

Corpus e interpretación biosanitaria: extracción terminológica basada en bitextos del campo de la Neurología para la fase documental del intérprete¹

Miriam Seghiri*

Resumen: En el presente trabajo ilustraremos una metodología de extracción terminológica basada en bitextos para la creación de un glosario bilingüe y bidireccional (inglés-español/español-inglés) para la interpretación biosanitaria y, más concretamente, del campo de la Neurología. Con objeto de ilustrar esta metodología nos centraremos en los parámetros de diseño del corpus de bitextos y, a continuación, en el protocolo de compilación en seis fases, a saber, búsqueda, descarga, conversión de formato, codificación, almacenamiento y alineación. Finalmente se procederá a la extracción semiautomática de los términos que integrarán el glosario bilingüe de neurología con el programa LexTerm.

Palabras clave: bitextos, documentación, glosario, interpretación, lingüística de corpus, terminología.

Corpora and medical interpreting: terminology extraction based on bitexts for the interpreter's documentation process

Abstract: In this paper we apply the benefits of exploiting corpora, and particularly bitexts, for extracting terminology for interpreters. We present a systematic methodology for extracting bilingual and bidirectional glossaries (English-Spanish/Spanish-English) based on bitexts for medical interpreters working in Neurology. In order to illustrate this methodology, we focus on corpus design and on the compilation protocol in six steps: searching, downloading, text formatting, giving a code number, saving data and alignment. Finally, we illustrate how to extract the terms semiautomatically in order to build a bilingual and bidirectional glossary with LexTerm.

Keywords: bitexts, corpus linguistics, documentation, glossary, interpreting, terminology.

Panace@ 2017; 18 (46): 123-132

Recibido: 15.XI.2017. Aceptado: 06.XII.2017.

1. Introducción

La preparación documental tanto para un traductor como para un intérprete resulta fundamental para poder ofrecer un producto final de calidad; en este sentido, tal y como apunta Gile (2009: 144-158), a pesar de las diferencias existentes entre las dos profesiones (y disciplinas), un buen conocimiento lingüístico y terminológico es esencial para el ejercicio de ambas. Así, según Choi (2005), la preparación previa del intérprete cuenta con dos fases, una primera de corte temático, que le permitirá comprender los contenidos y conceptos del tema objeto de preparación; y, una segunda, de corte terminológico, que consiste en familiarizarse con la terminología, léxico y expresiones del ámbito objeto de estudio. De esta forma, la preparación terminológica y la temática coinciden en gran medida y ambas pertenecen, como no puede ser de otro modo, a la fase documental de carácter previo a la interpretación. Sin embargo, como subraya Gile (2009: 144-158) en el caso concreto del intérprete, la fase documental previa se perfila como fundamental en tanto que este no cuenta con muchas posibilidades de poder documentarse durante el proceso de ejecución del encargo, mientras que el traductor sí puede hacerlo. De hecho, el intérprete, durante la interpretación misma, precisa de materiales que

sirvan para la comunicación inmediata y que puedan ser consultados rápidamente. Como bien expone Lievens (2015: 30), «la solución para el problema léxico es la construcción de un glosario». Esta afirmación coincide con los resultados de una encuesta realizada a intérpretes elaborada por Corpas Pastor y Fern (2015: 54) que reveló que el recurso preferido por estos profesionales son los glosarios o diccionarios bilingües, tanto para la fase previa como durante el proceso de interpretación.

Sin embargo, a pesar de la predilección de los intérpretes por el uso de glosarios, el principal problema es que, en muchas ocasiones, no existen sobre la temática específica sobre la que tienen que interpretar ni en el par de lenguas con el que trabajan. De este modo, no les queda otra solución que elaborar su propio glosario durante la fase documental que, a su vez, podrán consultar durante el proceso de interpretación. Estos glosarios pueden construirse de dos formas, bien de manera manual, o bien de forma (semi-) automática. En este sentido, no cabe duda de que, habida cuenta del escaso tiempo con el que cuenta el intérprete en la fase previa a la interpretación, a lo que se suma, en ocasiones, la falta de información sobre la temática objeto de interpretación, la opción de elaborar un glosario de forma (semi-) automática en escasos minutos

* Departamento de Traducción e Interpretación de la Universidad de Málaga, Málaga (España). Dirección para correspondencia: seghiri@uma.es

—incluso, momentos antes de entrar en cabina— se perfila como una solución óptima.

Son múltiples los trabajos que señalan las virtudes del uso de corpus para traducir (*cf.* Bowker, 1998; Varantola, 2000; Seghiri, 2011; Sánchez Ramos, 2016; Sánchez Ramos y Vigier Moreno, 2016; Vigier Moreno, 2016; Postigo y Calleja, 2017; Postigo y Mira, 2017), incluso algunos van más allá y subrayan las ventajas del empleo de bitextos para la generación de glosarios para la traducción y, de forma más incipiente, para la interpretación (*cf.* Seghiri, 2006, 2011, 2015, 2017; Gallego y Tolosa, 2012; Pérez Pérez, 2013; Lievens, 2015; Corpas Pastor y Seghiri, 2016 o Zaretskaya *et al.*, 2016, entre otros).

Así, en este trabajo, presentaremos una metodología protocolizada para la extracción terminológica a partir de la explotación de bitextos. Esta metodología adaptará los postulados de Seghiri (2006, 2011, 2015 y 2017), orientados a traductores, para la generación de glosarios de forma (semi-) automática para intérpretes. Se ejemplificará a través de la creación de un corpus bilingüe (español-inglés) de textos paralelos o bitextos, esto es, textos originales y sus traducciones, del ámbito biosanitario, y más concretamente del campo de la Neurología. Esta elección no es casual, pues viene dada por el hecho de que «[d]esde que en 2007 se publicó el actual sistema formativo de la especialidad de Neurología [...] la Neurología es una de las especialidades preferidas por los primeros números del MIR y en una de las que primero se cubren la totalidad de las plazas ofertadas. [De hecho], la demanda de plazas de Neurología ha crecido un 86% entre los 500 primeros números del MIR» (Pérez Abril, 2015).

2. Definición de Neurología

La Neurología, que ya se había consolidado como disciplina en las últimas décadas del siglo XIX, es una especialidad médica (Zarranz, 2007: 9) que se incluye en la clasificación universal de la UNESCO (1983: 8). Es posible localizarla en esta clasificación dentro de las ciencias médicas (categoría 32, *Medical Sciences*) y, más concretamente, se ubica en medicina interna (nivel 3205, *Internal medicine*) como una subdisciplina (subdisciplina 3205.07, *Neurology*). La Sociedad Española de Neurología (2010: 1) la define como sigue:

[1]a Neurología es la especialidad médica que estudia la estructura, función y desarrollo del sistema nervioso (central, periférico y autónomo) y muscular en estado normal y patológico, utilizando todas las técnicas clínicas e instrumentales de estudio, diagnóstico y tratamiento actualmente en uso o que puedan desarrollarse en el futuro. La Neurología se ocupa de forma integral de la asistencia médica al enfermo neurológico, de la docencia en todas las materias que afectan al sistema nervioso y de la investigación, tanto clínica como básica, dentro de su ámbito.

Asimismo, según el Consejo Nacional de Especialidades Médicas (1996: 432), el neurólogo sería el médico que tiene como especialidad la Neurología y el campo de acción de la

Neurología se daría tanto en el ámbito asistencial (la evaluación de las enfermedades del sistema nervioso central y periférico), como en el docente (la enseñanza de la especialidad en los distintos niveles formativos) y en el investigador (la investigación básica y clínica aplicada).

Por lo que se refiere a las enfermedades neurológicas, son aquellas que afectan al sistema nervioso central y periférico (Sociedad Española de Neurología, 2010: 1) y se estima que existen más de 600 entidades, o patologías, diferentes (MedlinePlus, 2013). Para la clasificación de las patologías neurológicas se suele recurrir a la novena revisión (Modificación Clínica) de la Clasificación Internacional de Enfermedades (Organización Mundial de la Salud, 2012), también conocida como CIE-9-MC. En la siguiente figura (*cf.* Figura 1) se pueden observar los apartados principales que se enumeran en las *Enfermedades del Sistema Nervioso* (sin incluir los Órganos Sensoriales), según la mencionada CIE-9-MC (Organización Mundial de la Salud, 2012: 691-715):

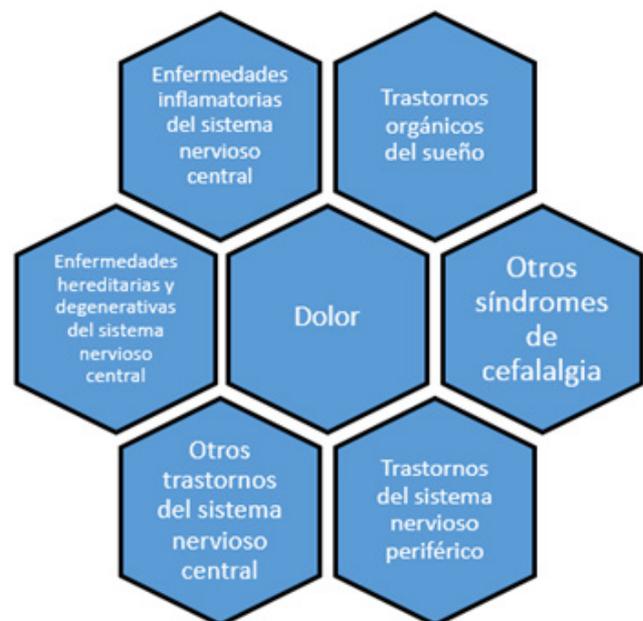


Figura 1. Clasificación de enfermedades neurológicas según la clasificación CIE 9 (elaboración propia)

3. Elaboración de un corpus de bitextos de Neurología

Un corpus puede definirse como una colección de textos escogidos según unos criterios de diseño (Atkins *et al.*, 1991: 1) que se pueden utilizar como muestra de una lengua o parte de ella (Tengku *et al.* 2010: 6-7; Atkins *et al.*, 1991: 2), aunque debe ser representativo para ser considerado un corpus propiamente dicho (Corpas, 2001: 156; Seghiri, 2006). Dentro de todos los tipos de corpus existentes, en el presente trabajo nos centraremos en aquel «virtual» que, en palabras de Seghiri (2011: 16 y 17), es aquel que se crea únicamente a partir de recursos electrónicos en Internet con el objetivo de realizar la labor de documentación de forma económica (únicamente se precisa de un ordenador con conexión a Internet), rápida (los textos se pueden descargar en cuestión de segundos) y fiable (se pueden consultar textos del mismo género y actuales).

Asimismo, nuestro objetivo será compilar un corpus paralelo bilingüe, o de bitextos. Este último término, bitexto, fue acuñado por Harris (1988: 8) quien, desde su punto de vista, considera que un texto original y su traducción no son en realidad dos textos sino que conforman «a single text in two dimensions, each of which is a language». Otros autores completan esta definición indicando que un corpus paralelo o de bitextos puede estar compuesto bien por textos originales en un idioma y sus traducciones en otra lengua, bien por textos traducidos sin su texto origen (Tengku *et al.*, 2010: 12).

Una vez aclarados estos conceptos, procedemos a abordar los criterios de diseño (*cf.* 3.1.) y el protocolo de compilación (*cf.* 3.2) del corpus virtual de bitextos de Neurología.

3.1. Criterios de diseño

Con objeto de construir un corpus de bitextos del campo de la Neurología, será necesario encontrar un género en el que abunden los textos originales y sus respectivas traducciones, a la par que terminología especializada. De este modo, en los artículos de investigación médica es frecuente encontrar resúmenes en varios idiomas que, habitualmente, han sido sometidos a un proceso de revisión por los evaluadores de las revistas donde se publican, por lo que su extracción supondría un valioso recurso como base para la construcción de un futuro glosario para el intérprete. Así, se pretende elaborar un corpus de bitextos de *abstracts* o resúmenes de artículos especializados de Neurología en inglés y español. Para este corpus, los criterios de diseño que se establecen son el género (artículos de investigación científicos y, en particular, sus resúmenes o *abstracts*), la autoría (especialistas que publican en revistas de medicina, y en particular de Neurología) y la temática de los mismos (Neurología), así como su antigüedad (de 2010 hasta la fecha). El corpus será *virtual*, pues se recurrirá exclusivamente a resúmenes publicados en revistas electrónicas de medicina. Por otro lado, el medio de los textos que se recogerán en el corpus será *escrito*, en concreto, se tratará de resúmenes redactados por médicos procedentes de España o que trabajen en este país (con objeto de limitar diatópicamente el corpus al español peninsular), así como los *abstracts* de esos mismos artículos traducidos al inglés, por lo que el corpus será, a su vez, bilingüe y monodireccional. En cuanto a la antigüedad de los documentos, como ya hemos indicado, solamente se han utilizado aquellos artículos que se publicaron a partir del 2010 para contar con un número considerable de muestras para el corpus pero que, a su vez, no estén desfazados con respecto a los avances médicos que se han producido en este campo. Del mismo modo, el corpus será parcial, ya que no se recurrirá a artículos científicos completos, sino que se extraerán exclusivamente sus resúmenes, por ser la única parte del trabajo que se encuentra, por lo general, en dos lenguas. Ya que todos los textos recogidos son resúmenes y tienen la misma temática, el corpus será homogéneo. Asimismo, será representativo a nivel cualitativo, pues los resúmenes fueron escritos por especialistas en el campo de la Neurología.

Así, nuestro objetivo será compilar un corpus de bitextos, bilingüe y monodireccional (español-inglés), textual, virtual, actual, parcial, homogéneo y representativo de resúmenes de artículos científicos de investigación en Neurología, que servirá de base para la elaboración de un glosario bilingüe y bidireccional para el intérprete.

3.2. Protocolo de compilación

Con el diseño del corpus de bitextos bien establecido, y para asegurar que el corpus sea representativo desde el punto de vista cualitativo, es importante seguir un protocolo de compilación. Se destacan, así, cinco fases basadas en los postulados de Seghiri (2006, 2011, 2015 y 2017), que se describirán con más detalle a continuación, a saber, localización de bitextos, descarga, conversión de formato, codificación y almacenamiento. A estas fases, propias de la compilación de corpus comparable, le sigue una sexta, la alineación (*cf.* Castillo Rodríguez, 2009), necesaria para la futura gestión y explotación de los bitextos.

El primer paso en el proceso de compilación consiste en la localización de bitextos, a saber, de los resúmenes de los artículos de investigación en español e inglés que conformarán el corpus de Neurología. En la actualidad hay una gran variedad de fuentes disponibles en Internet donde se pueden encontrar textos médicos. No obstante, para este corpus se realiza una búsqueda institucional, pues las fuentes de las cuales se obtengan los textos que compondrán el corpus de bitextos deben ser fiables para asegurar la representatividad cualitativa de la muestra. Por este motivo, todos los resúmenes se han extraído de la revista electrónica *Neurología*,² perteneciente a la Sociedad Española de Neurología, una de las principales asociaciones médicas en su campo de especialidad.

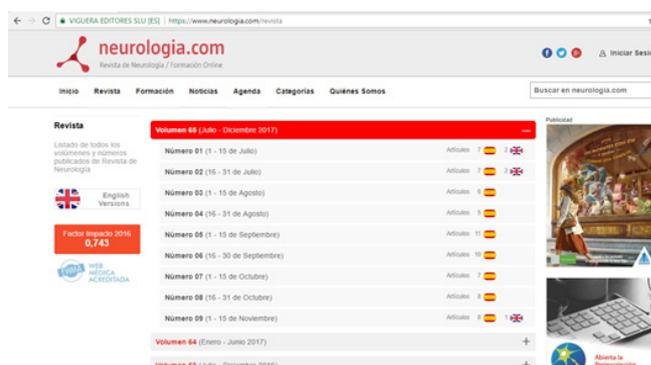


Figura 2. Página principal de la revista electrónica *Neurología*

Tras localizar los resúmenes, el segundo paso consistirá en la descarga de los bitextos para poder acceder a ellos sin necesidad de una conexión a Internet. En el caso de la revista *Neurología* se puede acceder de forma gratuita a todos los *abstracts* y resúmenes publicados sin necesidad de suscripción alguna. Las descargas se pueden realizar bien de forma manual, a través de la opción de búsqueda de la revista, bien por lotes (o grupos de resúmenes) gracias a programas como HTTrack.³

Una vez descargados los resúmenes en inglés y en español, se ha podido observar que los documentos que se han obtenido se encuentran en formato HTML (.html) o PDF (.pdf). En este sentido, el tercer paso consistirá en una conversión de formato de los documentos descargados para que el programa que explote el futuro corpus de bitextos los pueda reconocer, pues, por lo general, estos programas solo gestionan documentos en formato de texto plano (.txt). Así, los resúmenes en formato HTML se pueden tratar de dos maneras, a saber, bien copiando el texto relevante en un formato de texto plano (.txt) de forma manual; o bien recurriendo a un programa para convertir los textos en HTML a texto plano, como Online-Converter.⁴ Por otro lado, para los resúmenes que se encontraban en PDF, se utilizó el programa de conversión de formato PDF (.pdf) a texto plano (.txt) *UniPDF*⁵ (cfr. Figura 3), a través de la selección de la opción *Text* (texto) disponible en las opciones de *Output format* (formato de salida):

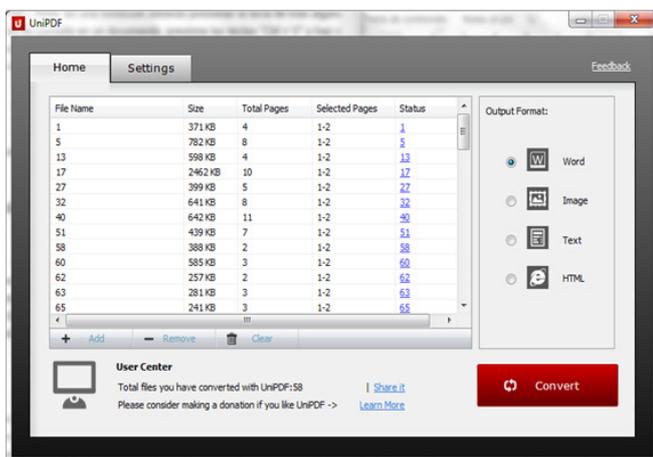


Figura 3. Conversión de resúmenes en formato PDF (.pdf) a texto plano (.txt) con el programa UniPDF

También debe revisarse el texto, ya que pueden aparecer errores al convertirse al nuevo formato, como la desaparición de la «ñ» o de algunas tildes, entre otros (cfr. Castillo Rodríguez, 2009: 6). Esta tarea de revisión puede automatizarse a través del programa Scleaner.⁶ El mencionado programa elimina las tabulaciones y espacios en blanco extras, corrige erratas, o divide las frases por el sitio correcto automáticamente, entre otras opciones.

Una vez descargados los documentos, el cuarto paso consiste en la codificación de los bitextos para que se puedan identificar con facilidad y, con objeto de que los programas para la gestión de bitextos puedan explotarlos en paralelo, esto es, que puedan identificar qué documento es el original y su correspondiente traducción. A este respecto, se pueden crear una serie de códigos compuestos por números y letras. En nuestro caso, hemos optado por la siguiente codificación (cfr. Figura 4), a saber:

1. Número: comenzando por «01», la numeración sirve para saber cuántos textos hay y para ordenarlos por

antigüedad de recopilación. Se numerarán igual tanto el texto original como su correspondiente traducción con objeto de que puedan gestionarse en paralelo.

2. TO o TM: «TO», es decir, «Texto Original», indica que el texto es original, mientras que «TM», «Texto Meta», expresa que el texto es traducido.
3. ES y EN: «ES» refleja que los textos están escritos en español, mientras que «EN» significa que los textos están escritos en inglés.
4. RE: indica que son resúmenes.

La codificación puede realizarse manualmente o bien, automatizarse, gracias a programas como LupasRename,⁷ lo que supone un ahorro considerable de tiempo.

Durante el quinto paso, el corpus debe almacenarse convenientemente en carpetas y subcarpetas que permitan su correcta localización e, incluso, futura ampliación. Así, se ha decidido crear una carpeta principal que indique la temática del corpus de bitextos —corpus de resúmenes de Neurología— que contiene dos subcarpetas que expresan el idioma de los mismos —español e inglés— y, del mismo modo, en su interior, se han creado otras dos carpetas que indican el formato de los documentos —PDF y texto plano ASCII— (cfr. Figura 4):

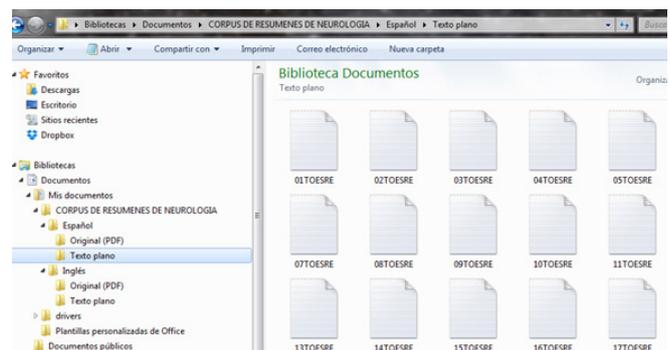


Figura 4. Codificación y almacenamiento del subcorpus en español

Cuando ya se han ejecutado los cinco pasos (localización de bitextos, descarga, conversión de formato, codificación y almacenamiento) tenemos como resultado un corpus de bitextos (español e inglés) compuesto por 200⁸ textos (100 resúmenes originales en español y sus correspondientes 100 traducciones al inglés,) que constan de un resumen o *abstract* de artículos de Neurología, así como sus palabras clave.

Finalmente, el último paso consiste en la alineación. Así, los documentos deben alinearse para poder procesarse sin errores por programas de gestión de corpus. La alineación de los textos facilita que se puedan establecer correspondencias entre los párrafos de los resúmenes recopilados. En la actualidad existen muchos programas (libres o comerciales) que pueden realizar esta tarea. Nosotros hemos elegido LF Aligner,⁹ gratuito, que permite crear memorias de traducción y corpus textuales, así como su exportación en formato de memoria de traducción (Translation Memory eXchange, .tmx) o en formato de texto plano (.txt) con codificación UTF-8. El programa

es muy intuitivo y su uso es relativamente sencillo: hay que especificar el formato y la codificación de los textos (en este caso, texto plano con codificación UTF-8), el par de lenguas de los textos que se van a alinear y, por último, seleccionar los documentos que se quieren alinear (cfr: Figura 5).

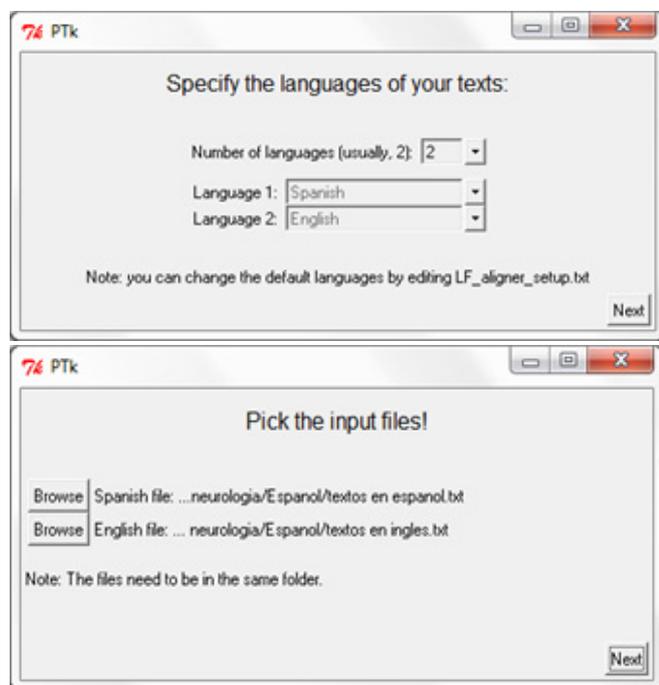


Figura 5. Interfaz de usuario de LF Aligner

Una vez que se ha acometido la alineación, LF Aligner muestra cuántas líneas se han creado para cada idioma y ofrece la opción de revisar la alineación de modo manual, ya sea con el editor gráfico del programa o mediante un archivo de Excel (.xls). Tras llevar a cabo la revisión con el editor gráfico, se muestran los textos alineados por frases, como se puede apreciar en la figura 6.

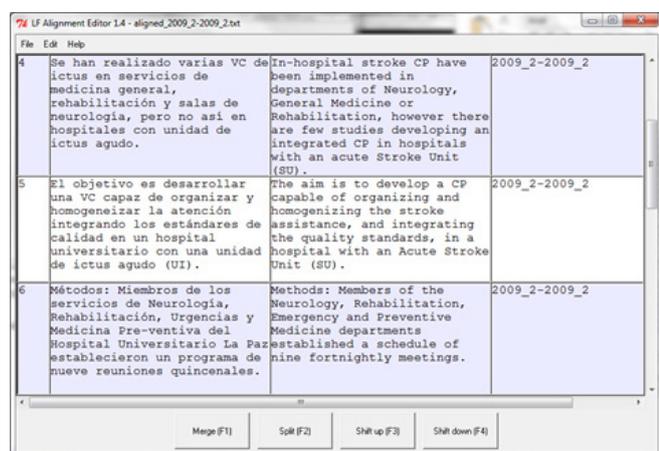


Figura 6. Alineación de los resúmenes de artículos científicos de Neurología (español-inglés) con LF Aligner

Durante la revisión es posible unir o separar los párrafos (*Merge* y *Split*), así como desplazarlos hacia arriba (*Shift up*)

o hacia abajo (*Shift down*), como se puede ver en los cuatro botones inferiores de la pantalla principal (cfr: Figura 6). Del mismo modo, en el menú desplegable *Edit* se pueden ejecutar diferentes acciones para editar la alineación, como cambiar el orden de las columnas, borrar las celdas que estén vacías y modificar el texto original, en caso de que se detecte algún error, entre otras opciones.

Una vez que se ha finalizado la etapa de revisión y edición de la alineación del texto original y su traducción, el programa ofrece la opción de crear un archivo con extensión .tmx, es decir, una memoria de traducción que se puede utilizar en programas TAO (Traducción Asistida por Ordenador). Asimismo, es posible extraer los documentos alineados con formato de texto plano con codificación UTF-8, que serán los que utilizaremos para la gestión con el programa de concordancias (cfr: Figura 7).

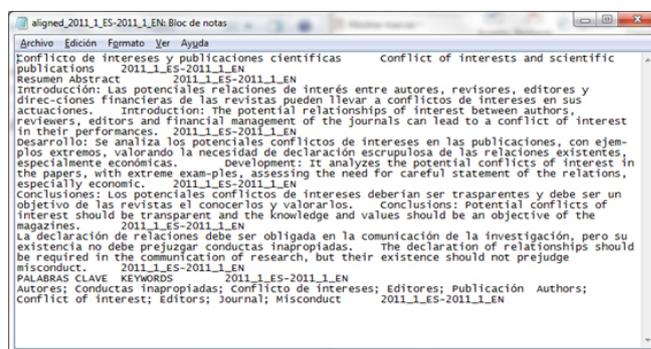


Figura 7. Texto alineado con LF Aligner

4. Extracción terminológica basada en bitextos para la elaboración de un glosario de Neurología

Para la extracción terminológica basada en bitextos con objeto de elaborar un glosario bilingüe y bidireccional (inglés-español/español-inglés) de Neurología recurriremos al programa LexTerm, por tratarse de un programa de extracción automática de terminología y equivalentes de traducción que se basa en técnica estadística, así como por ser gratuito, de libre distribución y de código abierto (Oliver *et al.* 2007).

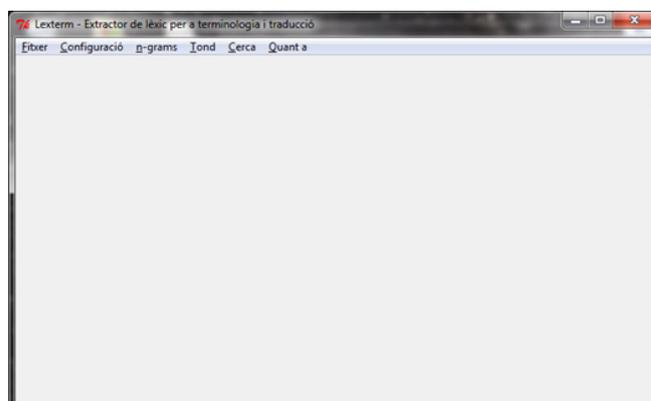


Figura 8. Interfaz de usuario y barra de herramientas de LexTerm

4.1. Extracción de candidatos a término y de equivalentes de traducción

Para empezar a utilizar LexTerm, en primer lugar, se procederá a abrir el corpus de bitextos. Para ello, se debe elegir la opción *Fitxer* para que se despliegue un menú y, en él, hay que seleccionar la opción *Obre corpus paral·lel*; o bien, utilizar el atajo de teclado «Ctrl + p».

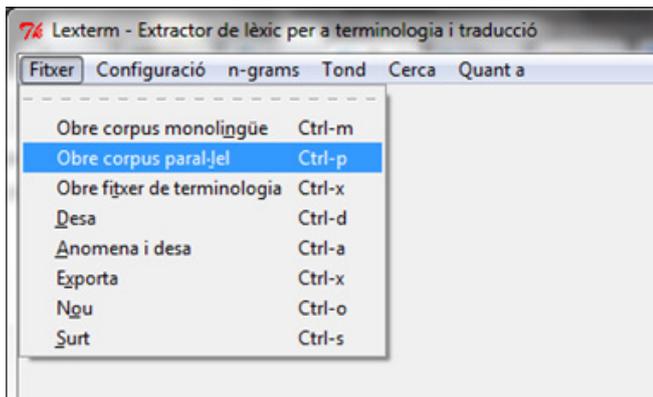


Figura 9. Menú desplegable de LexTerm

Después de subir el documento alineado que se quiere gestionar, y de configurar el programa de acuerdo a nuestras necesidades, podemos comenzar a extraer los términos del texto de origen (español) con la opción *n-grams* de la barra de herramientas (cfr. Figura 9). Al seleccionar esa opción, se cargarán un máximo de 250 términos (cfr. Figura 10):

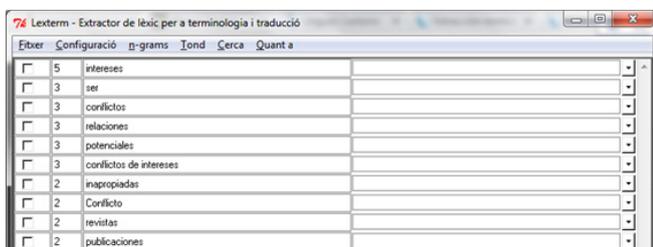


Figura 10. N-gramas en LexTerm

Debido a que al principio seleccionamos un máximo de 3-gramas, el número máximo de palabras combinadas será de tres. En la columna izquierda de la interfaz se muestran los candidatos a término del texto seleccionado, resultado de la búsqueda automática de n-gramas, y el programa los ordena por frecuencia de aparición, es decir, en primer lugar, se muestra el n-grama que más veces se repite en un texto, mientras que en el último lugar se puede ver el n-grama menos frecuente. El número a la izquierda del candidato a término indica el número de veces que dicho término aparece en el texto (cfr. Figura 10). En el caso de este corpus, la frecuencia de las palabras nunca será demasiado alta debido a que los textos se tratan de resúmenes y no de artículos completos. Del mismo modo, si no se hubiera utilizado una lista de palabras vacías, las palabras que aparecerían con más frecuencia serían aquellas que no aportan significado (cfr. Figura 11).

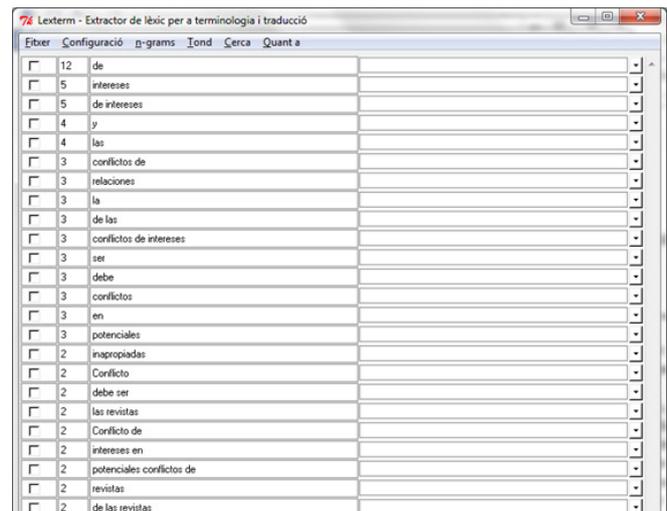


Figura 11. N-gramas sin stop-words en LexTerm

Después de que LexTerm haya extraído de forma automática los candidatos a término, continuamos con la búsqueda de equivalentes de traducción. Para buscar los equivalentes de traducción, se tiene que situar el cursor en la celda del término que se quiera traducir, hacer clic en ella para que el cursor no se mueva y, a continuación, seleccionar *Tond* en la barra de herramientas o bien utilizar el atajo de teclado «Alt + t». Cuando se realiza esta acción, el equivalente más probable aparece de inmediato, aunque podemos elegir otros posibles equivalentes si hacemos clic en el botón que se encuentra a la derecha de la celda del equivalente (cfr. Figura 12).

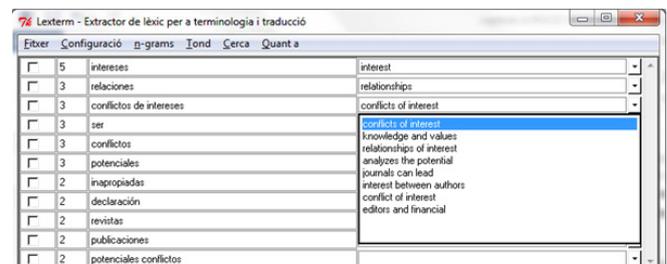


Figura 12. Equivalentes de traducción en LexTerm

Tras seleccionar ese botón, se despliega un cuadro donde se pueden ver otros posibles equivalentes de traducción. En el caso de que el equivalente que queramos seleccionar no se encuentre en ese cuadro, también podríamos escribirlo directamente, de forma manual, en la celda del equivalente.

En algunos casos, es evidente qué candidatos a términos presentados por el programa vamos a elegir para el glosario, así como qué equivalentes de traducción son los correctos. Sin embargo, también hay ocasiones en las que el intérprete puede tener dudas, bien porque desconoce el candidato a término, o bien porque hay más de un equivalente que parece correcto. Cuando esto ocurre, el intérprete puede dirigirse a los textos para ver el contexto en el que aparece el candidato a término o el equivalente. Esta acción se puede realizar desde el mismo programa. Así, si queremos conocer el contexto de

«intereses», por ejemplo, basta con dejar de nuevo el cursor sobre esa celda y seleccionar la opción *Cerca*.

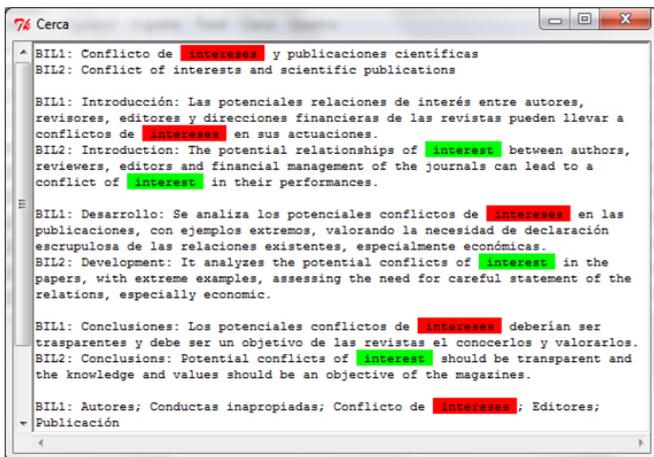


Figura 13. Equivalentes de traducción en LexTerm

Cuando se busca el contexto de los términos, aparece una nueva ventana donde se puede observar la palabra elegida en todos los párrafos del texto en los que aparece. De esta manera, después de la etiqueta «BIL1» aparecerá el párrafo del texto origen en español, y después de «BIL2» aparecerá el párrafo del texto traducido al inglés (cfr: Figura 13).

4.2. Generación del glosario bilingüe y bidireccional de Neurología

Una vez que se han extraído las listas de los candidatos a términos, y que se han seleccionado sus equivalentes de traducción, procedemos a la generación del glosario. En primer lugar, procedemos a seleccionar los términos definitivos y sus equivalentes. Así, descartaremos las palabras vacías que no haya filtrado la lista de *stop-words*, del mismo modo que no incluiremos las palabras generales que no tengan un significado específico en los resúmenes de artículos especializados, como «publicación», «aceptación» o «edición», por ejemplo. Para elegir qué términos se desean incluir en el glosario se debe seleccionar la casilla que se encuentra a la izquierda del número que indica la frecuencia de aparición de las palabras (cfr: Figura 14).



Figura 14. Selección de términos para el glosario

El siguiente paso consiste en exportar los términos escogidos que conformarán finalmente el glosario. Para ello, abriremos de nuevo el menú desplegable *Fitxer* y elegiremos *Exporta*, o bien utilizaremos el atajo de teclado «Ctrl + x» (cfr: Figura 15).

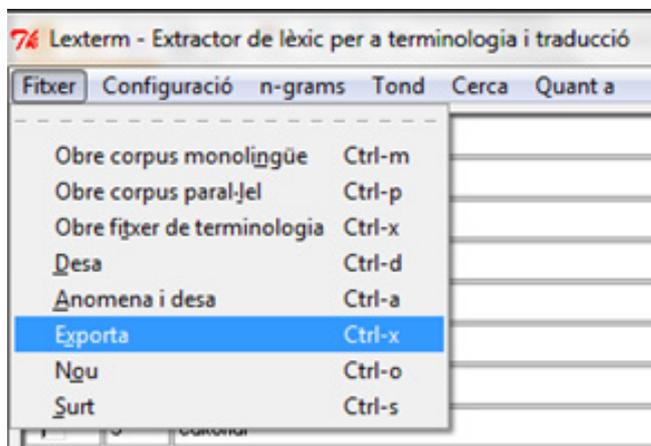


Figura 15. Menú desplegable de LexTerm

Se abrirá un explorador que permitirá elegir la ubicación del glosario generado, ponerle un nombre y guardarlo en formato de texto plano. Tras llevar a cabo estos pasos, el resultado es un documento en el que aparecen los términos en español a la izquierda y sus equivalentes de traducción en inglés a la derecha, separados por un tabulador, tal y como se puede ver en la figura 16.

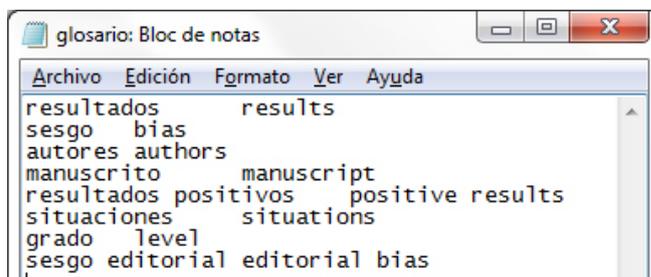


Figura 16. Términos exportados de LexTerm

Se aprovechará el texto plano con separación por tabuladores que ha creado LexTerm para copiar los términos y sus equivalentes y pegarlos en una hoja del procesador Microsoft Excel. Una vez copiado el glosario en Microsoft Excel, el intérprete puede agregar, si lo desea, otros campos, como el año de publicación del artículo, el autor o autores, la referencia completa del artículo, una pequeña definición del término, una frase en la que aparezca para contextualizar y la parte del artículo donde se encuentra el término (si se hubieran utilizado los artículos completos y no solo los resúmenes), entre otros muchos campos que podrían facilitar más información útil a la hora de utilizar un glosario.

En el caso de que se hayan generado varios glosarios, a partir de múltiples documentos alineados, deben pegarse todos los glosarios individuales en la hoja de Excel y, seguidamente, or-

ganizarlos automáticamente por orden alfabético descendente (de la A a la Z): esta acción tan sencilla hará que se puedan buscar y encontrar los términos deseados con más facilidad.

1	ESPAÑOL	INGLÉS
2	AAC	CAA
3	acetilcolina	acetylcholine
4	ácido valproico	valproic acid
5	actividad	activity
6	agudo	acute
7	AIT	TIA
8	alteración	abnormality
9	Alucinación	hallucination
10	ambulatoria	ambulatory
11	amígdala extendida medial	medial extended amygdala
12	amígdala	amygdala
13	amusia	amusia
14	angiopatía amiloidea cerebral	cerebral amyloid angiopathy
15	antibioterapia	antibiotherapy
16	anticuerpos IgG-NMO	NMO-IgG antibodies
17	anticuerpos	antibodies
18	antiepiléptico	antiepileptic

Figura 17. Glosario español-inglés ordenado alfabéticamente.

Hemos obtenido como resultado un glosario bilingüe y monodireccional (español-inglés) a partir del corpus de bitextos alineado con LF Aligner y gestionado con LexTerm (cfr: Figura 17). Para obtener el glosario inglés-español, basta con cambiar el orden de las columnas en Excel; esta acción se puede hacer seleccionando la columna donde se encuentran los equivalentes de traducción en inglés para cortarla, después se debe hacer clic con el botón derecho del ratón sobre la columna A (donde se encuentran los términos en español) y elegir la opción «Insertar celdas cortadas», de este modo en la columna de la izquierda se encontrarán los términos en inglés y a la derecha en español. Seguidamente, tal y como se hizo con el glosario español-inglés, ordenaremos alfabéticamente las entradas para que sea más fácil localizar los términos.

1	INGLÉS	ESPAÑOL
2	A	EA
3	abnormality	alteración
4	acetylcholine	acetilcolina
5	activity	actividad
6	acute	agudo
7	AD	DCL
8	AD	DTA
9	AED	FAE
10	ageing	envejecimiento
11	ALS	ELA
12	Alzheimer's disease	enfermedad de Alzheimer
13	Alzheimer-type dementia	demencia tipo Alzheimer
14	ambulatory	ambulatoria
15	amusia	amusia
16	amygdala	amígdala
17	amyotrophic lateral sclerosis	Esclerosis lateral amiotrófica

Figura 18. Glosario inglés-español ordenado alfabéticamente

Tras realizar estas acciones, hemos obtenido también un glosario bilingüe (inglés-español) a partir del corpus de bitextos creado (cfr: Figura 18).

El resultado final es un glosario bilingüe y bidireccional (español-inglés/inglés-español) de Neurología compuesto por un total de 283 términos en cada lengua, que se encuentra disponible para su consulta en la siguiente dirección URL: <https://goo.gl/F7PNa7>.

5. Conclusiones

En el presente trabajo hemos pretendido aplicar los beneficios de la lingüística de corpus a la interpretación. De este modo, hemos ilustrado una metodología protocolizada para la extracción terminológica basada en bitextos con objeto de generar un glosario, recurso predilecto por excelencia de los intérpretes, tanto en la fase documental previa a la interpretación como durante el proceso de interpretación mismo, tal y como puso de manifiesto la encuesta realizada por Corpas Pastor y Fern (2015) a intérpretes. Para ello, se ha tenido en cuenta, en primer término, el diseño del corpus de bitextos (español-inglés) y, seguidamente, se ha aplicado un protocolo de compilación con seis pasos, a saber, localización de bitextos, descarga, conversión de formato, codificación, almacenamiento y alineación, este último paso llevado a cabo con el programa gratuito LF Aligner. El resultado ha sido un corpus de bitextos monodireccional (español-inglés) de resúmenes de artículos de investigación de Neurología con 100 resúmenes en español y sus correspondientes 100 traducciones al inglés. La extracción terminológica bilingüe y la exportación de los textos candidatos a términos para la generación del glosario a partir del corpus de bitextos se ha ejecutado con el programa, también gratuito, LexTerm. El resultado es un glosario bilingüe y bidireccional (español-inglés/inglés-español) de Neurología, disponible en red, integrado por 283 términos en cada lengua, que puede resultar de gran utilidad para la fase documental previa del intérprete así como una extraordinaria herramienta de consulta durante el proceso de interpretación. La metodología para la generación de glosarios de forma semiautomática aquí ilustrada es totalmente gratuita, a la par que muy rápida una vez interiorizada, y puede aplicarse a cualquier temática y par de lenguas objeto de interpretación.

Notas

1. El presente trabajo se enmarca en el seno de los proyectos VIP (Sistema integrado Voz-texto para IntérPretes, ref. FFI2016-75831) e INTERPRETA 2.0 (aplicación de herramientas TIC para la enseñanza-aprendizaje de la interpretación, PIE17-015), así como parcialmente en los proyectos INTELITERM (ref. FFI2012-38881), TERMITUR (ref. HUM-2754) y POSTrad (ref. PIE 109, 2017-2018), además de la red temática TRAJUTEC, coordinada por la Universidad de Málaga.
2. La revista electrónica *Neurología* se encuentra disponible en la siguiente dirección URL: <https://www.neurologia.com/revista>.
3. El programa HTTrack puede descargarse en la siguiente dirección URL: <http://www.httrack.com/page/20/en/index.html>.

4. Online-Converter puede descargarse en la siguiente dirección URL: <https://documento.online-convert.com/es/convertir-a-txt>.
 5. Se puede descargar una versión gratuita del programa en la siguiente URL: <http://unipdf-converter.softonic.com>.
 6. El programa Scleaner, ideado por Hernani Costa, Gloria Corpas y Miriam Seghiri, se encuentra accesible a través de la siguiente dirección URL: <http://www.lexytrad.es/es/recursos/scleaner>.
 7. Se puede descargar el programa gratuitamente desde la siguiente dirección: <http://rename.lupasfreeware.org>.
 8. El corpus compilado, así, es representativo a nivel cualitativo, gracias al protocolo de diseño y compilación seguidos; por lo que se refiere a la representatividad cuantitativa, se eligió descargar 200 documentos por la limitada extensión que, por lo general, presentan los resúmenes de los artículos científicos y porque, al automatizar el proceso, no suponía un gran consumo de tiempo. Asimismo, se ha podido comprobar que el corpus es representativo a nivel cuantitativo gracias a la aplicación informática ReCor (*cf.* Seghiri 2006, 2014 y 2015). ReCor supone una solución eficaz para determinar *a posteriori*, por primera vez, el tamaño mínimo de un corpus o colección textual, independientemente de la lengua o género textual de dicha colección, estableciendo, por tanto, el umbral mínimo de representatividad a través de un algoritmo (N-Cor) de análisis de la densidad léxica en función del aumento incremental del corpus. El programa puede solicitarse de manera gratuita a través de una solicitud por correo electrónico a: alinares@uma.es.
 9. El programa para alineación de bitextos LF Aligner puede descargarse gratuitamente en la siguiente dirección URL: <https://lf-aligner.soft112.com>.
- Referencias bibliográficas**
- Atkins, Sue; Jeremy Clear y Nicholas Ostler (1991): «Corpus Design Criteria», *Literary and Linguistic Computing*, 7 (1): 1-16.
- Ballester, Leonor y José Antonio Capdevila (2009): «Meningitis bacteriana aguda en adultos», en José Antonio Capdevila (ed.): *Protocolos: Enfermedades infecciosas*. Barcelona: Elsevier.
- Bowker, Lynne (1998): «Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study» *Meta*, 43 (4): 631-651.
- Bradley, Walter G; Robert B. Daroff, Gerald M. Fenichel y Joseph Jankovic (2010): *Neurología clínica*. Barcelona: Elsevier.
- Castillo Rodríguez, Cristina (2009): «La elaboración de un corpus ad hoc paralelo multilingüe», *Revista Tradumática*, 7: 1-11.
- Choi, Jungwha (2005): «Qualité et préparation de l'interprétation. Évolution des modes de préparation et rôle de l'Internet», *Meta*, 50 (4): 1-15.
- Consejo Nacional de Especialidades Médicas (1996): *Guía de formación de especialistas: Neurología*. Madrid: Ministerio de Sanidad y Consumo y Ministerio de Educación y Cultura. <<http://www.msssi.gob.es/profesionales/formacion/docs/Neurologia.pdf>> [consulta 5.XI.2017].
- Corpas Pastor, Gloria (2001): «Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada», *TRANS. Revista de Traductología*, 5: 155-184.
- Corpas Pastor, Gloria y Lily Fern (2016): «A survey of interpreters' needs and practices related to language technology». Documento técnico [FFI2012-38881-MINECO/TI-DT-2016-1]. Málaga: Universidad de Málaga.
- Corpas Pastor, Gloria y Miriam Seghiri (eds.) (2016): *Corpus-based Approaches to Translation and Interpreting: From Theory to Applications*. Frankfurt: Peter Lang.
- Erazo-Torricelli, Ricardo (2013): «Miopatías estructurales congénitas», *Neurología*, 57 (Supl. 1): S53-S64.
- Gallego, Daniel y Miguel Tolosa (2012): «Terminología bilingüe y documentación *ad hoc* para intérpretes de conferencias. Una aproximación metodológica basada en corpus», *Estudios de traducción*, 2: 33-46.
- Gile, Daniel (2009): *Basic concepts and models for interpreting and translator training*. Amsterdam: John Benjamins.
- Harris, Brian. (1988): «Bitext, a New Concept in Translation Theory. Working Papers on Bilingualism», *Language Monthly*, 54: 810.
- Lievens, Jan. (2015): *Interpretación sanitaria: un glosario sobre infecciones de transmisión sexual*. Trabajo fin de Máster. Gante: Universidad de Gante.
- MedlinePlus (2013): *Enfermedades neurológicas*. <<http://www.nlm.nih.gov/medlineplus/spanish/neurologicdiseases.html>> [consulta 10.XI.2017].
- Oliver, Antoni; Mercè Vázquez y Joaquim Moré (2007): «Linguoc LexTerm: una herramienta de extracción automática de terminología gratuita», *Translation Journal*, 4 (11). <<http://translationjournal.net/journal/42linguoc.htm>> [consulta 10.XI.2017].
- Organización Mundial de la Salud (2010). *International Statistical Classification of Diseases and Related Health Problems*. 10th Revision. <http://apps.who.int/classifications/icd10/browse/Content/statchtml/ICD10Volume2_en_2010.pdf> [consulta 1.XI.2017].
- Organización Mundial de la Salud (2012): *Clasificación Internacional de Enfermedades. Modificación Clínica (CIE-9-MC)*. 8.ª edición. <https://www.msssi.gob.es/estadEstudios/estadisticas/docs/CIE9MC_8ed.pdf> [consulta 1.XI.2017].
- Pérez Abril, Juan (2015): «La demanda de plazas de Neurología ha crecido un 86% entre los 500 primeros números del MIR», *IM-MédicoHospitalario*, 17 abril. <<http://www.immedicohospitalario.es/noticia/5278/la-demanda-de-plazas-de-neurologia-ha-crecido-un-86-entre-los-500-primeros-numeros-del-mir>> [consulta 14.XI.2017].
- Pérez Pérez, Pablo Salvador (2013): «La enseñanza de la interpretación de conferencias: un estudio empírico con metodología de corpus». Tesis doctoral. Málaga: Universidad de Málaga.
- Postigo, Encarnación y Marina Calleja (2017): «A model to Enhance Interaction for People with Severe Intellectual Disability in Healthcare, Education and Interpreting», *Procedia*, 237 (21): 1189-1195.
- Postigo, Encarnación y Concepción Mira (2017): «Translation, Virtual Environments and ICT Tools for Achieving Competence in Language for Specific Purposes», en Rosa Muñoz-Luna y Lidia Taillefer (eds.) *Integrating Information and Communication Technologies in English for Specific Purposes*. Berlín: Springer, pp. 145-163.
- Sánchez Ramos, M.ª Mar (2016): «Community Healthcare Translator Training and Ad Hoc Corpora», *Current Trends in Translation Teaching and Learning*, 3: 119-139.
- Sánchez Ramos, M.ª Mar y Francisco Javier Vigier Moreno (2016): «Using Monolingual Virtual Corpora in Public Service Legal Translator Training», en Elena Martín-Monje, Izaskun Elorza, Blanca García Riaza (eds.), *Technological Advances in Specialized Linguistic Domains. Practical Applications and Mobility*. Londres/Nueva York: Routledge, pp. 228-239.

- Seghiri, Miriam (2006): *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Málaga: SPICUM.
- Seghiri, Miriam (2011): «Metodología protocolizada de compilación de un corpus de seguros de viaje: aspecto de diseño y representatividad», *Revista de lingüística teórica y aplicada*, 49 (2): 13-30.
- Seghiri, Miriam (2014): «Too Big or Not Too Big: Establishing the Minimum Size for a Legal Ad Hoc Corpus», *Hermes*, 53: 85-98.
- Seghiri, Miriam (2015): «Determinación de la representatividad cuantitativa de un corpus ad hoc bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos/Establishing the quantitative representativeness of an E-Reader User's Guide ad hoc corpus (English-Spanish)», en María Teresa Sánchez Nieto (ed.): *Corpus-based Translation and Interpreting Studies: From description to application*. Berlín: Frank & Timme, pp. 125- 146.
- Seghiri, Miriam (2017): «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores», *Babel*, 63 (1), pp. 43–64. DOI 10.1075/babel.63.1.04seg.
- Sociedad Española de Neurología (2010): ¿Qué es la neurología? Madrid: Sociedad Española de Neurología. http://www.sen.es/pdf/2010/que_es_la_neurologia.pdf [consulta 5.XI.2017].
- Tengku, Sepoa; Helia Vaezian y Mahmoud Akbari (2010): *Corpora in Translation*. Berna: Peter Lang.
- UNESCO (1983): *Guía para la redacción de artículos científicos destinados a la publicación*. París: UNESCO.
- Varantola, Krista (2000): «Translators, dictionaries and text corpora», en Silvia Bernardini y Federico Zanettin (eds.): *I corpora nella didattica della traduzione*. Boloña: CLUEB, pp. 117-133.
- Vigier Moreno, Francisco Javier (2016): «Teaching the Use of ad hoc Corpora in the Translation of Legal Texts into the Second Language», *Language and Law/Linguagem e Direito*, 3 (1): 100-113.
- Zaretskaya, Anna; Gloria Corpas Pastor y Miriam Seghiri (2016): «Corpora in computer-assisted translation: a users' view», en Gloria Corpas Pastor y Miriam Seghiri (eds.): *Corpus-based Approaches to Translation and Interpreting: From Theory to Applications*. Frankfurt: Peter Lang, pp. 253-276.
- Zarranz, Juan (2007): *Neurología*. Barcelona: Elsevier.

