», en Gunilla Anderman, Margaret Rogers (eds.), *Incorporating Corpora. The Linguist and the Translator*, Clevedon/Buffalo/Toronto, Multilingual Matters, pp. 18-31. En línea, <<https://doi.org/10.21832/9781853599873-005>> [Consulta: 11/11/2020].
- MCENERY, Tony; XIAO, Richard (2011): «What Corpora Can Offer in Language Teaching and Learning», en Eli Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*, New York, Routledge, vol. 2, pp. 364-380. En línea, <https://www.lancaster.ac.uk/fass/projects/corpus/ZJU/xpapers/McEnery_Xiao_teaching.PDF> [Consulta: 11/11/2020].
- OLÍMPIO DE OLIVEIRA SILVA, M.^a Eugênia; PENADÉS, Inmaculada (2016): «Linguee como herramienta de enseñanza-aprendizaje de las unidades fraseológicas», en Gloria Corpus Pastor et al. (eds.), *Computerised and Corpus-Based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, Genève, Editions Tradulex, pp. 367-374. En línea, <<http://www.tradulex.com/varia/Europhras2015.pdf>> [Consulta: 11/11/2020].
- PÉREZ HERNÁNDEZ, María Chantal (2002): *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. ELiEs 18*, Barcelona, RedIris/UAB. En línea, <<http://elies.rediris.es/elies18/>> [Consulta: 11/11/2020].

- RESNIK, Philip *et al.* (1999): «The Bible as a Parallel Corpus: Annotating the “Book of 2000 Tongues”», *Computers and the Humanities*, vol. 33, pp. 129-153.
- RÖMER, Ute (2008): «Corpora and Language Teaching», en Anke Lüdeling, Merja Kytö (eds.), *Corpus linguistics: an international handbook*, Berlin/New York, Walter de Gruyter, vol. 1, pp. 112-130. En línea, <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.455.9749&rep=rep1&type=pdf>> [Consulta: 11/11/2020].
- SARTOR, Elisa (2019): «Corpus textuales y mapas digitales: un estudio de la circulación de *Arte de los metales* (1640) de Álvaro Alonso Barba en Europa y América durante la Edad Moderna a través de las humanidades espaciales», en Matteo De Beni (ed.), *Imagen y discurso técnico-científico en español. Miradas interdisciplinarias*, Mantova, Universitas Studiorum (Colección *Pliegos Hispánicos*), pp. 197-219.
- ST. JOHN, Elke (2001): «A Case for Using a Parallel Corpus and Concordancer for Beginners of a Foreign Language», *Language Learning and Technology*, vol. 5, n. 3, pp. 185-203. En línea, <https://scholarspace.manoa.hawaii.edu/bitstream/10125/44573/1/05_03_stjohn.pdf> [Consulta: 11/11/2020].
- TRIGO, Beatriz (2021): «Investigar en el contexto de las Humanidades Digitales y el español LE/L2», en Mar Cruz Piñol (ed.), *e-Research y español LE/L2. Investigar en la era digital*, London/New York, Routledge, pp. 24-39.
- WANG, Lixun (2001): «Exploring Parallel Concordancing in English and Chinese», *Language Learning and Technology*, vol. 5, n. 3, pp. 174-184. En línea, <https://scholarspace.manoa.hawaii.edu/bitstream/10125/44572/1/05_03_wang.pdf> [Consulta: 11/11/2020].
- WEI, Weiqian (2020): *Propuesta de aplicación didáctica de los corpus paralelos en el aula de ELE para los estudiantes sinohablantes*, Universitat de Barcelona, Trabajo de Fin de Máster dirigido por Mar Cruz Piñol. En línea, <<http://hdl.handle.net/2445/172281>> [Consulta: 11/11/2020].
- WU, Fan (2016): «Fraseología comparada del español y del chino: su aplicación a la enseñanza en la clase de español como lengua extranjera», *marcoELE, Revista de Didáctica Español Lengua Extranjera*, n. 22. En línea, <https://www.redalyc.org/pdf/921/Resumenes/Resumen_92153185011_1.pdf> [Consulta: 11/11/2020].
- ZANETTIN, Federico (1994): «Parallel Words: Designing a Bilingual Database for Translation Activities», *Corpora in Language Education and Research: A Selection of Papers from Talc*, vol. 94, n. 4, pp. 99-111. En línea, <https://www.academia.edu/3490875/Parallel_Words_Designing_a_Bilingual_Database_for_Translation_Activities> [Consulta: 11/11/2020].
- ZIEMSKI, Michal *et al.* (2016): *The United Nations Parallel Corpus*, Language Resources and Evaluation (LREC'16), Portorož. En línea, <<https://conferences.unite.un.org/UNCORPUS/es>> [Consulta: 11/11/2020].

Corpus

- Corpus Paralelo de las Naciones Unidas*. En línea, <<https://conferences.unite.un.org/uncorpus/>> [Consulta: 11/11/2020].
- Corpus Paralelo de Español Inglés y Chino (CPEIC)*. En línea, <<http://140.116.245.146/cpeic/TriLin.html?role=teacher>> [Consulta: 11/11/2020].
- Eshelper*. En línea, <<https://www.esdict.cn/>> [Consulta: 11/11/2020].
- Glosbe*. En línea, <<https://es.glosbe.com/>> [Consulta: 11/11/2020].
- Linguee*. En línea, <<https://www.linguee.com/>> [Consulta: 11/11/2020].
- MyMemory*. En línea, <<https://mymemory.translated.net/>> [Consulta: 11/11/2020].
- OPUS*. En línea, <<http://opus.nlpl.eu/>> [Consulta: 11/11/2020].
- PaGeS*. En línea, <<https://www.corpuspages.eu/corpus/about/about?lang=en>>. [Consulta: 11/11/2020].
- Reverso*. En línea, <<https://diccionario.reverso.net/>> [Consulta: 11/11/2020].
- ReversoContext*. En línea, <<https://context.reverso.net/traduccion/>> [Consulta: 11/11/2020].
- RST Spanish-Chinese Treebank*. En línea, <<http://ixa2.si.ehu.es/rst/zh/index.php>> [Consulta: 11/11/2020].
- The Holy Bible*. En línea, <<http://users.umiacs.umd.edu/~resnik/parallel/bible.html>> [Consulta: 11/11/2020].