



Robinson Crusoe: Creación de la edición electrónica bilingüe. El proyecto InLéctor. Estudios de Traducción

Marta Grueso Coy y Antoni Oliver¹

Recibido: 3 de mayo de 2017 / Aceptado: 4 de septiembre de 2017

Resumen. Actualmente el libro electrónico va ganando territorio en el mercado editorial pero sigue siendo un proceso lento y costoso. Para incentivar el uso del libro electrónico se pueden ofrecer prestaciones adicionales que mejoren la experiencia lectora del usuario. Una de estas mejoras puede ser ofrecer conjuntamente el texto original y el traducido. El presente trabajo explica el proceso de creación del libro electrónico bilingüe de la novela *Robinson Crusoe* de Daniel Defoe para ser incorporada al proyecto InLéctor. Este proyecto pretende crear herramientas para la creación de libros electrónicos bilingües interactivos. Este proyecto se lleva a cabo en el grupo de investigación Language Processing Group (LPG) de la Universitat Oberta de Catalunya (UOC). Las obras y las traducciones literarias que se publican son de dominio público. Los programas desarrollados se basan en software libre y se publicarán también bajo una licencia libre.

Palabras clave: libro electrónico, literario, texto bilingüe, proyecto InLéctor

Creation of the bilingual electronic edition. Inlector's project.

Abstract. Nowadays the electronic book is gaining territory in the literary market but it is still a slow and expensive process. To encourage the use of the electronic book it is possible to offer additional features that will improve the user's reading experience. One of these improvements can be the offer of the original text and the translated text together. This project consists of the creation of interactive bilingual e-books. This project is carried out in the investigation group Language Processing Group (LPG) of the Universitat Oberta de Catalunya (UOC). The published originals and translations are in the public domain. The developed programs are based on free software and will also be published under a free license.

Keywords: electronic book, literary market, bilingual text, interactive glossary, InLéctor project

Sumario. 1. Introducción. 2. Versiones y traducciones. 3. Proceso técnico. 4. Generación de memorias de traducción. 5. Conclusiones.

Cómo citar: Grueso Coy, M. (2017) *Robinson Crusoe*: Creación de la edición electrónica bilingüe. El proyecto InLéctor. Estudios de Traducción, en *Estudios de Traducción* 7, 205-216.

¹ Universitat Oberta de Catalunya (UOC)
martagruescocoy@gmail.com / aoliverg@uoc.edu

1. Introducción

En este trabajo se detalla el proceso de creación del libro bilingüe correspondiente a la novela *Robinson Crusoe* de Daniel Defoe, para ser incluida al proyecto InLéctor (Oliver, Coll-Florit y Climent 2012). Este proyecto pretende crear herramientas que permitan la creación de libros electrónicos bilingües interactivos de una manera automática. Este proyecto se lleva a cabo en el grupo de investigación Language Processing Group (LPG) de la Universitat Oberta de Catalunya (UOC). El proyecto tiene el objetivo de fomentar la lectura en versión original a partir de libros bilingües, en formato texto y, en un futuro, también en audio. Estos libros permiten leer el texto original y pasar a la traducción mediante un clic. El original y la traducción se presentan como textos paralelos a nivel de oración. Los libros electrónicos bilingües creados en el proyecto InLéctor se publican en los formatos de libros electrónicos más habituales: html, epub y mobi. En este proyecto se publican obras literarias que tengan derechos de autor y de traducción libres, es decir, de dominio público, pero los mismos programas y procesos se pueden utilizar para cualquier libro. Los programas desarrollados se publicarán también bajo una licencia libre por lo que cualquier persona u organización podrá crear sus propios libros bilingües.

En este artículo explicaremos el proceso de creación del libro bilingüe inglés-español correspondiente a la obra de Daniel Defoe: *Robinson Crusoe*. Este es el primer libro que se desarrolla a partir de la colaboración con la colección de libros paralelos en Cybertextos² de Miguel Garci-Gomez. En esta colección se pueden encontrar un gran número de ediciones bilingües, especialmente entre el inglés y el español. En Cybertextos las obras están alineadas a nivel de párrafo, mientras que en InLéctor la alineación se realiza a nivel de oración. Partir de las obras alineadas a nivel de párrafo de Cybertextos facilita enormemente la alineación a nivel de oración.

2. Versiones y traducciones

Robinson Crusoe es un clásico de la literatura publicado ya hace casi 300 años. A causa del éxito inmediato y universal que tuvo, la novela ha sido traducida a lo largo de la historia. A finales del siglo XIX ningún otro libro de la historia de la literatura occidental tenía más ediciones, traducciones y versiones que *Robinson Crusoe*. Actualmente hay más de 700 versiones y traducciones de dicha novela. En este punto, cabe mencionar la distinción entre versión y traducción.

En este trabajo consideramos que una versión es cada una de las formas que adopta un texto o la interpretación de una obra. Se trata de una adaptación de una obra con el objetivo de presentarla de forma distinta de la original u ofrecerla a otro destinatario. Por otro lado, la traducción se considera el proceso de trasvase de unas ideas expresadas en una lengua de partida a una lengua distinta de llegada. Ahora bien, la traducción puede ser directamente de la versión original, puede ser de una versión del original del inglés, e incluso puede ser de una traducción a partir de otra lengua, que a su vez puede ser traducción del original o de otra versión.

² <http://mgarci.aas.duke.edu/cibertextos/EDICIONES-BILINGUES/>

La primera traducción española de *Robinson Crusoe* de Daniel Defoe data de 1835, más de un siglo después del original inglés³. Dicha traducción fue traducida a través del italiano y no del francés como era costumbre, idioma a partir del cual la realizó el traductor italiano. Anterior a la traducción de 1835, hubo una en 1745 pero no vio la luz a causa de la censura. Así pues, se considera la traducción de 1835 como la primera versión impresa en español de la obra *Robinson Crusoe*. Bajo el título *Aventuras de Crusoe, escritas en inglés por Daniel Defoe*, no consta el nombre del traductor. En esta primera traducción predomina la reducción textual a causa de la autocensura dieciochesca y de la síntesis. Uno de los temas censurados es, por ejemplo, la crueldad de los conquistadores españoles y el trato para con los nativos. Además de reducciones por censura también aparecen modificaciones. Es el caso de la instrucción religiosa de Robinson a Viernes que se presenta orientada hacia las enseñanzas de la Iglesia Católica. Por último, cabe hacer mención al estilo propio de Defoe, un estilo sencillo pero reiterativo que también se ve sintetizado en esta primera traducción impresa en español. Por ejemplo, se sintetizan descripciones minuciosas y reflexiones religiosas.

Como hemos comentado anteriormente, toda versión de una obra literaria se ve sometida a cierta manipulación textual. En el caso particular de la primera traducción impresa al español, esta manipulación viene acentuada por la época en que se publicó; en el siglo XVIII regía una censura civil y eclesiástica.

Julio Cortázar tradujo la obra en 1944, por encargo de la editorial Vial de Buenos Aires, a la que le falta aproximadamente un 30 por ciento del texto. Este hecho fue descubierto varias décadas después cuando los traductores Carmen M. Cáceres y Andrés Barba volvieron al original de *Robinson Crusoe* para la versión que les había solicitado la editorial Sexto Piso, y se percataron de las omisiones que hizo Cortázar. Con lo cual, la traducción de Cortázar puede considerarse otra versión. La polémica de su versión fue causada por la supresión que realizó de párrafos enteros y la división del libro en capítulos con su correspondiente título, cosa que nunca hizo Defoe en el original. Estos cambios en la traducción pudieron haber estado motivados: por exigencias de la editorial argentina, porque quizá Cortázar tradujo a partir de otra edición que era una versión o por las libertades recreadoras que se tomó Cortázar.

El éxito colosal de la obra causó que varias versiones de la obra difieran en algunos aspectos del original, como hemos observado en los ejemplos anteriores a causa de la censura. Pero también disponemos de versiones modificadas. Por ejemplo, en la edición de 1959 de la editorial Espasa Calpe el traductor —cuyo nombre no figura ni en la portada ni en los créditos del libro— bautizó como Domingo a Viernes. Todas las versiones de *Robinson Crusoe* a lo largo de casi trescientos años contienen una de dos limitaciones (o ambas): excluyen el segundo volumen, o incluyen los dos pero recortan el texto, a excepción de una versión datada a mediados del siglo XIX, que glosa el texto en exceso a causa de una intención didáctica. La segunda limitación es la existencia de un tercer volumen que nunca ha sido traducido.

No es hasta 2012 cuando el escritor Enrique de Hériz, tras más de año y medio de trabajo, presenta los ejemplares de *Robinson Crusoe* y *Nuevas aventuras de*

³ Infante, E. P. (2012). La primera traducción española del Robinson Crusoe, de Daniel Defoe. En *Caleidoscopio de traducción literaria* (pp. 95-110).

Robinson Crusoe, en la primera traducción al español íntegra y completa, formada por tres volúmenes, el último de los cuales, *Serias reflexiones de Robinson Crusoe*, ha sido publicado por la editorial Edhasa⁴.

Como ya hemos comentado, en este libro bilingüe se ha utilizado la colección de libros paralelos en Cybertextos por Miguel Garci-Gomez. En esta obra disponer de los textos alineados a nivel de párrafo ha supuesto una gran ayuda. Los primeros intentos de crear el libro paralelo no dieron fruto por las grandes diferencias entre el texto original y las traducciones disponibles libres de derechos. En Cybertextos, la labor de seleccionar la traducción más fiel al original (de las disponibles libres de derechos) y además relacionar cada párrafo del original con su correspondiente traducción, ya ha sido realizada. A partir de esta alineación a nivel de párrafo, la alineación a nivel de oración resultará mucho más sencilla.

3. Proceso técnico

En este apartado se explicará el proceso técnico que se lleva a cabo para la creación de un libro electrónico bilingüe. Algunos de los pasos difieren si se realizan a partir de obras en texto sin alinear a nivel de párrafo, o si se realizan a partir de obras ya alineadas a nivel de párrafo (como lo es en el caso que nos ocupa). Todos los algoritmos de creación se han desarrollado en Python y se distribuyen con una licencia libre desde la página de SourceForge del proyecto InLéctor⁵.

Buscar libros con derechos de autor libre

En este trabajo se considera la obra de Daniel Defoe: *Robinson Crusoe*, cuyo original y traducciones han sido extraídas de la colección de libros paralelos en Cybertextos⁶ por Miguel Garci-Gomez. Cybertextos es un sitio web donde aparecen publicadas obras libres de derechos de autor. Se ha llegado un acuerdo entre el proyecto Cybertextos y el proyecto InLéctor por lo que podemos utilizar sus obras ya alineadas a nivel de párrafo.

En otros casos las obras se han obtenido del proyecto Gutenberg⁷ y de Wikisource⁸.

Importación de los textos

Los textos se importan a una hoja de cálculo de Libre/Open Office, en la que se tiene un párrafo en cada celda. Si el texto es bilingüe, es decir, proviene de Cybertextos, se importa en original y la traducción en la segunda y tercera columna respectivamente. La primera columna se reserva para poner las marcas correspondiente a elementos del lenguaje de marcaje Docbook. Para un libro bilingüe la hoja de cálculo tendría el siguiente aspecto:

⁴ <http://www.lavanguardia.com/libros/20120113/54244890461/enrique-de-heriz-traduce-por-primera-vez-al-castellano-todo-robinson-crusoe.html>

⁵ <https://sourceforge.net/projects/inlector/>

⁶ <http://mgarci.aas.duke.edu/cgi-bin/celestina/en/index.cgi?libroId=2038>

⁷ <http://www.gutenberg.org/>

⁸ <https://wikisource.org>

	Robinson Crusoe	
	Daniel	
	Defoe	
Chapter	Start in life	
	I was born in the year 1632, in the city of York, of a good family, though not of that country, my father being a foreigner of Bremen, who settled first at Hull. He got a good estate by merchandise, and leaving off his trade lived afterward at York, from whence he had married my mother, whose relations were named Robinson, a good family in that country, and from whom I was called Robinson Kreutznear; but by the usual corruption of words in England we are now called, nay, we call ourselves, and write our name, Crusoe, and so my companions always called me.	Nací en 1632, en la ciudad de York, de una buena familia, aunque no de la región, pues mi padre era un extranjero de Brema que, inicialmente, se asentó en Hull. Allí consiguió hacerse con una considerable fortuna como comerciante y, más tarde, abandonó sus negocios y se fue a vivir a York, donde se casó con mi madre, que pertenecía a la familia Robinson, una de las buenas familias del condado de la cual obtuve mi nombre, Robinson Kreutznaer. Mas, por la habitual alteración de las palabras que se hace en Inglaterra, ahora nos llaman y nosotros también nos llamamos y escribimos nuestro nombre Crusoe; y así me han llamado siempre mis compañeros.
	I had two elder brothers, one of which was lieutenant-colonel to an English regiment of foot in Flanders, formerly commanded by the famous Colonel Lockhart, and was killed at the battle near Dunkirk against the Spaniards; what became of my second brother I never knew, any more than my father and mother did know what was become of me.	Tenía dos hermanos mayores, uno de ellos fue coronel de un regimiento de infantería inglesa en Flandes, que antes había estado bajo el mando del célebre coronel Lockhart, y murió en la batalla de Dunkerque contra los españoles. Lo que fue de mi segundo hermano, nunca lo he sabido al igual que mi padre y mi madre tampoco supieron lo que fue de mí.

Tabla 1. Representación del inicio de Robinson Crusoe una vez importado a hoja de cálculo

Si nos fijamos en la tabla 1, observaremos que las tres primeras líneas de la segunda columna (correspondiente al original inglés) tenemos respectivamente el título de la obra, el nombre del autor, y el apellido del autor. Esta información puede estar o no en la tercera columna (correspondiente a la traducción castellana). En la cuarta fila, en la primera columna, tenemos la marca correspondiente a inicio de capítulo (*chapter*), y en la segunda columna tendremos el título del capítulo en la lengua original (que puede estar vacío si el capítulo no tiene título), y en la tercera el título del capítulo en la lengua correspondiente a la traducción (que puede estar vacío como en este caso). En las siguientes filas tendremos los diferentes párrafos tanto del original como de la traducción, hasta que encontremos un nuevo capítulo.

En la primera columna se pueden poner diversas marcas correspondientes al lenguaje Docbook, que tanto pueden corresponder a divisiones la obra en partes (*part*),

capítulos (*chapter*) o secciones (*section*); como a visualizaciones especiales del contenido (como por ejemplo literallayout que fuerza que el contenido se muestre tal y como está escrito).

Para las obras no alineadas a nivel de párrafo el procedimiento es el mismo, pero se crean dos hojas de cálculo: una para el original y la otra para la traducción. Cada una de estas hojas de cálculo dispondrá de dos columnas, la primera para las marcas de Docbook y la segunda para el texto de la obra. Las primeras tres filas de la hoja de cálculo también contendrán el título de la obra, el nombre y el apellido del autor.

Con el uso de la hoja de cálculo se pretende facilitar la posterior creación de los libros en formato docbook por personas sin conocimientos específicos de XML en general ni de docbook en particular.

Entrenamiento de segmentadores

Los libros electrónicos bilingües de la colección InLéctor están alineados a nivel de oración (o segmentos parecidos a una oración). Para poder realizar esta alienación es imprescindible segmentar el texto, tanto en la lengua de partida como en la de llegada. Este paso es crítico, ya que una segmentación deficiente provocará problemas serios de alineación posterior. Estos problemas tendrán como consecuencia saltos inesperados entre el original y la traducción en el libro final.

Para optimizar el proceso de segmentación, se crea un segmentador específico para cada obra y lengua. Para entrenar este segmentador específico se utiliza la implementación que proporciona el *Natural Language Toolkit* (NLT) (Bird, 2006) del algoritmo de Kiss and Strunk (2006). Para entrenar el segmentador para cada lengua se utiliza como entrada el propio texto de la obra que se está tratando en la lengua correspondiente. Posteriormente el segmentador se enriquece con una lista de abreviaturas habituales en la lengua correspondiente.

Creación del Docbook bilingüe

En este paso se crea el libro bilingüe en formato Docbook, que es un formato estándar basado en XML para la creación de documentos. Se ha optado por el uso de este formato ya que al tratarse de un formato estándar, y además basado en XML, existen una gran cantidad de herramientas de software libre que nos permiten su tratamiento y manipulación.

El Docbook bilingüe tiene la particularidad que incluye un enlace entre cada oración del original y de la traducción. A continuación podemos observar un fragmento de docbook bilingüe:

```
<para><phrase id="ss-98"><link linkend="ts-98">I had a neighbor, a Portuguese
of Lisbon, but born of English parents, whose name was Wells, and in much such
circumstances as I was.</link></phrase>
<phrase id="ss-99"><link linkend="ts-99">I call him my neighbor, because his plan-
tation lay next to mine, and we went on very sociably together.</link></phrase>
<phrase id="ss-100"><link linkend="ts-100">My stock was but low, as well as
his; and we rather planted for food than anything else, for about two
years.</link></phrase>
```

<phrase id="ss-101"><link linkend="ts-101">However, we began to increase, and our land began to come into order; so that the third year we planted some tobacco, and made each of us a large piece of ground ready for planting canes in the year to come.</link></phrase>

<phrase id="ss-102"><link linkend="ts-102">But we both wanted help; and now I found, more than before, I had done wrong in parting with my boy Xury.</link></phrase>

</para>

.....

<para><phrase id="ts-98"><link linkend="ss-98">Tenía un vecino, un portugués de Lisboa, hijo de ingleses, que se llamaba Wells y se encontraba en una situación similar a la mía.</link></phrase>

<phrase id="ts-99"><link linkend="ss-99">Digo que era mi vecino, ya que su plantación colindaba con la mía y nos llevábamos muy bien.</link></phrase>

<phrase id="ts-100"><link linkend="ss-100">Mis existencias eran tan escasas como las tuyas, pues, durante dos años, sembramos casi exclusivamente para subsistir.</link></phrase>

<phrase id="ts-101"><link linkend="ss-101">Con el tiempo, comenzamos a prosperar y aprendimos a administrar mejor nuestras tierras, de manera que, al tercer año, pudimos sembrar un poco de tabaco y preparar una buena extensión de terreno para sembrar azúcar al año siguiente.</link></phrase>

<phrase id="ts-102"><link linkend="ss-102">Ambos necesitábamos ayuda y, entonces, me di cuenta del error que había cometido al separarme de Xury, mi muchacho.</link></phrase>

</para>

Como podemos observar, el párrafo en cuestión tiene 4 segmentos tanto en castellano como en inglés. El segmento original (ss) número 98 está enlazado con el segmento traducido (ts) número 98 (la coincidencia de número no siempre se da).

4a. Creación del Docbook bilingüe (a partir de alineados por párrafos)

En el caso de los documentos alineados por párrafos provenientes de la colección Cybertextos, el procedimiento de alineación a nivel de oración se simplifica:

- Para cada párrafo se segmenta el original y la traducción con el segmentador específico para esa lengua entrenado en el paso anterior.

- Si el número de segmentos del original y la traducción coinciden, se marcan estos segmentos con la marca *phrase* i el identificador numérico del párrafo y se crea el enlace con el correspondiente segmento traducido.

- Si el número de segmentos del original y la traducción no coinciden, se lleva mantiene la alineación a nivel de párrafo y se marca todo el párrafo con la marca *phrase*, con su identificador numérico correspondiente y su enlace al correspondiente a la lengua de llegada.

En los libros creados a partir de textos alineados a nivel de párrafo, los identificadores numéricos del original y la traducción siempre coinciden.

4b. *Creación del Docbook bilingüe (a partir de textos sin alinear a nivel de párrafo)*

Cuando no disponemos de los textos alineados a nivel de párrafo el procedimiento a seguir difiere en buena parte. Los pasos son los siguientes:

- A partir de la hoja de cálculo que contiene los libros tanto en original como en traducción se segmentan los párrafos con el segmentador específico entrenado para cada lengua. El programa crea una nueva hoja de cálculo que contiene ahora más columnas: marca docbook (si la hay) incluyendo la marca de párrafo, número de segmento, texto del segmento. En la siguiente tabla podemos observar un ejemplo en inglés:

para	6	
	7	To Sherlock Holmes she is always THE woman.
	8	I have seldom heard him mention her under any other name.
	9	In his eyes she eclipses and predominates the whole of her sex.

Y su correspondiente en español:

para	4	
	5	Ella es siempre, para Sherlock Holmes, la mujer.
	6	Rara vez le he oído hablar de ella aplicándole otro nombre.
	7	A los ojos de Sherlock Holmes, eclipsa y sobrepasa a todo su sexo.

- La alineación de documentos se realiza con la herramienta Hunalign (Varga, 2007). Esta herramienta es capaz de realizar una alineación totalmente automática de dos documentos. Para realizar la alineación puede utilizar diccionarios bilingües entre las lenguas implicadas. Hemos creado diccionarios bilingües en el formato adecuado a partir de los diccionarios de transferencia del sistema de traducción automática Apertium (Forcada, 2011). Las hojas de cálculo segmentadas para cada lengua se transforman en el formato adecuado para esta herramienta. En inglés:

<p>

To Sherlock Holmes she is always THE woman.

I have seldom heard him mention her under any other name.
In his eyes she eclipses and predominates the whole of her sex.

Y en español:

<p>

Ella es siempre, para Sherlock Holmes, la mujer.

Rara vez le he oído hablar de ella aplicándole otro nombre.

A los ojos de Sherlock Holmes, eclipsa y sobrepasa a todo su sexo.

El programa retorna un archivo de alineación que indica los números de línea del original que se corresponden con la traducción. Veamos el correspondiente al ejemplo:

6	4	0
7	5	1.45135
8	6	1.21656
9	7	0.906857

A partir de los dos archivos de hoja de cálculo más el archivo con el resultado de la alineación se puede crear el libro en formato docbook bilingüe:

```
<para><phrase id=>ss-7>><link linkend=>ts-5>>To Sherlock Holmes she is
always THE woman</link></phrase>
<phrase id=>ss-8>><link linkend=>ts-6>>I have seldom heard him mention her
under any other name.</link></phrase>
<phrase id=>ss-9>><link linkend=>ts-7>>In his eyes she eclipses and predomi-
nates the whole of her sex.</link></phrase>
...
<para><phrase id=>ts-5>><link linkend=>ss-7>>TElla es siempre, para Sherlock
Holmes, la mujer.</link></phrase>
<phrase id=>ts-6>><link linkend=>ss-8>>Rara vez le he oído hablar de ella apli-
cándole otro nombre.</link></phrase>
<phrase id=>ts-7>><link linkend=>ss-9>>A los ojos de Sherlock Holmes, eclipsa
y sobrepasa a todo su sexo.</link></phrase>
```

Transformación a EPUB

Para transformar el Docbook bilingüe a epub utilizaremos la herramienta *dbtoepub*, que se distribuye con el *The Docbook project*⁹. Para evitar que los enlaces entre los segmentos originales y sus traducciones aparezcan subrayados se utiliza una hoja de estilo css específica, que contiene el siguiente código:

```
.link {
color: #000;
```

⁹ <http://docbook.sourceforge.net>

```

cursor: inherit;
line-height: 1.2;
text-decoration: none
}

```

Transformación a mobi

La transformación a mobi se realiza a partir del epub utilizando el programa «Kindlegen»¹⁰. Mobipocket, conocido popularmente como mobi, es un formato para archivos de libro electrónico. En su formato de libro digital, se marca el contenido, pero no se delimita su formato, que se adapta a los diferentes tamaños de las pantallas.

Generación del html

Para generar el html a partir del DocBook se utiliza *xsltproc*. Una vez transformado es importante modificar la sección *head* del html para evitar que los enlaces aparezcan subrayados.

Publicación de los libros

Todo los libros de InLéctor se publican en su página web¹¹ Hasta el momento se han publicado los siguientes libros:

Lenguas	Autor	Título
eng-spa	Sir Arthur Conan Doyle	The Adventures of Sherlock Holmes
eng-spa	Daniel Defoe	Robinson Crusoe
eng-spa	Jane Austen	Sense and Sensibility
eng-spa	Robert Louis Stevenson	Strange Case of Dr Jekyll and Mr Hyde
fra-spa	Alexandre Dumas	Les Trois Mousquetaires,
fra-cat	Anatole France	L'île des pingouins
rus-spa	Фёдор Михайлович Достоевский	Игрок

¹⁰ <https://www.amazon.com/gp/feature.html?docId=100076521111>

¹¹ <https://inlector.wordpress.com/>

4. Generación de memorias de traducción

Una vez alineados los textos para construir el libro bilingüe es trivial crear memorias de traducción y de esta manera construir un conjunto de memorias que puedan tener diversas utilidades: ayuda en el aprendizaje de lenguas o en el aprendizaje de la escritura literaria, estudios de traducción o como corpus de entrenamiento de sistemas de traducción automática (Yu, 2012). No se espera una utilidad de estas memorias como ayuda a la traducción en cuanto a la recuperación de coincidencias exactas o parciales, ya que estas son muy poco frecuentes en los textos literarios. No obstante, utilizando las funciones estándar de búsqueda en memorias de traducción de las diversa herramientas de traducción asistida, se podrá encontrar información interesante de cómo traducir una determinada expresión.

Así pues, las memorias de traducción son un producto secundario, aunque no por ello menos interesante, del proyecto InLéctor y se pueden descargar de la página de SourceForge del proyecto¹²

Lenguas	Título	Segmentos	Tokens L1	Tokens L2
eng-spa	https://sourceforge.net/projects/inlector/		100.419	100.905
eng-spa	Robinson Crusoe	1.498	122.478	104.779
eng-spa	Moby Dick	6.340	213.718	218.478
eng-spa	Sense and Sensibility	5.011	120.809	120.965
eng-spa	Strange Case of Dr Jekyll and Mr Hyde	1.107	24.637	21.017
fra-spa	Les Trois Mousquetaires,	11.731	208.217	198.674
fra-cat	L'île des pingouins	42.17	79.318	79.458
rus-spa	Ирпок	1.685	17.326	18.186

5. Conclusiones

En este trabajo se ha presentado el proceso de creación de libros electrónicos bilingües en el proyecto InLéctor, aprovechando la reciente publicación del libro inglés-español de la obra *Robinson Crusoe* de Daniel Defoe. Esta obra nos ha plantado una nueva problemática, que no había ocurrido antes en el proyecto: la obra en cuestión tiene una gran cantidad de versiones diferentes, y las traducciones han sufrido diversas mutilaciones y cambio. Este hecho ha significado que haya sido imposible alinear la obra por los métodos convencionales. Gracias a la cesión de los archivos de esta obra en la colección Cybertextos, la creación del libro electrónico bilingüe ha sido posible. La colección Cybertextos consiste en originales de obras literarias y

¹² <https://sourceforge.net/projects/inlector/>

sus traducciones, alineadas manualmente a nivel de párrafo. Fruto de esta primera colaboración, el responsable de la colección Cybertextos ha hecho extensible el acuerdo de uso para el resto de obras de la colección, lo que sin duda significará un impulso considerable al proyecto InLéctor.

En el artículo se detalla el proceso técnico de creación de los libros electrónicos bilingües, tanto si se parten de obras en bruto como si se parte de obras alineadas a nivel de párrafo. En el proceso se utilizan únicamente herramientas libres o gratuitas y se ha simplificado al máximo para que cualquier persona o entidad pueda crear sus propios libros.

Referencias bibliográficas

- Cuando el Domingo es Viernes, o al revés.* (13 de mayo de 2012). *El infierno de Barbusse*. Recuperado en <http://www.elinfiernodebarbusse.com/2012/05/cuando-el-domingo-es-viernes-o-al-reves.htm>
- Bird, S. (2006, July). NLTK: the natural language toolkit. In *Proceedings of the COLING/ACL on Interactive presentation sessions* (pp. 69-72). Association for Computational Linguistics.
- de Hériz, E. (7 de febrero de 2012). En busca de Robinson Crusoe. *Letras libres*. Recuperado en <http://www.letraslibres.com/mexico-espana/en-busca-robinson-crusoe>
- Enrique de Hériz traduce por primera vez al castellano todo 'Robinson Crusoe'* (13 de enero de 2012). *La Vanguardia*. Recuperado en <http://www.lavanguardia.com/libros/20120113/54244890461/enrique-de-heriz-traduce-por-primera-vez-al-castellano-todo-robinson-crusoe.html>
- González, A. O. (2015). TMX: intercambio de memòries de traducció. *Revista Tradumàtica: tecnologies de la traducció*, (13), 608-622.
- Infante, E. P. La primera traducción española del Robinson Crusoe, de Daniel Defoe. In *Caleidoscopio de traducción literaria* 2012, 95-110.
- Kiss, T., & Strunk, J. Unsupervised multilingual sentence boundary detection. *Computational Linguistics*, 32(4), 2006, 485-525.
- Mikel L. Forcada, Mireia Ginestí-Rosell, Jacob Nordfalk, Jim O'Regan, Sergio Ortiz-Rojas, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Gema Ramírez-Sánchez and Francis M. Tyers: «Apertium: a free/open-source platform for rule-based machine translation». In *Machine Translation: Volume 25, Issue 2* (2011), 127-144.
- Oliver, A., Climent, S., & Coll-Florit, M. (2012). InLéctor: Sistema de lectura bilingüe interactiva. *Procesamiento del Lenguaje Natural*, 49, 203-207.
- Varga, D., Halácsy, P., Kornai, A., Nagy, V., Németh, L., & Trón, V. Parallel corpora for medium density languages. *AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES* 4, 2007, 292, 247.
- Yu, Q., Max, A., & Yvon, F. Aligning bilingual literary works: a pilot study. In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, 2012, 36-44.