

<https://doi.org/10.32735/S0718-22012022000541030>

157-178

HACIA ESTUDIOS TRADUCTOLÓGICOS DEL CHINO CLÁSICO: ELABORACIÓN DE UN CORPUS PARALELO CHINO CLÁSICO-ESPAÑOL

*Towards the translation studies of classical Chinese: the elaboration of a classical
Chinese-Spanish parallel corpus*

WEI SUN

Universidad de Estudios Extranjeros de Beijing (China)

Universitat Autònoma de Barcelona (España)

felipe@bfsu.edu.cn

INNA KOZLOVA

Universitat Autònoma de Barcelona (España)

inna.kozlova@uab.cat

FULIANG CHANG

Universidad de Estudios Extranjeros de Beijing (China)

juanchang@163.com

Resumen

El corpus, siendo una herramienta eficaz, se está empleando en múltiples ámbitos, entre los que se encuentran los estudios de traducción. Sin embargo, dicho recurso no se utiliza con mucha frecuencia en las investigaciones traductológicas cuando una de las lenguas en cuestión es chino clásico (la lengua escrita empleada antiguamente en China), y autores que sí hacen estas investigaciones suelen no mencionar los pasos a seguir cuando se trata un corpus paralelo *ad hoc*. El objetivo del presente artículo es presentar cómo se elabora un corpus paralelo chino clásico-español tomando como ejemplo la novela histórica *Romance de los Tres Reinos* y sus traducciones al español. Para ello, hemos descrito la problemática de creación y explotación de este tipo de corpus, siguiendo el orden de las fases del proceso. Resulta que algunas fases del proceso son universales, mientras que otras son propias de lenguas clásicas y de la combinación lingüística chino-otra lengua, en cuyo caso sí faltan aplicaciones eficaces. Con el trabajo, esperamos deparar información a los estudiosos que quieran realizar y profundizar las investigaciones basadas en corpus paralelo y en las obras clásicas de China.

Palabras clave: Corpus paralelo; Estudios de traducción; *Romance de los Tres Reinos*; Chino clásico; Traducción chino-español

Abstract

The corpus, being an effective tool, is utilized in multiple fields. One of them is the Translation. However, this resource isn't used frequently in translation studies when one of the languages in question is classical chinese (written language used anciently in China). And when it

Recibido: 3 noviembre 2020

Aceptado: 7 abril 2021

comes to an *ad hoc* parallel corpus, the steps of how to create it are usually omitted by researchers who make such investigations. The objective of this article is to present how to elaborate a classic Chinese-Spanish parallel corpus taking as example the historic novel *Romance of the Three Kingdoms* and its two translations into Spanish. To reach the objective, we have described the problem of creation and exploitation of this type of corpus, following the order of the phases of the process. We've found that some of the phases are universal while other phases are unique for old languages and for the linguistic combination Chinese-other language. And in this latter case, there's a lack of effective applications. With the current work, we hope to offer useful information to those scholars who are willing to carry out their own investigations based on self-built parallel corpus and at the same time related to the classic Chinese works.

Key words: Parallel corpus; Translation studies; *Romance of the three kingdoms*; Classical chinese; chinese-spanish translation

1. INTRODUCCIÓN

El chino clásico puede definirse como la lengua escrita empleada antiguamente en China (Wang, 2006). Es de vital importancia, porque es la lengua en la que están escritas numerosas obras clásicas del país. Incluso en videojuegos populares derivados de dichas obras se puede hallar la presencia del chino clásico, tal como nos muestra la Figura 1.



Figura 1. Chino clásico en El Destino del Dragón: Los Tres Reinos (*傲世三国*, Aoshi Sanguo)

Estudiar el chino clásico y obras clásicas quizá sea una de las mejores maneras de conocer de cerca la cultura antigua de China. A esto han contribuido los estudiosos del ámbito de Traducción: por un lado, hay cada vez más traductores que intentan trasladar a otros idiomas las obras clásicas de China; por el otro lado, estudiosos vienen contrastando

desde diversos ángulos estas obras y sus traducciones a diferentes lenguas. Sin embargo, como el chino clásico difiere mucho del chino moderno utilizado hoy, traducir obras clásicas de China y emprender investigaciones acerca de su traducción supone un gran reto, y los trabajos al respecto llevados a cabo todavía son limitados. Para ello, una posible solución puede ser la elaboración del corpus paralelo, recurso que nos permite obtener información útil mediante análisis de ejemplos vivos.

Del término corpus paralelo, existe cierta discusión terminológica. Un corpus paralelo, según algunos expertos, está compuesto de textos originales de diferentes idiomas seleccionados por los mismos criterios (Schmied y Schäffler, 1996; Granger, 1996), mientras que otros consideran que un corpus paralelo consiste en un corpus de textos originales con su traducción a otras lenguas (Baker, 1995; McEnery y Oakes, 1996; Hunston, 2002; Wang, 2004; Doval, 2018). En el presente trabajo nos regiremos por la segunda definición, ya que “está más aceptada en el mundo académico en comparación con la primera” (Hu, 2011, p. 33, traducción nuestra).

“Un corpus se puede definir, sucintamente, como una herramienta muy fiable que permite validar de forma empírica teorías a partir de una muestra lingüística real y representativa” (Molés-Cases, 2016, p. 151). Frente a los demás tipos de corpus, la mayor ventaja del corpus paralelo consiste en que nos permite visualizar las relaciones correspondientes a nivel de léxico, de sintaxis y de texto entre dos o más lenguas. Observar dichas relaciones correspondientes nos permite realizar la comparación, haciendo posible establecer equivalencias contextuales. En este sentido, el corpus paralelo tiene bastante valor en cuanto a la lingüística contrastiva, a la lexicografía bilingüe, a la traducción automática, entre otros temas (Hu, 2011).

A pesar de que los estudios traductológicos basados en el corpus pueden remontarse a principios de los años noventa del siglo pasado (Kruger, Wallmach y Munday, 2011) y que son muchos los corpus paralelos ya existentes, dicha herramienta todavía se utiliza poco en lo que se refiere a las lenguas clásicas. En los contados trabajos en que sí se emplea, se encuentran las publicaciones de Cichosz (2012, 2013), que versan respecto de problemas de traducción entre el inglés clásico, el alto alemán antiguo y el latín. También destacan el trabajo de Eckhoff *et al.* (2018) y el de Haug y Jøhndal (2008), en los que hablan de la construcción del *Treebank* (corpus parseado) paralelo de las traducciones indoeuropeas más antiguas del Nuevo Testamento. Mediante dicho *Treebank*, intentan obtener un esquema de anotación semántica común que después se puede aplicar en la anotación de otros corpus de estos idiomas antiguos. Los autores afirman que el tratamiento de las lenguas antiguas requiere mucho la intervención humana.

Cuando la combinación lingüística del corpus bilingüe es chino clásico-otra lengua, el par lingüístico predominante es chino clásico-inglés. Por ejemplo, con el apoyo del corpus paralelo Liu (2013) ha comparado desde múltiples ángulos las cuatro

traducciones al inglés de *A la Orilla del Agua*; Zhao (2014) ha estudiado la prosodia semántica en la traducción de *Sueño en el Pabellón Rojo*; Fan (2017) se ha enfocado en las palabras culturales en las cinco traducciones al inglés de *Las Analectas*. En lo tocante al corpus paralelo chino clásico-español, los trabajos que lo tratan son aún menos abundantes. A continuación explicamos por qué.

En caso de estudiar la traducción de obras clásicas de China, muchas veces un investigador no dispone de un corpus utilizable y lo necesita componer para satisfacer sus necesidades investigadoras. Para ello, es necesario saber la manera óptima de operar en cada fase implicada durante la creación. Sin embargo, los contados autores que han mencionado la elaboración de un corpus paralelo chino clásico-otra lengua en su trabajo tienden a no explicar con detalle el procedimiento. Como resultado, muchos investigadores que quieran emplear este tipo de corpus en su trabajo quedan sin saber qué hacer.

Es cierto que existen unas fases universales para la creación de un corpus paralelo, por ejemplo, recopilación de datos, almacenamiento, conversión de formatos y alineación de los textos (véase Castillo Rodríguez, 2009; González Rey, 2014). No obstante, la composición de un corpus chino clásico-otra lengua presenta ciertas particularidades. Es decir, no todas las pautas a seguir para construir un corpus paralelo inglés-español, por ejemplo, son reproducibles en la elaboración de un corpus bilingüe chino-español, por no mencionar la creación de un corpus chino clásico-español. La razón principal consiste en que el chino clásico es poco estudiado. Es un idioma analítico donde las palabras tienden a ser monoformáticas y que no existen fronteras claras entre sí. Como consecuencia, merece tratamientos especiales.

Por todo ello, en el presente trabajo nos parece oportuno presentar cómo se elabora un corpus paralelo chino clásico-español, tomando *Romance de los Tres Reinos* y sus traducciones al español como ejemplo. Nos enfocaremos en los procesos a seguir y los problemas afrontados en cada paso. Esperamos que el procedimiento que proponemos pueda proporcionar información útil a los que quieran realizar y profundizar las investigaciones concernientes a *Romance de los Tres Reinos* u otras obras clásicas de China.

Es oportuno mencionar mencionar que en este trabajo abundan palabras de origen chino. Acerca de su uso hemos seguido las normas propuestas por Casas-Tost, Fustegueres i Rosich, Xianghong Qu, Rovira-Esteva y Vargas-Urpi (2015).

2. RECOPIACIÓN DE *ROMANCE DE LOS TRES REINOS* Y SUS TRADUCCIONES AL ESPAÑOL

2.1. SELECCIÓN DE TEXTOS

La novela histórica *Romance de los Tres Reinos* (suele ser llamada *Los Tres Reinos*) es la primera narrativa larga con división clara entre capítulos de la época antigua de China. Narra los acontecimientos registrados entre el año 184 y el año 280 d.C., es decir, desde la rebelión de los Turbantes Amarillos hasta la reunificación de China por el emperador Wudi de la dinastía Jin.

Dicha obra está redactada en un lenguaje mixto entre el chino clásico y el chino moderno. Siendo una de las cuatro obras clásicas más populares y trascendentales del país, no solo ocupa un puesto preponderante en la literatura china, sino que también ejerce mucha influencia en todo el mundo. Merece ser estudiada desde muchos ángulos, entre los que se encuentra el ámbito de Traducción. Por consiguiente, la hemos seleccionado como ejemplo en el presente trabajo.

Para la elaboración del corpus paralelo arriba mencionado, antes que todo precisa recopilar tanto el texto original de la obra como sus traducciones al español.

Con el objetivo de encontrar las traducciones de la obra al castellano, hemos realizado un barrido de los principales motores de búsqueda, los comercios electrónicos de importancia y también la base de datos “La literatura china traducida en España” (Rovira-Esteva, Casas-Tost, Tor-Carroggio y Vargas-Urpi, 2019). Esta última es una herramienta de acceso abierto que permite a todos los interesados buscar información (título, autor, ISBN, año, traductor y editorial, entre otras cosas) relacionada con las obras escritas originalmente en chino publicadas en España.

Como resultado, hemos encontrado tres versiones de traducción de la obra *Romance de los Tres Reinos* al español: la primera fue hecha por María Teresa Ortega y Olga Marta Pérez, dos traductoras consagradas de Cuba; la segunda fue realizada por Ariel Allevi; y la última, hasta la fecha, es una traducción no completa (solo se han traducido 59 de los 120 capítulos), por Ricardo Cebrián Salé, escritor y traductor español y también aficionado a la literatura asiática.

Tras estudiar las traducciones, hemos decidido utilizar la primera y la tercera en el presente trabajo, ya que se puede hallar suficiente información de su elaboración, la editorial por ejemplo, la que en cierto sentido garantiza la calidad de la traducción; la segunda, en cambio, no dispone de semejantes datos, por tanto, hemos decidido no incluirla en nuestro corpus paralelo.

En lo que se refiere a *Romance de los Tres Reinos* en chino, también existe una cantidad considerable de versiones. La que hemos elegido como texto original es la versión con comentarios de Mao Zonggang. La razón de esta decisión es sencilla: por un lado, es una versión bien divulgada y que goza de una gran popularidad; por el otro lado, es la base

sobre la cual se han realizado las versiones al español que hemos presentado antes. En este sentido, nuestra opción asegura una mayor vinculación entre el texto original y las traducciones, y reducirá la posibilidad de que surjan problemas, entre otros, discrepancias y falta de paralelismo, en las siguientes fases de la elaboración del corpus.

2.2. OBTENCIÓN DE TEXTOS

La versión original de *Romance de los Tres Reinos* en chino arriba mencionada puede adquirirse por medio de internet en formato “txt”. La traducción realizada por Ricardo Cebrián Salé no es difícil de conseguir, porque puede obtenerse en <https://www.amazon.es/>. No obstante, la obtención de la traducción realizada por María Teresa Ortega y Olga Marta Pérez en formato digital presenta mayor dificultad: no tenemos posibilidad de acceder al texto en este formato. Como solución, hemos adquirido el libro impreso (seis tomos en total), hemos escaneado el texto en español (alrededor de 1.800 páginas) guardándolo como PDF (compuesto por imágenes) en el ordenador y lo hemos convertido en formato editable mediante *software* de OCR. Para que el derecho de autor quede sin ser violado, hemos conseguido el consentimiento de Olga Marta Pérez de que la traducción (con la información de autor y fuente indicada) la utilizaremos solo y exclusivamente para investigaciones personales sin fines de lucro.

El *software* de OCR que utilizamos es ABBYY FineReader, que nos permite reconocer todos los caracteres de una imagen, extraerlos, copiarlos y trabajar con ellos como si fueran texto plano. Dicha aplicación es compatible con más de 190 idiomas y una de sus ventajas más destacadas consiste en que tiene una tasa de acierto elevada, en especial cuando lo usamos para tratar letras latinas. Aquí abajo está expuesta una captura de pantalla del interfaz de dicha aplicación (Figura 2).

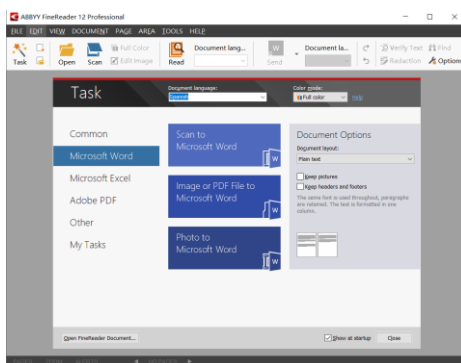


Figura 2. Interfaz del *software* OCR ABBYY FineReader

Podemos elegir la tarea que realizar, indicando las lenguas implicadas y el tipo de formato que queremos obtener como resultado. En nuestro caso, necesitamos elegir “Imagen o Archivo de PDF a Microsoft Word” y el idioma involucrado es el español. Después, en la ventanilla que se abre seleccionamos los archivos pendientes de procesar y la aplicación comenzará la conversión de manera automática.

Finalizado dicho proceso, guardamos los textos obtenidos en tres archivos distintos en formato Microsoft Word para poder proceder con el siguiente paso: tratamiento previo del texto.

3. TRATAMIENTO PREVIO DEL TEXTO

Para construir cualquier corpus paralelo, el tratamiento previo del texto es de vital importancia. La razón consiste en que los textos obtenidos, sean descargados del internet, introducidos en forma manual o frutos de la conversión, no son impecables. Puede haber problemas de diferentes tipos:

- Los formatos pueden no ser uniformes.
- Pueden hallarse signos, imágenes o espacios innecesarios.
- Los caracteres o letras pueden estar equivocados.
- Los acentos pueden estar mal marcados.

Por tanto, será necesario tomar medidas para conseguir un texto “limpio”, lo que en gran medida garantizará una búsqueda de calidad y así contribuirá a la precisión y objetividad del resultado final.

En cuanto al tratamiento previo que se necesita llevar a cabo en nuestro caso, por un lado, es comparar el texto en chino que hemos conseguido por internet con el *Romance de los Tres Reinos* (2012) en chino publicado por Ediciones en Lenguas Extranjeras (外文出版社, *Waiwen Chubanshe*), y en caso de hallar diferencias (omisión de párrafos y errores de caracteres, entre otras), realizamos modificaciones a base de este último.

Aparte de eso, necesitamos ajustar el formato: dejamos que cada frase ocupe una línea. En el presente caso, respecto de la detección del límite oracional nos regimos por el criterio propuesto por Haug y Jøhndal (2008). Es decir, damos por entender que una frase es la que acaba con los signos de puntuación “。”, “!” y “?”. Dicho ajuste se puede realizar utilizando la función “Buscar” y “Reemplazar” de Word y facilitará la futura alineación. Las operaciones concretas para realizar pueden verse en la Tabla 1 expuesta abajo:

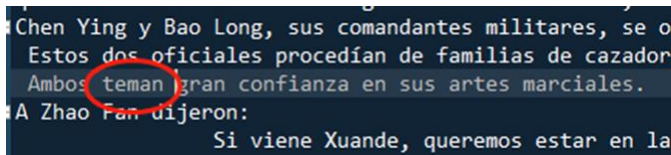
Tabla 1. Operaciones con la función “Buscar” y “Reemplazar” de Word

Paso	Operación		Ejemplo del resultado de la operación	
	Signo original	Sustitución	Antes	Después
1	。 (。 es punto en chino)	。 ^p (^p representa la marca de párrafo)	五原山岸，尽皆崩裂。种种不祥，非止一端。	五原山岸，尽皆崩裂。 种种不祥，非止一端。
2	!	! ^p	害民贼！认得我么？	害民贼！ 认得我么？
3	?	? ^p	朝廷滋蔓极广，安能尽诛？倘机不密，必有灭族之祸：请细详之。	朝廷滋蔓极广，安能尽诛？ 倘机不密，必有灭族之祸：请细详之。
4	。 ^p”	。 ”^p	重用臣等：大事可图矣。 ”董太后大喜。	重用臣等：大事可图矣。” 董太后大喜。
5	! ^p”	! ”^p	进曰：“此计大妙！” ”便发檄至各镇，召赴京师。	进曰：“此计大妙！” ”便发檄至各镇，召赴京师。
6	? ^p	? ”^p	庄主问曰：“二少年谁家之子？” ”帝不敢应。	庄主问曰：“二少年谁家之子？” ”帝不敢应。

Más concretamente, primero sustituimos “。” por “。 ^p”, “!” por “! ^p” y “?” por “? ^p”. Eso quiere decir que después de cada “。”, “!” y “?” iniciamos un nuevo párrafo. Hecho esto, tenemos que planeamos si hay excepciones que deberían haberse eximido de dicha operación: hay casos en que “。” está seguido de “” para indicar la finalización de un estilo directo, por lo tanto “。” y “” nunca deberían estar separados. Lo mismo ocurre con “!” y “?”. Así que necesitamos reemplazar “。 ^p” por “。 ”^p”, “! ^p” por “! ”^p” y “? ^p” por “? ”^p”.

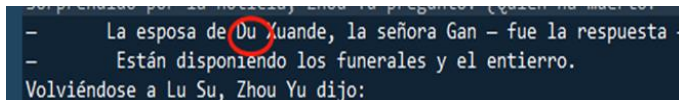
También nos hace falta examinar con detenimiento la traducción de María Teresa Ortega y Olga Marta Pérez. Además de ajustar el formato y quitar los elementos paratextuales (signos, gráficos, etc.), nuestro objetivo será remendar los errores generados en el proceso de conversión. A pesar de que el *software* ABBYY FineReader tiene un rendimiento relativamente bueno en lo que se refiere a la tasa de acierto, de ninguna manera es perfecto. A esto si se suma la calidad del escaneo, es natural encontrar casos donde ha habido reconocimiento equivocado de letras y símbolos, tal como muestran las siguientes figuras (3, 4, 5, 6, 7 y 8):

Figura 3. Reconocimiento equivocado de letras (“Kong Kong” debería ser “Kong Rong”)



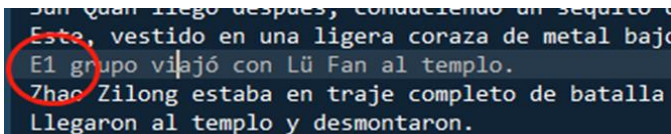
Chen Ying y Bao Long, sus comandantes militares, se o
Estos dos oficiales procedían de familias de cazador
Ambos **teman** gran confianza en sus artes marciales.
A Zhao Fan dijeron:
Si viene Xuande, queremos estar en la

Figura 4. Reconocimiento equivocado de letras (“teman” debería ser “tenían”)



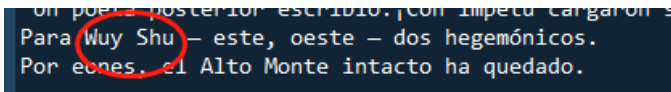
La esposa de **Du Xuande**, la señora Gan – fue la respuesta –
Están disponiendo los funerales y el entierro.
Volviéndose a Lu Su, Zhou Yu dijo:

Figura 5. Reconocimiento equivocado de letras (“Du Xuande” debería ser “Liu Xuande”)



San Juan luego después, conduciendo un séquito
Este, vestido en una ligera coraza de metal bajo
El grupo viajó con Lü Fan al templo.
Zhao Zilong estaba en traje completo de batalla
Llegaron al templo y desmontaron.

Figura 6. Reconocimiento equivocado de símbolos (“El grupo” debería ser “El grupo”)



Para **Wuy Shu** – este, oeste – dos hegemónicos.
Por eso, el Alto Monte intacto ha quedado.

Figura 7. Reconocimiento equivocado de letras (“Wuy Shu” debería ser “Wu y Shu”)



los soldados de **Kongrning** se burlaban
s hombres: “¡Probemos por última vez!”

Figura 8. Reconocimiento equivocado de letras (“Kongrning” debería ser “Kongming”)

Frente a este problema, puede ser necesario pasar el texto a un revisor humano. Si lo hacemos nosotros, lograríamos no solo “limpiar” el texto sino también familiarizarnos con la traducción. Respecto de la traducción de Ricardo Cebrián Salé, no hay muchas

cosas que hacer, porque se trata de una publicación oficial, y ha sido revisada por el mismo traductor. Nos basta con ajustar el formato.

Por último, necesitamos guardar los textos procesados. Hay dos puntos a que tenemos que prestar especial atención: la denominación de los textos y el formato en que están guardados. Para facilitar la identificación, es necesario seguir un determinado patrón (Kozlova y Rodríguez-Inés, 2017, p. 215).

1. Como *Romance de los Tres Reinos* es una obra larga (con alrededor de 640.000 caracteres chinos distribuidos en 120 capítulos), no nos parece sensato guardar el texto original entero o una traducción entera como un solo archivo, porque dificultaría la alineación y la búsqueda. Por consiguiente, hemos dividido el texto original en 120 archivos separados, y cada capítulo corresponde a un archivo. Para que los archivos estén más ordenados, los denominamos CH001, CH002, CH003...CH120, respectivamente (“CH” hace referencia al idioma chino, y el número que le sigue representa el capítulo correspondiente), y los guardamos en una carpeta. Hacemos lo mismo con las dos versiones españolas.

2. Todos los archivos deben ser guardados en el mismo formato que sería adecuado para las aplicaciones que uno pretende emplear más adelante para realizar la alineación y la consulta. Si no, dichos archivos podrían no visualizarse en forma correcta. En nuestro caso, los hemos guardado en “txt” con codificación “UTF-8”.

Así que, por último, tenemos tres carpetas con sendos archivos, tal como muestra la Figura 9 de abajo:

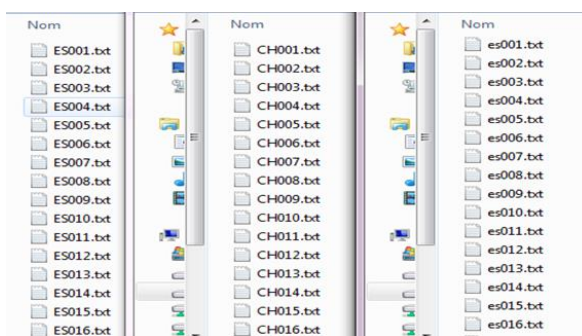


Figura 9. Carpetas y archivos

Con esto, se da por terminado el tratamiento previo del texto y se puede proceder con su segmentación y etiquetación.

4. SEGMENTACIÓN

Según Hu (2011), en español las palabras suelen considerarse como unidades mínimas de la lengua, entre las que se mantiene siempre espacio de intervalo, mientras que en chino son los caracteres que suelen ser unidades mínimas de la lengua, entre los que no hay espacio de intervalo alguno. A causa de esta discrepancia entre los dos idiomas, programas dedicados al estudio del corpus suelen no reconocer los textos crudos en chino. Por tanto, en caso de que las aplicaciones lo exijan, hará falta realizar la segmentación entre los caracteres chinos.

Hay un número considerable de aplicaciones para la segmentación, entre otras, NLPiR, HanLP, El tartamudo (结巴分词, Jieba Fenci), El monigote de nieve (雪人分词, Xueren Fenci) y Segmentador Pangu (盘古分词, Pangu Fenci). Quizá la más destacada sea el *software* ICTCLAS, desarrollado por el Instituto de Tecnología de Computación de la Academia de Ciencias de China, que ofrece no solo el servicio de segmentación sino también la de etiquetado gramatical. La tasa de acierto puede alcanzar al 97% (Hu, 2011). Sin embargo, todos estos programas son para el chino moderno y no para el chino clásico.

Visto el buen rendimiento que tiene ICTCLAS, hemos intentado llevar a cabo la segmentación del texto original de *Romance de los Tres Reinos* (tomamos el primer capítulo como ejemplo). No obstante, el experimento no ha sido tan fructífero como esperábamos y el resultado (véase la Figura 10) evidencia la afirmación arriba mencionada:

第一/m 回/v
 宴/ng 桃园/n 豪杰/n 三/n 结/v 义/ng
 斩/v 黄/nr1 巾/g 英雄/n 首/m 立功/vi

话说/v 天下/n 大势/n ... /wd 分/v 久/ng 必/d 合/v , /wd 合/v 久/ng 必/d 分/v , /wj
 周末/t 七/m 国/v 分/qt 争/v /wd 并/v 入/v 于/p 秦/tg ... /wj
 及/cc 秦/tg 灭/v 之/v 后/t , /wd 楚/tg , /wm 汉/g 分/v 争/v /wd 又/d 并/v 入/v 于/p 汉/tg , /wj
 汉朝/t 自/p 高祖/n 斩/v 白/nr1 蛇/m 而/cc 起义/vm , /wd 一统天下/v , /wd 后来, 光/n 武/n 中/nz 兴/v 帝/ng /wd 建/d 分
 为/v 三/m 国/n , /wj

Figura 10. Resultado de la segmentación utilizando El Asistente de la Segmentación de Chino (汉语分词小助手, Hanyu Fenci Xiaozhushou) (Programa derivado de ICTCLAS aplicable a Windows)

La Figura 10 nos demuestra que sí que existen bastantes aciertos en la segmentación del texto original, en esta secuencia concreta, 74%. Pero hay muchas situaciones donde la segmentación falla, tal como indican los círculos. Son casos que requieren ajustes manuales. Un motivo puede ser la abundancia de topónimos y antropónimos que se utilizan con poca frecuencia. Por ejemplo, “光武” es el nombre del reinado de un emperador, por tanto los dos caracteres nunca deben entenderse separados.

Otro motivo puede residir en el tipo de lenguaje que utiliza *Romance de los Tres Reinos*, ya que es un lenguaje mixto entre el chino clásico y el chino mandarín, y que dista mucho del chino moderno que se usa hoy.

5. ANOTACIÓN

Un siguiente paso para preparar el texto para su posterior explotación mediante herramientas de corpus sería su anotación, la que se entiende como la representación de la información implícita de un corpus textual. “Como recurso de datos, las ventajas de un texto anotado en comparación con uno crudo son evidentes, y la utilización de principios idénticos en la anotación garantiza la reusabilidad de los datos y su valor para la comunidad de investigación” (Mikhailov y Cooper, 2016, p. 34). Según la función del corpus, la anotación se puede realizar a diversos niveles, entre otros, el nivel fonético, el sintáctico, el gramatical y el pragmático.

Si nos interesa realizar una investigación pertinente a la traducción de las palabras funcionales empleadas en *Romance de los Tres Reinos*, por ejemplo, un tipo de anotación útil sería el etiquetado gramatical. Este, a diferencia de la segmentación, sería aconsejable ser aplicado tanto al chino como al español y nos ayudaría a comparar el texto original y sus traducciones a nivel léxico. “El hecho de que el chino clásico no tenga morfología flexiva hace aún más informativo el etiquetado gramatical” (Lee, 2012, p. 77, traducción nuestra).

El etiquetado gramatical del chino clásico también se puede llevar a cabo mediante El Asistente de la Segmentación de Chino que hemos mencionado en el apartado anterior. Sin embargo, dicha anotación depende mucho del resultado de la segmentación. En otras palabras, si la segmentación no da resultados satisfactorios, el etiquetado gramatical tampoco es correcto y necesita la intervención humana. El experimento realizado en la sección anterior puede servir de prueba (véase la Figura 11).

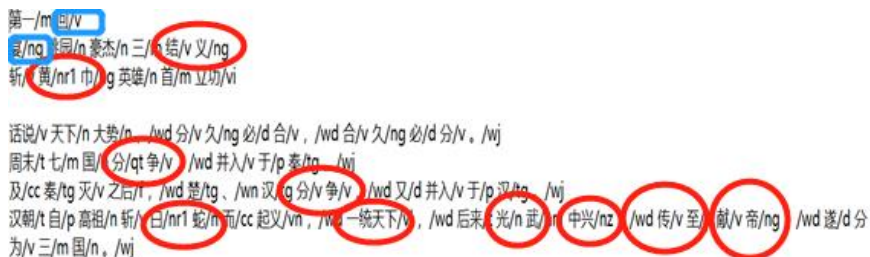


Figura 11. Resultado del etiquetado gramatical utilizando El Asistente de la Segmentación de Chino (*汉语分词小助手, Hanyu Fenci Xiaozhushou*)

Como hemos explicado, los círculos señalan los fallos de la segmentación. También son casos que requieren corrección en lo tocante al etiquetado gramatical. Por ejemplo, “*肅宗*” es el nombre de un emperador, por lo consiguiente los dos caracteres deben comprenderse como un conjunto, que llevaría la etiqueta “/n”.

Aparte de las inadecuaciones causadas por la segmentación, también hay situaciones esporádicas donde la aplicación falla solo en el etiquetado gramatical, tal como nos indican los rectangulares azules. Por ejemplo, “*宴*” en el texto no es un sustantivo que quiere decir “banquete”, sino un verbo que significa “celebrar un banquete”. Así que debería llevar la etiqueta “/v”. Eso se debe a que “más de la mitad de las palabras tienen dos o más posibles categorías léxicas y los cambios dinámicos de categorías léxicas son los fenómenos gramaticales más comunes en chino clásico” (Huang, Peng, Wang y Wu, 2002, p. 2, traducción nuestra). Por tanto, resulta difícil determinar en forma automática la categoría gramatical de una palabra.

En resumen, el etiquetado gramatical del chino clásico suele exigir la intervención humana, ya que las aplicaciones existentes para el tratamiento de chino moderno no presentan un comportamiento satisfactorio al respecto y que “lamentablemente hasta la fecha, todavía no ha visto la luz ninguna herramienta dedicada al etiquetado gramatical del chino clásico” (Liu y Shen, 2013, p. 37, traducción nuestra).

En cambio, el etiquetado gramatical del texto español, si precisa, puede efectuarse por medio del *software* Tree Tagger de versión Windows, desarrollado por Helmut Schmid de la Universidad de Stuttgart. Un extracto de la traducción “cruda” y la misma anotada, en este caso con etiquetado gramatical, se puede comparar observando las Figuras 12 y 13 expuestas abajo (tomamos como ejemplo el Capítulo 1 de la traducción realizada por María Teresa Ortega y Olga Marta Pérez):

CAPÍTULO 1

Tres espíritus osados se juran fe mutua en el Jardín de los Melocotoneros;

Héroes y paladines ganan los primeros honores en combate contra los Turbantes Amarillos.

<No fuente>Aquí comienza nuestra historia.

El imperio, largo tiempo dividido, deberá unificarse; el imperio, largo tiempo unido, deberá fragmentarse. Así ha sucedido siempre.

En los años finales de la dinastía Zhou, siete reinos lucharon entre sí y sólo el reino de Qin prevaleció sobre los demás y se adueñó de los otros seis.

Pero pronto el reino de Qin se desmoronó y sobre sus ruinas combatieron dos reinos rivales, Chu y Han, por la hegemonía. El reino Han salió victorioso y absorbió a su enemigo, como antes había hecho el de Qin.

El Antepasado Supremo mató la serpiente blanca y esta fue la señal que inspiró un levantamiento que culminó con la creación de un imperio unificado bajo el dominio de los Han. Doscientos años más tarde, después de la usurpación de Wang Mang, el emperador Guang Wu restauró la dinastía y los emperadores Han gobernaron durante otros doscientos años hasta el reinado de Xian, después el reino se dividió en tres partes.

Es posible hallar la causa de la caída de los Han durante los reinados de Huan y Ling, los predecesores de Xian.

Huan echó de sus cargos y persiguió a funcionarios íntegros y hábiles, mientras ponía toda su confianza en sus eunucos.

Luego cuando Ling sucedió a Huan en el trono imperial, el regente mariscal Dou Wu y el guardián imperial Chen Fan, fieles al trono,

se confabularon para ajusticiar al eunuco Cao Jie y sus cohortes, para así poner fin a sus abusos de poder. Pero el plan fue descubierto, entonces Dou Wu y Chen Fan fueron ejecutados.

Figura 12. La traducción de *Romance de los Tres Reinos*

CAPÍTULO NC 1 CARD Tres CARD espiritus NC osados VLadj se SE juran VLfin fe NC mutua ADJ en PREP el ART Jardín NP de PREP los ART Melocotoneros NC ; SEMICOLON héroes NC y CC paladines NC ganen VLfin los ART primeros ORD honores NC en PREP combate NC contra PREP los ART Turbantes ADJ Amarillos NC . FS <No fuente> Aquí ADV comienza VLfin nuestra PPO historia NC . FS El ART imperio NC CM largo ADJ tiempo NC dividido VLadj CM deberá VMfin unificarse VLinf ; SEMICOLON el ART imperio NC CM largo ADJ tiempo NC unido ADJ CM deberá VMfin fragmentarse VLinf . FS Así CSUBX ha VHfin sucedido VLadj siempre ADV . FS En PREP los ART años NC finales ADJ de PREP la ART dinastía NC Zhou NP CM siete CARD reinos NC lucharon VLfin entre PREP sí PPX y CC sólo ADV el ART reino NC de PREP Qin NC prevaleció VLfin sobre PREP los ART demás QU y CC se SE adueñó VLfin de PREP los ART otros QU seis CARD . FS Pero CCAD pronto ADV el ART reino NC de PREP Qin NC se SE desmoronó VLfin y CC sobre PREP sus PPO ruinas VLfin combatieron VLfin dos CARD reinos NC rivales NC CM Chu NP y CC Han NP CM por PREP la ART hegemonía NC . FS El ART reino NC Han NP salió VLfin victorioso ADJ y CC absorbió VLfin a PREP su PPO enemigo NC CM como CSUBX antes ADV había VHfin hecho VLadj el ART de PREP Qin NC . FS El ART Antepasado NC Supremo NC mató VLfin la ART serpiente NC blanca ADJ y CC esta DM fue VSfin la ART señal NC que COQE inspiró VLfin un ART levantamiento NC que COQE culminó VLfin con PREP la ART creación NC de PREP un ART imperio NC unificado VLadj bajo PREP el ART dominio NC de PREP los ART Han NP . FS Doscientos CARD años NC más ADV tarde ADV CM después ADV de PREP la ART usurpación NC de PREP Wang NP Mang NP CM el ART emperador NC Guang NP Wu NP restauró VLfin la ART dinastía NC y CC los ART emperadores NC Han NP gobernaron VLfin durante PREP otros QU doscientos CARD años NC hasta PREP el ART reinado NC de PREP Xian NP CM después ADV el ART reino NC se SE dividió VLfin en PREP tres CARD partes NC . FS Es VSfin posible ADJ hallar VLfin la ART causa NC de PREP la ART caída NC de PREP los ART Han NP durante PREP los ART reinados VLadj de PREP Huan NP y CC Ling NP CM los ART predecesores NC de PREP Xian NP . FS Huan NP echó VLfin de PREP sus PPO cargos NC y CC persiguió VLfin a PREP funcionarios NC íntegros ADJ y CC hábiles ADJ CM mientras CSUBX ponía VLfin toda QU su PPO confianza NC en PREP sus PPO eunucos NC . FS Luego CSUBF cuando CSUBX Ling NP sucedió VLfin a PREP Huan NP en PREP el ART trono NC imperial ADJ CM el ART regente VLfin mariscal NC Dou NP Wu NP y CC el ART guardián VLfin imperial ADJ Chen NP Fan NP CM fieles ADJ al PAL trono NC CM se SE confabularon VLfin para CSUBI ajusticiar VLfin al PAL eunuco NC Cao NP Jie NP y CC sus PPO cohortes NC CM para PREP así ADV poner VLfin fin NC a PREP sus PPO abusos NC de PREP poder VMfin . FS Pero CCAD el ART plan NC fue VSfin descubierto VLadj CM entonces ADV Dou NP Wu NP y CC Chen NP Fan NP fueron VSfin ejecutados VLadj . FS

Figura 13. Resultado del etiquetado gramatical utilizando Tree Tagger

De la Figura 12 y la Figura 13, se puede observar que, a rasgos generales, el resultado es satisfactorio y los errores solo se producen cuando se trata de nombres propios. Otro punto mejorable consiste en que el formato que tenía el archivo no se mantiene y que requiere reajustes del formato.

6. ALINEACIÓN DEL TEXTO

Para poder construir un corpus paralelo, la alineación del texto constituye un paso núcleo (McEnery, Xiao y Tono, 2006). La alineación puede efectuarse a nivel de párrafo y de documento, pero también se lleva a cabo a nivel oracional y léxico (Alonso *et al.*, 2001). Mediante la alineación, podemos contemplar de manera clara las similitudes y las diferencias entre el texto original y el texto de traducción. Desde otro punto de vista, “usuarios del corpus al fin y al cabo necesitarán consultar la misma oración en diferentes lenguas, lo que hace necesario la alineación oracional” (Haug y Jøhndal, 2008, p. 28, traducción nuestra). Por todo ello, en el presente trabajo efectuaremos la alineación oracional.

La alineación se puede realizar en forma manual o semiautomática con la ayuda de aplicaciones, entre otras, SDL Trados WinAlign, Tmxmall y ABBYY Aligner. Decimos “semiautomáticamente” y no “automáticamente”, porque al igual que los programas de OCR, estas aplicaciones no son impecables y tendremos que intervenir para realizar la posesición: revisar el resultado y modificarlo en casos necesarios. De todas formas, estas aplicaciones sí que pueden facilitar la alineación, que suele ser un trabajo duro que consume mucho tiempo y esfuerzo.

Hemos intentado acudir a este tipo de herramientas para llevar a cabo la alineación del presente trabajo, no obstante, el resultado de modo alguno ha sido satisfactorio. Eso lo podemos contemplar en la siguiente Figura 14:

1	第一回	CAPITULO 1
2	赛桃园豪杰三结义	Tres espíritus osados se juran fe mutua en el Jardín de los Melocotoneros;
3	斩黄巾英雄首立功	Héroes y paladines ganan los primeros honores en combate contra los Turbantes Amarillos.
4	话说天下大势，分久必合，合久必分。	Aquí comienza nuestra historia.
5	周末七国分争，并入于秦。	El imperio, largo tiempo dividido, deberá unificarse; el imperio, largo tiempo unido, deberá fragmentarse.
6	及秦灭之后，楚、汉分争，又并入于汉。	Así ha sucedido siempre.
7	汉朝自高祖斩白蛇而起义，一统天下，后来光武中兴，传至献帝，遂分为三国。	En los años finales de la dinastía Zhou, siete reinos lucharon entre sí y sólo el reino de Qin prevaleció sobre los demás y se adueñó de los otros seis.
8	推其致乱之由，殆始于桓、灵二帝。	Pero pronto el reino de Qin se desmoronó y sobre sus ruinas combatieron dos reinos rivales, Chu y Han, por la hegemonía. El reino Han salió victorioso y absorbió a su enemigo, como antes había hecho el de Qin.
9	桓帝禁锢善类，崇信宦官。	El Antepasado Supremo mató la serpiente blanca y esta fue la señal que inspiró un levantamiento que culminó con la creación de un imperio unificado bajo el dominio de los Han. Doscientos años más tarde, después de la usurpación de Wang Mang, el emperador Guang Wu restauró la dinastía y los emperadores Han gobernaron durante otros doscientos años hasta el reinado de Xian, después el reino se dividió en tres partes.

Figura 14. Prueba de alineación con Tmxmml

De la Figura 14 podemos observar que el producto que obtenemos después de la alineación semiautomática no nos sirve, ya que en la gran mayoría de las líneas el texto en chino y en español no coinciden. En realidad, de las 9 líneas mostradas, solo las primeras tres en rectángulo dan resultado positivo. Y tal como indican las flechas, desde la línea 4, el texto original y su correspondiente traducción en español están en diferentes líneas. Eso en cierto sentido corrobora la afirmación de Liu (2010), quien considera que todavía no existen programas que tengan un buen rendimiento en lo que se refiere a la alineación automática del chino clásico y sus traducciones.

Creemos que la razón principal del bajo porcentaje de aciertos consiste en que el chino y el español en general son dos idiomas lejanos, lo que se puede reflejar en la distancia entre sus estructuras gramaticales. Por tanto, en la alineación no es fácil encontrar con exactitud la traducción al español de una frase en chino. Además, las diferencias que existen entre el lenguaje utilizado en *Romance de los Tres Reinos* y el chino moderno conducen a una dificultad aún más grande al respecto.

Como la alineación semiautomática no es de gran utilidad para nosotros, no tenemos otro remedio que emplear el método más rudimentario pero eficaz: alinear los textos manualmente. Para ello, utilizamos el *software* EditPlus, programa de edición de textos para Windows. Para alinear los textos, lo que necesitamos hacer primero es poner el texto original en el centro y las dos traducciones en español a los dos lados, como lo que podemos ver en la Figura 15 expuesta abajo:



Figura 15. Los textos crudos antes de la alineación

Luego, realizamos la alineación manualmente. Como hemos mencionado antes, cada frase del texto original ya ocupa una línea. Estas frases ordenadas nos sirven como referencia, sobre la base de esto ajustamos las traducciones al español, segmentando o uniendo las expresiones. Este proceso no implica muchas operaciones, pero requiere atención suprema, porque un error en una línea provocaría el desorden en las oraciones restantes. En este sentido, el número que nos ofrece EditPlus al principio de cada línea es de gran utilidad, ya que nos ayuda a comprobar si estamos haciendo la alineación de manera correcta.

Durante el proceso, es posible que nos encontremos con dos situaciones especiales:

1. Cuando una frase del texto original no está traducida, ponemos una etiqueta <No traducción> en la línea correspondiente, como lo que hemos hecho en la siguiente Figura 16:

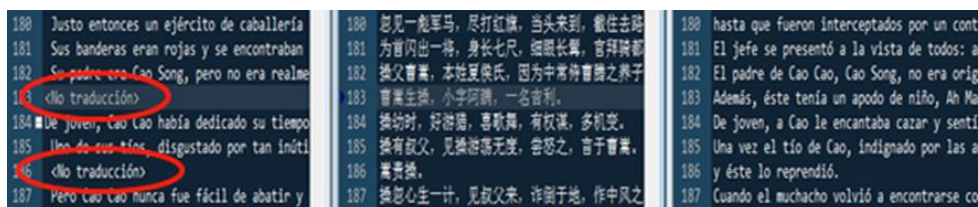


Figura 16. Etiqueta de <No traducción>

2. Cuando hallamos una frase en la traducción, pero no podemos encontrar en el texto original la oración que le corresponde, ponemos una etiqueta <No fuente> al comienzo de la frase y añadimos una línea de espacio en el texto original (también agregamos una en la otra traducción si procede), como lo que hemos hecho en la Figura 17 de abajo:

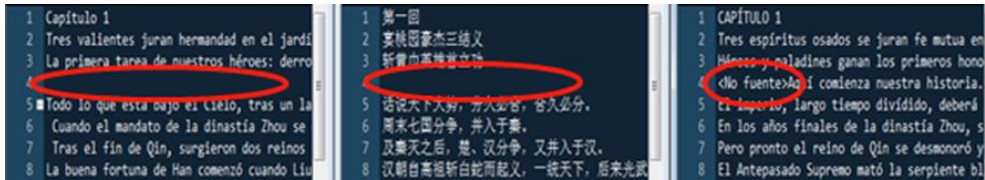


Figura 17. Etiqueta de <No fuente>

Tras alinear todas las líneas, efectuamos una revisión para asegurarnos de que todo esté bien. Con esto, damos por terminada la alineación. Los textos procesados ya pueden servir para la consulta, el análisis y la explotación del corpus. Como se trata de un corpus paralelo, aplicaciones como ParaConc, HyConc, BFSU ParaConc y CUC_ParaConc podrían ser útiles.

7. CONCLUSIONES

El corpus paralelo constituye una herramienta eficaz, “no solo como fuente documental de traducciones basadas en contextos reales, sino para llevar a cabo estudios contrastivos entre lenguas” (Castillo Rodríguez, 2009, p. 9). Sin embargo, mientras existen bastantes corpus monolingües de chino clásico (véase Xu, 2015), es muy reducido el número de corpus paralelos cuya combinación lingüística es chino clásico-otra lengua, y los pasos a seguir suelen ser omitidos cuando se trata un corpus paralelo propiamente construido.

En el presente trabajo hemos presentado la problemática de construcción del corpus paralelo chino clásico-español. La disponibilidad de un corpus en dicha combinación lingüística responde a las necesidades investigadores de aquellos que quieren estudiar traducciones de obras clásicas chinas al español. Este tipo de estudios ayudarían a futuros traductores a construir el puente no solo lingüístico sino también temporal para los lectores de estas obras.

Se tiene que matizar que la composición de corpus electrónico en lenguas antiguas es un reto en sí, ya que programas de reconocimiento de texto no siempre son útiles. Hemos tenido suerte de contar con versión digitalizada del texto original, por lo que, en la composición de nuestro corpus bilingüe, solo nos tuvimos que ocupar de los textos en castellano, uno de estos hemos obtenido en versión ya digital y el otro mediante OCR. El tema del texto y los nombres propios han complicado la tarea de reconocimiento automático, por lo que la intervención humana de un conocedor de la lengua china era imprescindible. A continuación, tanto el original como la traducción (en nuestro caso, ambas traducciones), antes de incorporarse a formar un corpus bilingüe, requerían de ajustes del formato para que cada frase ocupase una línea. Por suerte, el chino utiliza el mismo sistema de puntuación que las lenguas occidentales. A ello hay que añadir la necesaria comprobación de que las versiones de original y traducción se difieran en

cuanto al contenido, porque existe una multitud de versiones del original y era necesario asegurarse de que el contenido era al menos paralelo.

Las siguientes fases de composición de nuestro corpus bilingüe-diacrónico han puesto de relieve tanto la complejidad de trabajar con idioma antiguo como las drásticas diferencias entre el chino y el español. En cuanto a este último, primero hay que tener en cuenta que, si las aplicaciones de corpus lo exigen, es necesaria la fase de segmentación porque, a diferencia de idiomas occidentales, en chino normalmente no existe separación entre caracteres. Los resultados de segmentación para los textos escritos en chino clásico no son inaceptables del todo, pero requieren la intervención humana, la que es sobre todo necesaria cuando abundan topónimos relacionados con la realidad que ya no existe y por tanto poco comunes en chino moderno.

La siguiente fase, la anotación, se puede realizar a distintos niveles de complejidad. En el caso de lenguas antiguas, esta fase presenta problemas debido a que no existe una base de datos suficientemente amplia para el entrenamiento de la aplicación, como sucede en lenguas indoeuropeas antiguas. De la misma manera, existen aplicaciones para chino que realizan al mismo tiempo la segmentación y el etiquetado gramatical, pero el resultado de ambos no es satisfactorio cuando se trata de chino clásico, por la misma razón que acabamos de mencionar. En cambio, en la traducción al español, el resultado de etiquetado gramatical es satisfactorio, con la excepción notable en el caso de nombres propios.

Por último, hemos tenido que realizar la alineación totalmente a mano, porque las aplicaciones existentes son casi incapaces de alinear el texto en chino clásico y sus traducciones correspondientes. En vez de simplificar el proceso, pueden complicarlo aún más. De hecho, es un problema pendiente de resolver cuando una de las lenguas involucradas es chino clásico, y no importa si la otra es chino moderno o algún idioma occidental. La razón también reside en la escasez de base de datos para el entrenamiento de *software*.

A pesar de los logros obtenidos, el corpus del presente trabajo puede mejorarse respecto de procedimiento. De manera específica, sería interesante adentrarnos más en la anotación en el futuro: proponer una vía factible para llevar a cabo el etiquetado gramatical del chino clásico manualmente, incidiendo en los pasos a seguir y las etiquetas que poner.

El presente trabajo forma parte de la tesis doctoral en marcha “Análisis de la traducción en español de las palabras funcionales empleadas en *Romance de los Tres Reinos*”. Está financiado por la CSC (China Scholarship Council).

OBRAS CITADAS

- Baker, Mona (1995). Corpora in translation studies: an overview and some suggestions for future research. *Target* 7, 223-243. <https://doi.org/10.1075/target.7.2.03bak>.
- Alonso, Araceli, Mireia Martínez, Noemí Polo, Gal·la Puertas, Lluís de Yzaguirre, Carles Tebé y Teresa Cabré (2001). “La utilización de corpus paralelos alineados en la docencia de la traducción y de los lenguajes de especialidad”. En *Studies in contrastive linguistics.: proceedings of the 2nd International Contrastive Linguistics Conference, Santiago de Compostela, October, 2001*: 71-82. Servicio de Publicaciones.
- Castillo Rodríguez, Cristina (2009). “La elaboración de un corpus ad hoc paralelo multilingüe”. En *Revista tradumàtica: traducció i tecnologies de la informació i la comunicació*, (7), 3.
- Cichosz, Anna (2013). “V-1 main declarative clauses in Old English and Old High German biblical translations”. En *Periphrasis, Replacement and Renewal: Studies in English Historical Linguistics*, ed. Hegedus Iren & Dora Podor. Cambridge: CSP.
- (2012). “A parallel Latin - Old English corpus of biblical translations as a tool for word order studies”. En *Corpus Data Across Languages and Disciplines* (Łódź Studies in Language 28), ed. Piotr Pezik. Frankfurt am Main: Peter Lang.
- Doval, Irene (2018). “Corpus paralelos en la enseñanza de lenguas extranjeras: un ejemplo de aplicación basado en el corpus PaGeS”. En *CLINA*, 4(2), 65-82.
- Eckhoff, H., Bech, K., Bouma, G., Eide, K., Haug, D., Haugen, O. E., y Jøhndal, M. (2018). “The PROIEL treebank family: a standard for early attestations of Indo-European languages”. En *Language Resources and Evaluation* 52(1), 29-65.
- Fan, Min (范敏) (2017). “Jiyu yuliaoku de Lunyu wu yiben wenhua gaopin ci fanyi yanjiu (基于语料库的《论语》五译本文化高频词翻译研究)” [Estudio basado en corpus sobre las palabras culturales en las cinco traducciones de *Las Analectas*]. En *Foreign Language Education* (外语教学), 38(6), 80-83.
- González Rey, María Isabel (2014). “Creación de un corpus literario paralelo como herramienta didáctica en fraseología bilingüe francés-español: criterios de composición.” En *Fraseología y paremiología: enfoques y aplicaciones*, Instituto Cervantes, Biblioteca Fraseológica y Paremiológica, Serie Monografías 5, 153-175.
- Granger, Sylviane (1996). “From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora”. En K. Aijmer, B. Altenberg, & M. Johansson (Eds.) *Languages in Contrast. Papers from a Symposium on Text-Based Cross-Linguistic Studies: Lund, 2-5 March 1994*. Lund, Sweden: Lund University Press.

- Haug, Dag y Marius Jøhndal (2008). "Creating a parallel treebank of the old Indo-European Bible translations." En *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*.
- Hunston, Susan (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524773>.
- Hu, Kaibao (胡开宝) (2011). *Yuliaoku Fanyixue Gailun* (语料库翻译学概论) [Introducción a estudios de traducción basados en corpus]. Shanghai: Shanghai Jiaotong Daxue Chubanshe.
- Huang, Liang, Yinan Peng, Huan Wang y Zhenyu Wu (2002). "PCFG parsing for restricted classical Chinese texts." En *COLING-02: The First SIGHAN Workshop on Chinese Language Processing*, 1-6.
- Kozlova, Inna y Patricia Rodríguez-Inés (2017). "EFECT (Expert Field Environment Collaborative Training) corpus: Cooperation between the academic and professional spheres". En Magdalena Sowa, Jaroslaw Krajka (Eds.) *Innovations in Languages for Specific Purposes-Present Challenges and Future Promises*. Frankfurt am Main: Peter Lang, 207-230.
- Kruger, Alet, Kim Wallmach, y Jeremy Munday (Eds.) (2011). *Corpus-based translation studies: Research and applications*. Bloomsbury Publishing.
- Lee, John (2012). "A classical Chinese corpus with nested part-of-speech tags". En *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, 75-84.
- Liu, Keqiang (刘克强) (2013). *Shui Hu Zhuan Si Yingyiben Fanyi Tezheng Duoweidu Duibi Yanjiu* (《水浒传》四英译本翻译特征多维度对比研究) [Análisis contrastivo multidimensional de las características de traducción de las cuatro versiones al inglés de *A la Orilla del Agua*]. Tesis doctoral, Universidad de Estudios Internacionales de Shanghai.
- (2010). *Hongloumeng Zhongyingwen Yuliaoku de Chuangjian ji Yingyong Yanjiu* (红楼梦中英文语料库的创建及应用研究) [Estudio sobre la elaboración y la aplicación del corpus chino-inglés de *Sueño en el Pabellón Rojo*]. Beijing: Guangming Ribao Chubanshe.
- Liu, Keqiang (刘克强) y Yingmei Shen (沈映梅) (2013). «Sanguoyanyi zhongyingwen pingxing yuliaoku de yanzhi yu chuangjian» (《三国演义》中英文平行语料库

- 的研制与创建)» [Elaboración del corpus paralelo chino-inglés sobre *Romance de los Tres Reinos*]. En *Honghe Xueyuan Xuebao* (红河学院学报), (2), 34-39.
- Luo, Guanzhong (罗贯中) (2012). *Romance de los Tres Reinos* (三国演义). Beijing: Ediciones en Lenguas Extranjeras Cía. Ltda.
- (罗贯中) (2012). *Romance de los Tres Reinos*, traducido por María Teresa Ortega y Olga Marta Pérez. Beijing: Ediciones en Lenguas Extranjeras Cía. Ltda.
- (罗贯中). *Romance de los Tres Reinos*, traducido por Ricardo Cebrián Salé. (En proceso de traducción. El primer tomo se publicó en 2014.)
- (s/f). *Romance de los Tres Reinos*, traducido por Ariel Allevi. (Solamente es una versión digital para Kindle que se puede conseguir a través de <https://www.amazon.es/>)
- McEnery, Tony y Michael Oakes (1996). “Sentence and word alignment in the CRATER project”. En *Using corpora for language research*, 211-231.
- McEnery, Tony, Richard Xiao y Yukio Tono (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.
- Mikhailov, Mikhail y Robert Cooper (2016). *Corpus linguistics for translation and contrastive studies: A guide for research*. Londres: Routledge.
- Molés-Cases, Teresa (2016). “Compilación y análisis de un corpus paralelo para la investigación en traducción: Proyecto con Déjà Vu, Treetagger e IMS Open Corpus Workbench”. En *RLA. Revista de lingüística teórica y aplicada*, 54(1), 149-174.
- Rovira-Esteva, Sara, Helena Casas-Tost, Sílvia Fustegueres i Rosich, Xianghong Qu y Mireia Vargas-Urpi (2015). *Guía de estilo para el uso de palabras de origen chino*. Adeli Ediciones.
- Rovira-Esteva, Sara, Helena Casas-Tost, Irene Tor-Carroggio y Mireia Vargas-Urpi (2019). La literatura china traducida en España. Base de datos en acceso abierto. Disponible en: <https://dtieao.uab.cat/txicc/lite>. DOI: 10.5565/ddd.uab.cat/214778. [Última consulta: 28-08-2020].
- Schmied, Josef y Hildegard Schäffler (1996). “Approaching translationese through parallel and translation corpora”. En *Language and computers*, 16, 41-58.
- Wang, Kefei (王克非) (2004). *Shuangyu Duiying Yuliaoku: Yanzhi yu Yingyong* (双语对应语料库: 研制与应用) [Corpus paralelo bilingüe: diseño y aplicación]. Beijing: Waiyu Jiaoxue y Yanjiu Chubanshe.
- Wang, Li (王力) (2006). *Gudai Hanyu* (古代汉语) [Chino clásico]. Shanghai: Zhonghua Shuju.

- Xu, Jiajin (许家金) (2015). "Corpus-based Chinese studies: A historical review from the 1920s to the present." En *Chinese Language and Discourse* 6.2, 218-244.
- Zhao, Chaoyong (赵朝永) (2014). "Jiyu hanying pingxing yuliaoku de fanyi yuyiyun yanjiu-yi Honglouneng "mang XX" jiegou de yingyi wei li (基于汉英平行语料库的翻译语义韵研究——以《红楼梦》“忙XX”结构的英译为例)" [Estudio basado en corpus paralelo chino-inglés sobre la prosodia semántica-el caso de la estructura "忙XX" y sus traducciones al inglés]. En *Waiyu Jiaoxue Lilun yu Shijian* (外语教学理论与实践), 4: 75-82.