

# Corpus paralelos en la enseñanza de lenguas extranjeras: un ejemplo de aplicación basado en el corpus PaGeS

## *Parallel Corpora in Foreign Language Learning and Teaching: an Example of Use Based on the Corpus PaGeS*

**Irene DOVAL**

*Universidad de Santiago de Compostela*

*i.doval@usc.es*

Recibido: julio 2018. Revisado: septiembre 2018. Aceptado: octubre 2018.

**Resumen:** Este artículo trata sobre las aplicaciones de los corpus paralelos en la enseñanza de lenguas extranjeras a un nivel avanzado. Después de una visión general de los principios didácticos del uso de corpus como método de aprendizaje inductivo, se señalan las características básicas que ha de reunir un corpus para su inserción exitosa en la clase. A continuación, se presenta el corpus paralelo bilingüe español/alemán, PaGeS, y se describen brevemente las sucesivas fases de elaboración así como las opciones de búsqueda. El potencial del corpus PaGeS se ilustra mediante un ejemplo de aplicación sobre las equivalencias alemanas del verbo *salir*.

**Palabras clave:** corpus paralelos en la enseñanza de lenguas extranjeras; aprendizaje de lenguas extranjeras; alemán como lengua extranjera.

**Abstract:** This paper explores the potential of parallel corpora in foreign language learning and teaching. After an overview of the corpus-based pedagogical method of data-driven learning, the main requirements of a corpus to be used in the classroom are outlined. Subsequently, the bilingual parallel corpus Spanish/German, PaGeS, is introduced, the steps in its creation are sketched and the different search possibilities are briefly described. The potential of the PaGeS corpus is illustrated by a practical case study: the German equivalents of the Spanish verb *salir*.

**Key words:** parallel corpora; foreign language learning and teaching; data-driven-learning; German as a foreign language.

## 1. OBSERVACIONES PRELIMINARES

En los últimos años, los corpus paralelos se han convertido en una fuente de datos indispensable no sólo para las investigaciones lingüísticas contrastivas, la traductología o la lexicografía, sino que también constituyen el material básico para el entrenamiento de los algoritmos usados en la traducción automática. Por otro lado, estos corpus pueden utilizarse también de modo muy productivo en la enseñanza de lenguas extranjeras (ELE) con estudiantes de nivel intermedio alto y avanzado, aunque hasta ahora su uso aquí ha sido más bien modesto (Boulton 2017, Brown 2017). El objetivo de este artículo es presentar una panorámica de las diversas posibilidades que ofrecen los corpus paralelos en ELE así como mostrar un ejemplo práctico de su uso.

En lingüística se entiende por corpus una colección de textos orales o escritos en una sola lengua o en varias, utilizados para analizar el uso de la lengua. Un corpus está formado por los datos primarios, esto es, los propios textos, así como los metadatos que describen estos datos y, en su caso, la anotación lingüística (Lemnitzer / Zinsmeister 2015, 40). Dado que los textos que forman los corpus poseen una cierta amplitud, posibilitan la búsqueda y el análisis sistemático de los contextos de las unidades lingüísticas (Lüdeling / Walter 2010, 2).

Los corpus se clasifican según diversos criterios. Uno de ellos son las lenguas que lo componen (Lemnitzer / Zinsmeister 2015, 102ff); según éste, los corpus pueden ser monolingües –contienen un solo idioma–, bilingües o multilingües –contienen dos o más idiomas. En el caso de estos últimos, se distingue entre corpus comparables y paralelos, según la terminología comúnmente aceptada (McEnery / Xiao 2007, 2-3). Los corpus comparables se componen de textos monolingües en diferentes idiomas que comparten tema, tipología textual y registro con un origen y un alcance similares. Los textos de un corpus comparable no son traducciones unos de otros, sino que se seleccionan según criterios comunes en cuanto a la tipología textual. Los corpus paralelos, por el contrario, contienen textos originales en una lengua determinada junto a sus traducciones a otra o a más lenguas, por lo que también son ocasionalmente denominados como corpus de traducción.

La lingüística de corpus ha revolucionado la descripción de las lenguas, ya que en la actualidad la mayoría de los estudios lingüísticos se basan en corpus que son los que suministran los datos primarios de la investigación lingüística. K. Aijmer (2013, 1) resume así este desarrollo a la vez que aboga por su extensión al aula: «Corpora have changed our views on language and language use and we can also expect to find them in the classroom». Debido a la gran cantidad de material auténtico que presentan, los corpus monolingües se han ido introduciendo paulatinamente en la enseñanza de lenguas extranjeras. A. Boulton (2017) ofrece una cronología de los hitos más importantes en la investigación sobre el uso de corpus en la enseñanza de lenguas. Desde sus inicios en los años 90 reflexionan sobre su inserción en la clase de ELE. Un creciente número de publicaciones exploran métodos basados en corpus y plantean actividades concretas, especialmente para el caso del inglés (Braun / Chambers 2006, Aijmer 2009, McEnery / Xiao 2011) aunque también se encuentran contribuciones para otras lenguas, como el alemán (Tschirner 2009, Chrissou 2010, Wallner 2013, Ahrenholz 2013).

Pero este creciente interés se ha focalizado casi exclusivamente en el uso de corpus monolingües, entre otros factores debido fundamentalmente a su mayor disponibilidad. Así, hasta la fecha sólo artículos esporádicos tratan concretamente los corpus paralelos y sus aplicaciones en la didáctica de lenguas extranjeras, y casi exclusivamente relacionados con la enseñanza del inglés (Bernardini 2004, Brown 2017). A este tema, a las aplicaciones de un corpus paralelo en la enseñanza del alemán y/o español como lengua extranjera, está dedicado este artículo. En la sección siguiente, se presenta una panorámica de las aplicaciones generales de los corpus paralelos en la enseñanza de lenguas extranjeras y se introducen brevemente tres corpus paralelos que contienen español y alemán, lenguas en las que se centra este trabajo. A continuación, se describe brevemente el corpus paralelo PaGeS (Parallel Corpus German / Spanish): la motivación para su creación, el diseño, la anotación y las opciones de búsqueda. El apartado 4 ilustra el potencial didáctico de este corpus con un ejemplo práctico: los equivalentes alemanes del verbo español *salir*. Por último, se hace un balance final.

## 2. CORPUS PARALELOS EN LA ENSEÑANZA DE LENGUAS EXTRANJERAS: ALGUNAS CONSIDERACIONES

Leech (1997:2) señaló que la convergencia entre los corpus y la enseñanza de lenguas se centra en dos aspectos: el uso indirecto de los corpus y su uso directo en la clase: enseñar acerca de los corpus, enseñar a explotarlos y explotarlos para enseñar (p. 3). Así, en su uso indirecto los corpus sirven de base para el diseño y la preparación de materiales didácticos, obras de referencia como manuales o diccionarios o el desarrollo de pruebas de nivel. Por otro lado, los corpus pueden ser usados directamente

en ELE, presuponiendo, eso sí, una fase inicial de introducción tanto para profesores como para alumnos. En la actualidad, la investigación sobre ELE hace hincapié en las ventajas del aprendizaje inductivo basado en corpus (Mukherjee 2002, Bernardini 2004, Lüdeling/Walter 2009). Se trata de que los aprendices descubran por sí mismos patrones de uso de la lengua extranjera utilizando datos lingüísticos auténticos proporcionados por los corpus, el llamado *data-driven-learning*, (DDL) descrito por Johns (1991, 2), uno de los pioneros de este enfoque, de la siguiente forma:

[...] the language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data – hence the term «data-driven learning» (DDL) to describe the approach.

Según Johns, el término DDL se refiere a la capacidad por parte de los alumnos de usar datos de corpus para descubrir -en lugar de ser instruido sobre ello- qué significan las palabras o cómo se utilizan y llegar a sus propias conclusiones simplemente por estar expuestos a los datos. Referido concretamente a los corpus paralelos, este enfoque fomenta la autonomía del aprendizaje al animar a los aprendices a explorar por sí mismos los hechos lingüísticos para descubrir regularidades en los equivalentes traductológicos o en los contextos gramaticales. Por ejemplo, los estudiantes de alemán pueden, sin necesidad de recurrir a un diccionario, sintetizar los diferentes significados de un verbo alemán altamente polisémico como *bekommen* a la vista de los diferentes contextos semánticos en que se presenta y de sus equivalentes traductológicos en español, obtenidos estos mediante la consulta de un corpus paralelo, tal como muestra la Figura 1.

ES ⇌ DE [SS] segmento\_txt\_de:bekommen 🔍 Europarl v7

Ergebnisse: 5235 Seiten: 175 Aktuelle Seite: 1

<p>War ich krank? Hatten die Geschwister etwas <b>bekommen</b>, was ich nicht <b>bekommen</b> hatte? Stand für den weiteren Verlauf des Tages Unangenehmes, Schwieriges an, das ich bestehen mußte? [0004, 7]</p>	<p>¿Quizá estaba enfermo? ¿Les habían dado a mis hermanos algo que no me habían dado a mí? ¿Me esperaba aquel día algún trance desagradable o difícil? [0004, 7]</p>
<p>»Dir auch«, sagte Harry. »Schau dir das mal an! Ich hab Geschenke <b>bekommen</b>!« »Was hast du erwartet, Runkelrüben?«, sagte Ron und machte sich an seinen eigenen Stapel, der um einiges größer war als der Harrys. [0047, 12]</p>	<p>—Para ti también —contestó Harry—. ¡Mira esto! ¡Me han enviado regalos! — ¿Qué esperabas, nabos? —dijo Ron, volviéndose hacia sus propios paquetes, que eran más numerosos que los de Harry. [0047, 12]</p>
<p>Er schafft seine Arbeit nicht mehr – die Arbeit schafft ihn. Wichtige Vorgänge bleiben auf der Strecke. Mitarbeiter <b>bekommen</b> keine Termine. Meetings fallen aus. Kundenanfragen bleiben ohne Antwort. [0053, 1]</p>	<p>Ya no saca adelante el trabajo, el trabajo puede con él. Las tareas importantes se quedan en el camino. Los empleados no reciben un calendario. Las reuniones se suspenden. Las solicitudes de los clientes se quedan sin respuesta. [0053, 1]</p>
<p>Mach ihm klar, daß das kaum schlimmer ist als die Operation eines Weisheitszahns. Erfinde ein paar nette Geschichten, zum Beispiel, daß Berlusconi's Tochter sich hier operieren ließ, oder irgendeinen anderen Blödsinn. Er muß Vertrauen <b>bekommen</b>. Wir sperren ihn nicht ein. Er bekommt ein anderes Zimmer, hell und groß. [0072, Die Messer...]</p>	<p>Déjale claro que es como si le quitaran una muela del juicio. Invéntate un par de historias bonitas, que la hija de Berlusconi se operó aquí, o cualquier otra tontería. Tiene que confiar. Nosotros no lo encerraremos. Lo alojaremos en otra habitación, luminosa y grande. [0072, Afilar los...]</p>

»Weiß der Teufel, wo ich ihn hingellegt habe.« »Was ist mit dem Geld?« fragte —El diablo sabrá dónde las he dejado. —¿Qué hay del dinero? -preguntó

Figura 1: Lista de ejemplos de traducción del verbo *bekommen* (Corpus PaGeS)

Este estilo de aprendizaje «descubridor» aumenta la motivación y, por lo tanto, el éxito del aprendizaje (Gut 2007,7), a la vez que fomenta en los estudiantes los hábitos

del trabajo científico. Los estudiantes de lenguas extranjeras pueden aprender las técnicas de observación e interpretación de datos al explorar una determinada cuestión lingüística utilizando un corpus, practicando fases del trabajo científico tales como la formulación de hipótesis y su verificación a partir de datos reales del idioma (Johns 1991, 3).

A través de los corpus paralelos se puede asimismo obtener información precisa sobre la frecuencia real de los diversos fenómenos lingüísticos en sus contextos y sus colocaciones más frecuentes, tanto en la lengua materna como en la lengua extranjera. También se obtienen datos fiables sobre la gramaticalidad o la frecuencia de uso de ciertas construcciones (Mukherjee 2004). Los corpus paralelos son, por otro lado, un complemento útil a los diccionarios bilingües, ya que sirven como fuente de sugerencias de traducción adecuadas a los contextos concretos y, si permiten búsquedas multipalabra, son claramente más eficaces a la hora de ofrecer propuestas de traducción de expresiones idiomáticas y de otras combinaciones de palabras.

En cuanto al uso indirecto, los corpus, dado que reflejan el uso real de la lengua, son una fuente ideal para la producción de materiales didácticos y ejercicios auténticos. En este sentido, los corpus paralelos suponen un complemento de los libros de texto y de los cuadernos de ejercicios.

Ahora bien, para poder hacer un uso efectivo de las posibilidades descritas, los corpus paralelos han de reunir ciertos requisitos en cuanto a tamaño, tipo y calidad de los textos originales y de las traducciones, anotación lingüística y facilidad de uso. Si carecen de las características adecuadas, las potencialidades didácticas descritas no se hacen efectivas y su uso va a quedar necesariamente relegado. A continuación, se presentan brevemente tres corpus paralelos, incidiendo en aquellos aspectos relevantes para su uso en la clase de ELE. Estos corpus son *Opus*, *Intercorp* y *Linguee*, que son los corpus en línea de libre acceso más amplios que incluyen al alemán y al español.

OPUS es una colección creciente de textos de la web traducidos en diferentes lenguas (<http://opus.lingfil.uu.se/>) que forman la mayor colección existente de corpus paralelos multilingües disponibles gratuitamente. Se trata de un recurso en constante crecimiento que también proporciona herramientas para el procesamiento de los datos, así como múltiples sistemas de búsqueda para consultas de corpus. Esta colección fue recopilada por Jörg Tiedemann y actualmente está compuesta por 43 corpus en más de 90 idiomas (cf. Tiedemann 2016). Los textos se descargan automáticamente de Internet. Obviamente, no todos los pares de lenguas están representados cuantitativamente por igual. El par de lenguas alemán/español consta de 36 millones de bisegmentos alineados (Tiedemann 2012, 2214ff.). Los principales ámbitos cubiertos por OPUS son el legislativo y el administrativo, pues buena parte de los textos proceden de las instituciones de la Unión Europea. Hay, además, aunque en menor medida, otros tipos de texto como subtítulos, textos periodísticos y algunas otras colecciones más pequeñas de diversas fuentes de internet.

El corpus multilingüe InterCorp (<http://ucnk.ff.cuni.cz/intercorp>) tiene como objetivo servir de fuente de datos para estudios lingüísticos, la lexicografía, aprendizaje de lenguas, aplicaciones informáticas, así como para traductores y el público en general. La lengua pivote es el checo, esto es, todos los textos del corpus tienen una versión checa, sea la lengua original o una traducción, y es con ella con la que se alinean primariamente. Este corpus contiene 40 idiomas que, combinados, forman un gran número de subcorpus bilingües. Consta de dos partes, el núcleo y las colecciones (Čermák 2019). El núcleo está formado por textos literarios que han sido alineados manualmente. Las colecciones contienen principalmente recursos multilingües de las instituciones europeas (PressEurop, Acquis y Europarl). El subcorpus alemán/español consta de unos cinco millones de palabras (<http://ucnk.korpus.cz/intercorp/?lang=en>) y está lematizado y morfosintácticamente anotado con TreeTagger (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>).

Por último, Linguee (<https://www.linguee.com/>) es una colección comercial multilingüe de textos que busca automáticamente sitios web bilingües. Debido a que combina las funcionalidades de un diccionario con las de un motor de búsqueda (<http://www.linguee.de/deutsch-englisch/page/about.php>) se ha convertido en una herramienta muy popular. Los resultados de la búsqueda son presentados en dos partes. En la parte superior, «Diccionario», se presentan las traducciones más comunes de los términos buscados, como en un diccionario en línea; en la parte inferior, «Fuentes externas», se enumeran un gran número de ejemplos bilingües con los términos buscados, la mayoría de ellos pertenecientes al campo administrativo, económico o comercial.

Sin embargo, estos tres corpus presentan una serie de limitaciones para su pleno aprovechamiento en ELE. La limitación más importante es que en su mayor parte se trata de textos de dominios muy específicos, principalmente el lenguaje administrativo y comercial. Este inconveniente ya fue señalado hace una década por Heid (2008, 137): «only few parallel corpora are available, and many of them are specialized in terms of text types (e. g. parliament debates) and/or domains (e. g. technical documentation)» y corroborado por Steinberg *et al.* (2014, 4): «Another restriction is linked to the text domain, which for the bulk of the corpus is legal and administrative». Debido a esta limitación, los citados corpus, aun conteniendo ingentes cantidades de textos, son muy monótonos tanto a nivel sintáctico como léxico y están bastante alejados de la lengua general, por lo que resultan poco adecuados para una clase de lengua extranjera, siempre que no sea con fines específicos (Doval 2017b). Por otra parte, en estos corpus no se puede determinar inequívocamente la lengua en la que fue originalmente creado el texto y, aun en ese caso, es incierto si se ha producido una traducción directa o ha habido una lengua puente de la que ulteriormente se hicieron traducciones a los diferentes idiomas. Para el par de lenguas alemán/español, se puede suponer que una traducción directa es más bien la excepción, y que la mayoría de las traducciones

se hacen indirectamente a través de un tercer idioma, generalmente inglés. Además, los datos, o bien carecen de anotaciones lingüísticas, como en Linguee, o, si están disponibles, como en Opus o InterCorp, no están diseñados para ser consultados de manera intuitiva, ya que han sido diseñados pensando en la investigación lingüística y no en su aplicación pedagógica. Estas críticas han sido bien sintetizadas por Michael Brown (2017):

A common critique of data-driven learning (DDL) and the direct use of corpora in the classroom is that many, if not most, of the corpora available today are inappropriate for language learners because the level of the language in them is too high and the language fragments presented in concordance lines confuse learners. [...] And the second one is that even if the language were of the appropriate level, the tools for manipulating and searching corpora are not learner-friendly; they are designed with researchers, not language learners, in mind.

Es evidente que, en el contexto de ELE, con una disponibilidad de tiempo limitada, si el uso de una determinada herramienta, en este caso un corpus, requiere por parte de profesores y alumnos una fase previa más o menos larga de familiarización en la mayoría de los casos terminará por disuadir de su uso.

### 3. SOBRE EL CORPUS PARALELO ALEMÁN / ESPAÑOL (PaGeS)

Ante estas limitaciones, comenzamos en la Universidad de Santiago con la elaboración de un corpus bidireccional bilingüe paralelo alemán/español, Parallel Corpus German/Spanish (PaGeS) con el objetivo de satisfacer lo mejor posible no sólo las necesidades de ELE, sino también las de otras aplicaciones. Teniendo en cuenta que la creación de un corpus es una vasta tarea, tanto en tiempo como en esfuerzo, se ha pretendido que PaGeS pueda ser explotado para múltiples propósitos. Además de ELE, se pueden distinguir otros cuatro grandes campos de aplicación de los corpus paralelos, cada uno de ellos con sus usuarios específicos: la investigación básica en lingüística contrastiva y translatoología, la lexicografía, la traducción y el procesamiento del lenguaje natural, especialmente para la traducción automática (Doval 2017b: 127).

Por un lado, el corpus puede ser usado de manera indirecta, como material base para la elaboración de diccionarios bilingües y para la producción de ejercicios o de ejemplos ilustrativos de distintos fenómenos. Por otro lado, puede constituir un recurso de enseñanza directo en la clase de lengua y traducción para aprendices de alemán o español de nivel intermedio a avanzado. A través de PaGeS pueden obtener un gran número de sugerencias de traducción para un determinado ítem, mostradas directamente en ejemplos de uso.

Pero para desempeñar satisfactoriamente estas funciones, el corpus ha de cumplir varios requisitos. En primer lugar, los datos han de poseer una calidad contrastada

y esto tanto para los textos originales como para las traducciones. Además, el texto de la lengua original ha de identificarse claramente y las traducciones han de ser preferiblemente directas de ese original y no a través de una lengua puente. En segundo lugar, el tipo de texto ha de cubrir lo más posible la variante estándar actual, un registro general. En tercer lugar, los datos deben estar lingüísticamente anotados (lematización y etiquetado PoS). Y, por último, y muy importante, el corpus ha de disponer de una interfaz de consulta intuitiva y sencilla accesible a usuarios legos.

El Corpus Paralelo Alemán/Español, PaGeS, ([www.corpuspages.eu](http://www.corpuspages.eu)) reúne una colección de textos originales en español y alemán y sus respectivas traducciones publicadas en editoriales de prestigio; también contiene una pequeña proporción de traducciones en español y alemán de textos de un tercer idioma (inglés, francés, sueco e italiano).

En noviembre de 2018, PaGeS contiene aproximadamente 25 millones de palabras (si se agrega puntuación, se obtienen más de 28 millones de tokens) y 858.470 bisegmentos alineados (oraciones o segmentos más pequeños), como se muestra en la Tabla 1.

Lengua	Obras	Bisegmentos	Tipos	Palabras
Alemán Original	62	386.314	158198	5,081,806
Alemán Traducción < Español	54	319.315	136543	5,057,274
Alemán Traducción < 3.ª Lengua	18	152.841	74339	2,143,959
Español Original	54	319.315	103674	5,073,514
Español Traducción < Alemán	62	386.314	102431	5,311,191
Español Traducción < 3.ª Lengua	18	152.841	57509	2,161,909
<b>Total</b>	<b>134 (x2)</b>	<b>858. 470</b>		<b>24.829.653</b>

Tabla 1: Número de obras, bisegmentos, tipos y palabras en PaGeS (Noviembre 2018)

Se trata de una colección en desarrollo de obras publicadas después de 1960, con especial atención a obras de las dos últimas décadas. Se incluyen libros de ficción y algunos de no ficción (ensayos, textos biográficos y de divulgación científica), ya sea como obras completas o fragmentos de cierta amplitud. Los textos de ficción constituyen la gran mayoría (actualmente alrededor del 85%), ya que son los que con preferencia se traducen a otros idiomas y, por lo tanto, representan la mayor parte de los recursos disponibles. Al seleccionar las obras originales, se pretendió una cierta diversidad dialectal. Así, el corpus contiene obras no sólo de escritores alemanes y españoles, sino también de autores latinoamericanos, austriacos o suizos. Dados los objetivos primarios del corpus, la investigación lingüística y la enseñanza de lenguas, resultaba imprescindible asegurar la calidad del material, tanto con respecto a los textos originales como a su traducción. Para garantizarla, los textos tenían necesariamente



que haber pasado algún control de calidad. Para ello la única vía disponible era acudir a materiales publicados por editoriales reconocidas, donde originales y traducciones han sido sometidas a un estricto control de calidad. Para cada obra se almacenan metadatos sobre el autor, el título, el año de publicación de la primera edición y el año de publicación de la edición utilizada, así como sobre el traductor y la variedad dialectal de la lengua original (Doval *et al.* 2019). La Figura 2 muestra el porcentaje de palabras según los idiomas originales y las traducciones, así como su número absoluto.

## Palabras en PaGeS

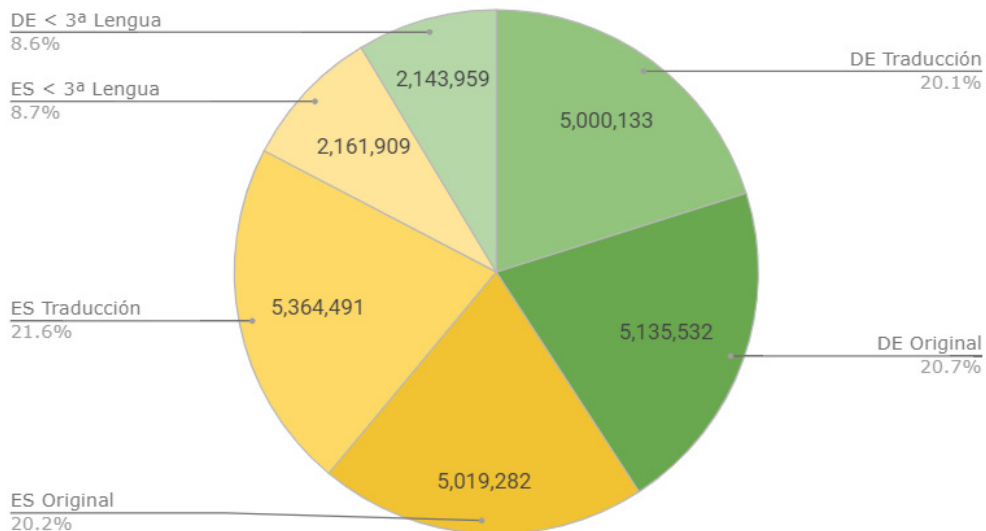


Figura 2: Composición actual del corpus PaGeS (Noviembre 2018)

Para garantizar la calidad, tanto el preprocesado como el alineado de los textos fueron revisados manualmente. El alineado es un paso crucial en la construcción y explotación de un corpus paralelo. Tiedemann (2011, 123) lo define «as a process of making symmetric correspondences explicit in order to enable further processing of parallel resources». Este conjunto de correspondencias de dos textos forma el bitexto alineado. En PaGeS hemos usado el programa LF-Aligner (<http://sourceforge.net/projects/aligner/>), ya que en varias pruebas esta herramienta de alineación alcanzó la mayor precisión. Su formato de salida es una tabla de Excel, que se importa posteriormente a las hojas de cálculo de Google que son las que editamos y sobre las que hacemos la revisión, por ser más manejables y fácilmente compartibles (Doval 2017b: 135).

PaGeS está anotado a nivel morfosintáctico con clases de palabras, a lo que la lingüística de corpus suele referirse como PoS (Part-of-Speech)-Tagging. Esta anotación ha sido realizada automáticamente por el TreeTagger,

mencionado anteriormente, para los textos en alemán y por Freeling (<http://nlp.lsi.upc.edu/freeling/node/1>) para los textos en español (Doval 2017a). El motor de búsqueda Apache Solr (<http://lucene.apache.org/solr/#intro>), de código abierto, es la herramienta utilizada para la indexación y búsqueda. Las consultas se realizan a través de una interfaz muy intuitiva (ver Fig. 2), sin que requiera ningún tipo de formación previa. Como ya se ha indicado, la facilidad y la inmediatez de uso de un corpus, junto con distintas posibilidades de búsqueda, son claves para su inserción exitosa en ELE.

El corpus PaGeS está disponible en línea desde marzo de 2016 y es de libre acceso para fines científicos. La Figura 3 muestra la interfaz de búsqueda simple o estándar. El usuario simplemente tiene que introducir la palabra o palabras buscadas en español o alemán, ya que el corpus es totalmente bidireccional. Estas búsquedas son por defecto lematizadas, a no ser que se inserte la palabra entre comillas, con lo cual se restringe la búsqueda a la palabra exacta. Si se hace una búsqueda multipalabra se obtienen contextos en que aparecen todas las palabras introducidas lematizadas, siempre que se encuentren a una distancia máxima de cinco palabras. Así se obtiene con celeridad una lista de resultados con su traducción, primero se muestran los resultados en los textos originales en la columna izquierda, luego en las traducciones en la columna derecha. La cadena buscada se resalta en negrita y se muestra junto con un pasaje de texto anterior y posterior, tal como se ve en la figura 3. Para más opciones de búsqueda, vid. Doval 2018 y [www.corpuspages.eu/help](http://www.corpuspages.eu/help). Los resultados de la búsqueda pueden ser posteriormente exportados a archivos de texto y descargados localmente.

The screenshot shows the PaGeS search interface. At the top left is the USC logo. In the center is the PaGeS logo with the text 'Parallel Corpus German Spanish'. To the right is an 'Entrar' button. Below the logo is a navigation bar with 'en | es | de | gl' and buttons for 'Buscar', 'Búsqueda avanzada', and 'Ayuda'. On the right side of the navigation bar are links for 'Sobre PaGeS', 'Lista de obras', 'Equipo', and 'Contacto'. The main search area shows 'ES ⇌ DE' and a search box containing 'intervenir'. Below the search box are 'Resultados: 933', 'Páginas: 32', and 'Página actual: 1'. The results are displayed in two columns. The left column shows the original text in Spanish with the search term highlighted in bold. The right column shows the German translation with the search term highlighted in bold. The search term 'intervenir' is highlighted in bold in both columns. The results are: 'Sólo que no te prestarían la menor atención. —Me lo imagino. **Intervino** Marina. —Yo también tengo una pregunta. Tal como iba la conversación, la curiosidad de la niña podía depararme cualquier novedad. [0120, 17]' and 'Nur würde man dir keinen Funken Aufmerksamkeit schenken.« »Kann ich mir vorstellen.« Da mischte sich Marina ein. »Ich habe auch eine Frage.« So wie das Gespräch verlief, konnte die Neugier des Mädchens jede Überraschung für mich bereithalten. [0120, 17]'. Below this is another result: 'Sólo conservé una pistola, y con una sola bala. Úrsula no **intervino**. La única vez que lo disuadió fue cuando él estaba a punto de destruir el daguerrotipo de Remedios que se conservaba en la sala, alumbrado por una lámpara eterna. [0023]' and 'Nur die Pistole behielt er, und nur eine Kugel. Ursula mischte sich nicht ein. Nur einmal widersprach sie, nämlich als er Remedios' Daguerreotyp vernichten wollte, der, von einem ewigen Lämpchen erleuchtet, im Wohnzimmer hing. [0023]'. The interface also shows 'Europarl v7' on the right side of the search results area.

Figura 3: Interfaz de la búsqueda standard de PaGeS

El diseño de PaGeS está orientado a crear un recurso multifuncional que satisfaga las necesidades de diferentes grupos de usuarios y sea adecuado para distintas

funciones (Doval 2017b). Aquí reside uno de los mayores desafíos en la confección de un corpus paralelo: saber conciliar las necesidades de usuarios exigentes como los investigadores, pero no a costa de perder, por ofrecer una interfaz demasiado compleja, a otros usuarios que simplemente quieren consultar el uso de una palabra o expresión, o encontrar rápidamente un equivalente de traducción.

Así, PaGeS presenta una vasta colección de usos reales de lengua que sirven de fuente de datos para la investigación en lingüística contrastiva, para estudios de traducción o para la lexicografía bilingüe. A la vez, pretende ser un recurso útil para los estudiantes avanzados de español o alemán como lengua extranjera, donde se pueden obtener diversos contextos de uso y ejemplos de traducción. El siguiente apartado está dedicado a este último grupo de usuarios, e ilustra el uso de corpus paralelos mediante un ejemplo práctico.

#### 4. UN CASO PRÁCTICO: EQUIVALENTES ALEMANES DEL VERBO ESPAÑOL *SALIR*

Como ilustración de un uso directo de los corpus paralelos en ELE, se presenta a continuación una propuesta de actividad práctica dirigida a aprendices de alemán de un nivel intermedio y avanzado. Se trata de analizar, mediante el corpus PaGes, los equivalentes alemanes del verbo español de movimiento *salir*.

El punto de partida de esta actividad ha sido la observación de que la expresión de los eventos de movimiento resulta particularmente difícil a los aprendices de alemán con lengua materna español. A menudo los estudiantes tratan en vano de encontrar un equivalente alemán (unívoco) para verbos de movimiento españoles tan típicos como *entrar*, *salir*, *subir* o *bajar*. Esta demanda no resulta en absoluto sorprendente, si se tiene en cuenta que en alemán no existen verbos simples comparables a estos verbos españoles, de ahí la dificultad que suponen.

Es bien sabido que el español suele expresar la dirección del movimiento (trayectoria) a través de los llamados verbos de trayectoria (*salir*, *bajar*, *cruzar*, etc.), que describen el curso del movimiento, sea de dentro afuera o de arriba abajo o de uno a otro lado. En alemán, sin embargo, la trayectoria se expresa por lo general a través de adverbios o partículas (satélites) y los verbos suelen describir el tipo de movimiento. En base a esta distinción, las lenguas pueden ser categorizadas tipológicamente según donde expresan prototípicamente la trayectoria (vid. Talmy 2000, 117 y ss.). Según esta clasificación, el español pertenece a las llamadas lenguas de marco verbal (trayectoria expresada en el verbo), mientras que el alemán pertenece a las lenguas de marco satélite (trayectoria expresada en los satélites verbales, normalmente preposiciones, adverbios o preverbios). Debido a esta divergencia, incluso los estudiantes hispanohablantes avanzados de alemán tienen considerables dificultades para liberarse de las

estructuras de su lengua materna. El objetivo de la actividad es, por tanto, llamar la atención de los alumnos sobre este problema para, de esta manera, profundizar en su conocimiento de la lengua alemana.

Con la ayuda del corpus PaGeS, los alumnos han de compilar de forma autónoma una lista de 100 ejemplos en los que aparezca el verbo *salir* en su uso espacial concreto. Dado que la configuración de búsqueda predeterminada es lematizada, la búsqueda es muy fácil de realizar. Los resultados de la búsqueda se muestran como oraciones completas, es decir, el verbo *salir* se presenta dentro de su contexto inmediato, tal como puede observarse en la Figura 4.

<p><b>Salió</b> del restaurante sin ánimos para preguntar si seguía allí el cocinero gallego al que casi había besado en su anterior visita, tras el inenarrable pavo relleno con salsa de granada. [0012]</p>	<p>Er verließ das Restaurant, ohne daß er Lust gehabt hätte zu fragen, ob der galicische Koch noch da sei, den er nach dem unbeschreiblichen gefüllten Truthahn in Granatapfelsauce bei seinem letzten Besuch beinahe geküßt hätte. [0012]</p>
<p><b>Salió</b> la indudable Patrice al frente de un cuerpo ya algo viejo y grueso pero muy bien retocado por la corsetería, con el rostro pintado al temple por una mano experta en restauraciones. [0012]</p>	<p>Dann trat die unbezweifelbare Patrice auf, mit einem Körper, der schon etwas alt und dick, aber mittels Korsettierung in Form gebracht war, und einem Gesicht, das eine kundige Hand mit Temperafarben restauriert hatte. [0012]</p>
<p><b>Salió</b> para volver con un holandés alto y fuerte que se parecía, como sólo puede parecerse un holandés alto y fuerte a otro holandés alto y fuerte, al inspector que le había interrogado en el hotel. [0012]</p>	<p>Sie kam wieder in Begleitung eines großen, korpulenten Holländers, der dem Inspektor, der ihn im Hotel befragt hatte, so sehr glich, wie nur ein großer, korpulenter Holländer einem anderen großen, korpulenten Holländer gleichen kann. [0012]</p>
<p>La espera tenía un algo de quimérico porque todo podía ser tan normal como parecía o, en cualquier caso, el matrimonio Singel podía tomarse bastantes horas de descanso y no <b>salir</b> a la calle hasta el día siguiente. [0012]</p>	<p>Das Warten hatte etwas Irreales, weil alles so normal sein konnte, wie es den Anschein hatte; auch konnte sich das Ehepaar Singel alle Zeit der Welt lassen und nicht vor dem nächsten Tag aus dem Haus gehen. [0012]</p>
<p>No podía pasar a otro estante porque hubiera perdido de vista la entrada en el hotel. A la media hora <b>salió</b> el señor Singel con su clavada humanidad orientada hacia la plaza Dam. Carvalho le siguió. [0012]</p>	<p>Zu einem anderen Regal konnte er nicht gehen, ohne den Hoteleingang aus den Augen zu verlieren. Nach einer halben Stunde kam Singel heraus und steuerte seine geklonte Korpulenz in Richtung Dam. Carvalho folgte ihm. [0012]</p>
<p>Carvalho le dio diez florines para empezar y la resistencia del taxista se diluyó. <b>Salió</b> del coche para fisgar en el motor por si alguien le llamaba la atención por</p>	<p>Carvalho gab ihm zehn Gulden Vorschuß, und sein Widerstand schmolz dahin. Er stieg aus und stocherte im Motorraum herum, falls es Ärger geben sollte, weil</p>

Figura 4: Lista de concordancias de *salir* en PaGeS

Sin embargo, los alumnos deben efectuar una labor previa de filtrado prescindiendo de los resultados erróneos debido a fallos ocasionados por el etiquetador. Por ejemplo, los sustantivos como *salida* o *sal* son clasificados erróneamente como formas verbales. Asimismo, hay que excluir los resultados en los que *salir* no denota un movimiento real, sino que tiene un significado metafórico (por ejemplo, casos como «le salió mal la jugada»). Una vez filtrados los 100 ejemplos relevantes, aprovechando la función de exportación que proporciona PaGeS, los alumnos han de descargarlos para poder procesarlos localmente. Las tareas que se señalan seguidamente y sus resultados, corresponden a un ejemplo de muestra.

A continuación, los estudiantes han de analizar los verbos alemanes correspondientes, clasificándolos según su categoría semántica, donde se pueden distinguir cuatro grupos: verbos generales de movimiento, verbos de manera de movimiento, verbos de trayectoria y, por último, correspondencia no verbal. De cada uno de estos grupos los estudiantes han de seleccionar unos cuantos ejemplos que ilustren cada clase. La Figura 5 muestra el porcentaje de cada grupo.

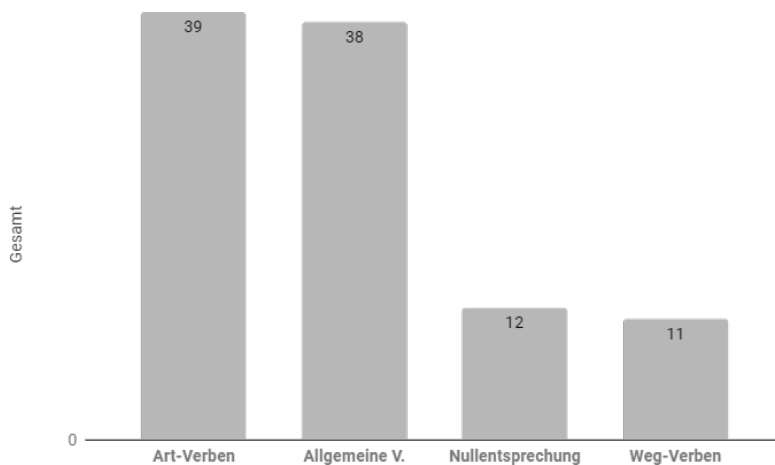


Figura 5: Distribución de los equivalentes alemanes por clase de verbos

Como puede verse en el gráfico, el verbo de trayectoria *salir* se traduce en aproximadamente el 80% de los casos o por un verbo general de movimiento (como *kommen* o *gehen*) o por un verbo de manera de movimiento. El 20 % de los casos restantes se reparte entre verbos de trayectoria y correspondencia no verbal. El único verbo de trayectoria documentado es *verlassen*, que aparece únicamente en 11 casos.

- (1) Salió del dormitorio y caminó hasta la cocina. [0103]<sup>1</sup>  
 Sie verließ das Schlafzimmer und ging in die Küche.

Especial atención se ha de dedicar a los casos de correspondencia cero, en los que la traducción alemana no contiene verbo de movimiento alguno. En alemán prácticamente cualquier verbo en combinación con una frase direccional, sea preposicional o adverbial, puede expresar fácilmente un movimiento, como en los ejemplos (2) con el verbo modal *können* y (3) con el verbo *wagen* «atreverse». En español, en cambio, los verbos de movimiento son absolutamente necesarios debido a su función central como portadores de la semántica de la trayectoria:

- (2) Mañana, cuando derribe la habitación, el niño podrá *salir*. [0006]  
 Morgen, wenn der Raum abgerissen wird, *kann* der Junge wieder *raus*.  
 (3) José Arcadio no se atrevió a *salir* de su casa en varios días. [0023]  
 Mehrere Tage *wagte* sich José Arcadio nicht mehr *aus dem Haus*.

1. Los números entre corchetes se refieren a la numeración de las obras del corpus Pa-GeS, de los que han sido extraídos los ejemplos. Vid. <https://www.corpuspages.eu/corpus/listofworks?lang=es>. El texto original se pone en primer lugar y después la traducción.

Además de los verbos modales, los verbos de sonido son a menudo reinterpretados en alemán como verbos direccionales, utilizándose para un desplazamiento que es acompañado de ese sonido:

- (4) Damit *polterte* der alte Zentaur aus dem Saal. [0001]  
Dicho esto, el viejo centauro salió *ruidosamente* de la sala.

La Tabla 2 enumera los verbos alemanes usados en la traducción de *salir* y el número de sus ocurrencias.

Clase verbal	Verbo (Ocurrencias)
V. generales de movimiento	gehen (23), kommen (10), bewegen, erreichen, führen, bringen, erscheinen
V. de trayectoria	verlassen (11)
∅ Verbo	Correspondencia cero (9), frases nominales (3)
Verbos de manera de movimiento	steigen (5), treten (4), rennen (4), laufen (4), stürzen (3), fahren (3), fliehen, springen, eilen, flattern, hasten, poltern, quellen, rasen, sausen, schieben, schleichen, schwingen, stapfen, stolpern, strömen, treiben

Tabla 2: Correspondencias verbales alemanas de *salir*

Tal como se recoge en la tabla, el alemán se vale con gran frecuencia para describir el movimiento de dentro afuera de un complemento directivo de origen, que indica la trayectoria, acompañado de una amplia variedad de verbos de manera de movimiento que describen la naturaleza del movimiento con más detalle. La tabla anterior sólo muestra los resultados en una única lista de 100 ejemplos, elegida al azar. Se podrían comparar los resultados obtenidos por los estudiantes, pudiendo, al final, recopilar en una única lista todos los equivalentes encontrados. La siguiente tarea podría ser una comparación entre el alemán y el español con respecto a la expresión de la manera de movimiento. Los alumnos analizarán si la manera de movimiento expresada en el verbo alemán se reproduce en español y, en caso afirmativo, a través de qué construcciones lingüísticas. Así se observará que en varios casos la precisión sobre la manera de movimiento en español se ignora. En el ejemplo (5), el movimiento rápido del verbo *rennen* no se recoge en la traducción española. Lo mismo ocurre en el ejemplo (6), en el que la manera especial de movimiento del verbo *stapfen* («mit festen Schritten gehen und dabei die Beine höher anheben und kräftig auftreten, sodass der Fuß im weichen Untergrund einsinkt» [caminar con pasos firmes levantando las piernas y pisando con fuerza de forma que el pie se hunde en el suelo blando] vid. [www.duden.de](http://www.duden.de)) no está recogida en la traducción española.

- (5) Selbstverständlich konnten wir nicht unten zur Vordertür *herausrennen* [...] [0019]  
Naturalmente, no podíamos salir por la puerta delantera [...]
- (6) Er nickte ihr zu und *stapfte* wortlos aus der Scheune, [0073]  
El le hizo una señal con la cabeza y salió del granero,

En otros casos, la manera de movimiento sí se especifica en el texto español. Aquí se pueden clasificar los ejemplos según los medios lingüísticos utilizados. En la mayoría de los casos la manera de movimiento se expresa mediante un sintagma adverbial, como en (7) *a hurtadillas* para *schleichen* y en (8) *a trompicones* para *stolpern*, con un participio, como en (9) *disparado* para *stürzen*, o con un gerundio, como en (10) *corriendo* para *hasten*.

- (7) Kurz vor neun *schlich* Laurenti aus dem Haus\_und warf hastig die Vespa seines Sohnes an.[0060]  
Poco antes de las nueve, Laurenti *salió* de la casa\_ *a hurtadillas* y se apresuró a marcharse en la Vespa de su hijo.
- (8) [...] kaum dass er aus der Kirche *gestolpert* war. [0032]  
En cuanto *salió a trompicones* de la iglesia, [...]
- (9) *Salió disparado* del baño, con el corazón al galope. [0064]  
Alex *stürzte* aus dem Klo, sein Herz raste.
- (10) El inglés la recogió, dio las gracias y *salió corriendo*. [0008]  
Der Engländer ergriff ihn, bedankte sich und *hastete* hinaus.

Esta actividad puede resultar útil también para ampliar el vocabulario en alemán no sólo en el campo de los verbos de manera de movimiento, sino también de los verbos de trayectoria y de los adverbios direccionales, así como para analizar patrones de uso gramatical de los complementos direccionales. Como se puede deducir de esta propuesta, las concordancias acompañadas de un cierto contexto que pueden obtenerse de un corpus paralelo resultan muy adecuadas para el aprendizaje, especialmente en el caso de construcciones que divergen sensiblemente en cuanto a su uso y estructura de la lengua materna a la extranjera.

## 5. BALANCE FINAL

Los corpus paralelos ofrecen una oportunidad única para observar en contraste estructuras de la lengua materna y extranjera y sus aplicaciones en ELE van desde el uso indirecto en el contexto de la preparación de las clases (por ejemplo, la creación de materiales de aprendizaje basados en usos lingüísticos auténticos), hasta el uso

directo por parte del alumno. Sobre la base del concepto de *data-driven-learning*, se argumenta aquí a favor de su uso directo en ELE, siempre que el nivel de los estudiantes lo permita. El alumno desarrolla así la capacidad de «learn how to learn» (Johns, 1991, 1) al asumir el papel de explorador. El aprendiz, al usar directamente un corpus, se transforma en un investigador lingüístico que, por su propia cuenta, a la vista de los datos, llega a sus propias conclusiones que, finalmente, le ayudarán a ganar conocimiento sobre la gramática, vocabulario y construcciones específicas de la lengua extranjera. El objetivo es sensibilizar a los alumnos y permitirles derivar independientemente reglas de uso utilizando ejemplos auténticos. Los corpus paralelos también se pueden utilizar para ampliar el vocabulario y fomentar la producción de textos, ya que proporcionan un apoyo muy útil donde un simple diccionario bilingüe se queda corto, pues no puede ofrecer toda la gama de posibles contextos. Ahora bien, para que los corpus paralelos puedan utilizarse de manera efectiva en la enseñanza, deben adaptarse a las necesidades de los profesores y alumnos, y tener un funcionamiento de la interfaz de búsqueda lo más sencillo e intuitivo posible.

Mientras que los ejemplos de uso de un corpus paralelo resultan demasiado complicados para alumnos principiantes, el trabajo con los datos lingüísticos auténticos de un corpus paralelo es de especial relevancia para la enseñanza avanzada de lenguas extranjeras, así como para la formación universitaria de los futuros profesores de lenguas extranjeras.

Finalmente, voy a reproducir aquí las palabras de Stig Johansson (2009, 38), pionero en la creación de corpus paralelos y en su uso en el campo de la lingüística contrastiva y de la didáctica de lenguas extranjeras, que resume sus ventajas de la siguiente manera:

Given appropriate search tools, corpora are a way of enhancing and focusing the input to the student. They provide authentic data. They encourage reflection. They are well suited for consciousness-raising activities and for the training of inferencing. They stimulate the student to work actively and independently, and in this way they probably increase both the motivation of the student and the learning effect.

## 6. BIBLIOGRAFÍA

- AHRENHOLZ, Bernt y Franziska WALLNER. 2013. «Digitale Korpora und Deutsch als Fremdsprache». En *Deutsch als Fremdsprache* (Deutschunterricht in Theorie und Praxis, Bd. 10) ed. por Bernt Ahrenholz e Ingelore Oomen-Welke, Schneider Hohengehren, 261-272.
- AJMER, Karin (ed.). 2009. *Corpora and Language Teaching*. Amsterdam: John Benjamins.



- BERNARDINI, Silvia. 2004. «Corpora in the classroom». En *How to Use Corpora in Language Teaching*, ed. por John Sinclair. Amsterdam: John Benjamins, 15-36.
- BOULTON, Alex. 2017. «Research timeline: Corpora in language teaching and learning». *Language Teaching*, 50 (4): 483-506. DOI 10.1017/S0261444817000167
- BRAUN, Sabine y Angela CHAMBERS. 2006. «Elektronische Korpora als Ressource für den Fremdsprachenunterricht». En *Praktische Handreichung für Fremdsprachenlehrer*, ed. por Udo Jung, Frankfurt/M: Peter Lang, 330-337.
- BROWN, Michael. 2017. «Using Parallel Corpora for Language Learning». En *Humanising Language Teaching*, 19, 3. Fecha de acceso 10 de enero de 2019. [https://www.researchgate.net/publication/328094165\\_Using\\_Parallel\\_Corpora\\_for\\_Language\\_Learning/download](https://www.researchgate.net/publication/328094165_Using_Parallel_Corpora_for_Language_Learning/download)
- ČERMÁK, Petr. 2019. «InterCorp. Parallel corpus of 40 languages». En *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications*, ed. por Irene Doval y M. Teresa Sánchez. Amsterdam: John Benjamins, 93-102.
- CHRISOU, Marios. 2010. «Mit Textkorpora im Unterricht arbeiten und Sprachförderung initiieren». Fecha de acceso 20 de noviembre de 2018. [https://www.uni-due.de/imperia/md/content/prodaz/chrisou\\_korpusarbeit.pdf](https://www.uni-due.de/imperia/md/content/prodaz/chrisou_korpusarbeit.pdf)
- DOVAL, Irene. 2017a. «POS-tagging a bilingual parallel corpus: methods and challenges». *Research in Corpus Linguistics* 5: 35-46.
- DOVAL, Irene. 2017b. «La construcción de un corpus paralelo bilingüe multifuncional». *Moenia. Revista lucense de lingüística y literatura* 27: 125-141.
- DOVAL, Irene. 2018. «Das PaGeS-Korpus, ein Parallelkorpus der deutschen und spanischen Gegenwartssprache». *Revista de Filología Alemana* 26: 181-197.
- DOVAL, I, S. FERNÁNDEZ LANZA, T. JIMÉNEZ JULIÁ, E. LISTE LAMAS y B. LÜBKE. 2019. «Corpus PaGeS: A multifunctional resource for language learning, translation and cross-linguistic research». En *Parallel Corpora for Contrastive and Translation Studies: New Resources and Applications*, ed. por Irene Doval y M. Teresa Sánchez. Amsterdam: John Benjamins, 103-121.
- GUT, Ulrike. 2007. «Sprachkorpora im Phonetikunterricht». *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 12: 2). Fecha de acceso 10 de diciembre de 2018. <https://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/viewFile/308/299>
- HEID, Ulrich. 2008. «Corpus linguistics and lexicography». En *Corpus Linguistics. An International Handbook. Volume 1. Handbücher zur Sprach- und Kommunikationswissenschaft*, ed. por A. Lüdeling y M. Kyto, Berlin: Walter de Gruyter, 131-153.
- JOHANSSON, Stig. 2009. «Some thoughts on corpora and second-language acquisition». En *Corpora and language teaching*, ed. por Karin Aijmer. Amsterdam: John Benjamins, 33-44.
- JOHNS, Tim. 1991. «Should you be persuaded – two samples of data-driven learning materials». En *English Language Research Journal, Vol. 4: Classroom Concordancing*, ed. por T. Johns y P. King, Birmingham: The University of Birmingham, 1-16.
- LEMNITZER, Lothar y Heike ZINSMEISTER. 2015. *Korpuslinguistik*. Eine Einführung. Tübingen: Narr.
- LEVIN, Beth y Malka RAPPAPORT HOVAV. 1992. «The lexical semantics of verbs of motion: The perspective from unaccusativity». En *Thematic Structure: Its Role in Grammar*, ed. por I. Roca, Berlin: de Gruyter, 247-269.
- LÜDELING, Anke y Maik WALTER. 2010. «Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung». En *Handbuch Deutsch als Fremd- und Zweitsprache, HSK 35*, ed. por Hans-Jürgen Krumm et al. De Gruyter. Fecha de acceso

- 10 de diciembre de 2018. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiter-innen/anke/pdf/LuedelingWalterDaF.pdf>
- MCENERY, Tony y Richard XIAO. 2007. «Parallel and comparable corpora: What are they up to?». En *Incorporating Corpora: Translation and the Linguist*. ed. por G. James y G. Anderman. Clevedon: Multilingual Matters). Fecha de acceso 10 de diciembre de 2018. [http://someyanet.com/104-IT\\_Kansai\\_Initiative/corpora\\_and\\_translation.pdf](http://someyanet.com/104-IT_Kansai_Initiative/corpora_and_translation.pdf)
- MCENERY, Tony y Richard XIAO. 2011. «What corpora can offer in language teaching and learning». En *Handbook of research in second language teaching and learning*, 2, 364-380.
- MUKHERJEE, Joybrato. 2004. «Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany». En *Applied Corpus Linguistics. A Multidimensional Perspective*, ed. por Ulla Connor & Thomas Upton Amsterdam: Rodopi, 239-250.
- STEINBERGER Ralf. et al. 2014. «An overview of the European Union's highly multilingual parallel corpora». *Language Resources and Evaluation Journal (LRE)*. <DOI: 10.1007/s10579-014-9277-0.>
- TALMY, Leonard. 2000. *Toward a Cognitive Semantics. Volume 1: Concept Structuring Systems*. Cambridge: A Bradford Book.
- TIEDEMANN, Jörg. 2011. *Bitext Alignment*. Toronto: Morgan & Claypool.
- TIEDEMANN, Jörg. 2012. «Parallel Data, Tools and Interfaces in OPUS». En *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*. Fecha de acceso 10 de diciembre de 2018. [http://www.lrec-conf.org/proceedings/lrec2012/pdf/463\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf)
- TIEDEMANN, Jörg. 2016. «OPUS – Parallel Corpora for Everyone». En *Baltic Journal of Modern Computing (BJMC), Vol 4, n.º 2, Special Issue: Proceedings of the 19th Annual Conference of the European Association of Machine Translation (EAMT)*. Fecha de acceso 10 de diciembre de 2018. [https://www.bjmc.lu.lv/fileadmin/user\\_upload/lu\\_portal/projekti/bjmc/Contents/4\\_2\\_28\\_Products.pdf](https://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/4_2_28_Products.pdf)
- TSCHIRNER, Erwin. 2009. «Korpuslinguistik und Fremdsprachenunterricht». En *Germanica Wratislaviensia* 129, 117-132.
- WALLNER, Franziska. 2013. «Korpora im DaF-Unterricht – Potentiale und Perspektiven am Beispiel des DWDS». *Revista Nebrija de Lingüística Aplicada a la enseñanza de lenguas* 13. Fecha de acceso 10 de diciembre de 2018. <https://www.nebrija.com/revista-linguistica/korpora-im-daf-unterricht-potentiale-und-perspektiven-am-beispiel-des-dwds.html>

## AGRADECIMIENTO

El proyecto, base de este artículo, ha sido financiado por la Agencia Estatal de Investigación del Ministerio de Ciencia, Innovación e Universidades (FFI2017-85938-R, IP: Irene Doval). Ha recibido también financiación de la Consellería de Economía y Competitividad de la Xunta de Galicia a través del Grupo de Investigación LitLinAI (2017-PG023). Nuestro agradecimiento por su apoyo.