

El Registro Público de Documentos del PE como recurso documental en la traducción especializada: Elaboración de bases de datos terminológicas con corpus en Multiterm

(The European Parliament's Register of Documents as a documentary resource in specialized translation: Using a corpus of texts to build a Multiterm terminology database)

MARÍA AZAHARA VEROZ GONZÁLEZ
z92vegom@uco.es

HUM887: Estudios de Traductología, Discurso y Cognición

Fecha de recepción: 12 de agosto de 2014
Fecha de aceptación: 1 de noviembre de 2014

Resumen: En este trabajo nos ocuparemos de las posibilidades que ofrece la enseñanza de la metodología propuesta por la lingüística de corpus en la formación de futuros traductores para la elaboración de bases de datos terminológicas a través del Registro Público de Documentos del Parlamento Europeo como recurso documental multilingüe para la traducción especializada. A lo largo de este trabajo explicaremos de manera detallada la propuesta de elaboración de bases de datos multilingües en Multiterm utilizada en la asignatura de Herramientas Profesionales para la Traducción de segundo curso en el Grado de Traducción e Interpretación de la Universidad de Córdoba.

Palabras clave: Registro Público de Documentos del Parlamento, Lingüística de corpus, Traducción, Bases de datos terminológicas, Glosarios multilingües, Herramientas Profesionales para la Traducción, Multiterm.

Abstract: In this paper we tackle the methodology proposed by corpus linguistics in the training of future translators for the elaboration of multilingual terminology databases through the Public Registry of Documents of the European Parliament as a multilingual documentary resource for specialized translation. Throughout this paper, we will explain in depth how futures translators can elaborate a multilingual terminology database in SDL Multiterm using this methodology in the subject Professional Tools for the Translation in Translation and Interpretation Degree at the University of Córdoba.

Key words: Public Registry of Documents of the European Parliament, multilingual terminology databases, multilingual glossaries, Professional Tools for the Translation, Multiterm.

1. ENFOQUE TEÓRICO: LA TRADUCCIÓN Y LA LINGÜÍSTICA DE CORPUS

El desarrollo de la lingüística de corpus ha revolucionado, sobre todo desde su resurgimiento en los años ochenta, la investigación en otras disciplinas como la enseñanza de idiomas, lexicografía, procesamiento del lenguaje natural, sociolingüística, psicolingüística y, como no, traducción (Baker, 1993; Corpas Pastor, 1995, 1997a, Laviosa 1998a) tal y como describe Kohn (1996: 48) (a través de Corpas Pastor, 2001):

This new linguistic discipline [=corpus linguistics] represents for Translation Studies the basis on which a complex and coherent theory, integrating the different approaches (descriptive and evaluative, literary and linguistic, etc., considered not opposed but complementary), could be elaborated, a theory accepted equally by both translator scholars and (which would be a premiere) by the community of professional translators.

Su introducción en traductología no se produjo hasta los años noventa (Baker, 1993), siendo esta autora la pionera en introducir las herramientas de análisis de la lingüística de corpus para investigar el producto y el proceso de la traducción desde un punto de vista descriptivo. Su evolución ha provocado grandes cambios metodológicos y teóricos (Laviosa, 2002). Este desarrollo proviene del afán por fomentar el empirismo y observar datos reales frente a la introspección de épocas pasadas (Corpas Pastor, 2001). El impacto que ha sufrido la traductología como disciplina, unida a los avances habidos en el ámbito de la informática, ha sido y es innegable (Tymozco, 1998).

Baker predijo que la disponibilidad de grandes corpora, tanto de textos originales como de textos traducidos unida al desarrollo de la metodología de corpus permitiría a los investigadores descubrir «the nature of translated texts as a mediated communicative event» (Baker, 1993: 243) a través de la investigación de lo conocido en la literatura como los universales de la traducción (Laviosa, 2002: 18). Cuando hablamos de universales nos referimos a los que suceden de manera habitual en las traducciones más que en los originales, y que es independiente de la influencia de las lenguas involucradas en el proceso traductor. Estos universales a los que se refiere Baker son:

Explicitación, en lo que se refiere a las formas de cohesión (Blum-Kulka, 1986), inserción de la información adicional en el texto meta y a nivel sintáctico en mayor redundancia (Baker, 1992).
 Desambiguación y simplificación (Vanderauwera, 1985: 97).
 Convencionalidad textual en traducción escritas (Vanderauwera, 1985) e interpretación (Shlesinger, 1991).
 Evitar repeticiones presentes en el texto origen (Shlesinger, 1991; Toury, 1991a).
 Tendencia a representar de manera excesiva características del texto meta (Toury, 1980, Vanderauwera, 1985).
 Distribución específica de elementos léxicos de cara a los textos originales en la lengua meta (Shamaa, 1978) (Traducción de Laviosa, 2002:18).

Desde las primeras propuestas, el estudio de la traducción a través del corpus ha suscitado diversos debates. En 1996, Baker discute sobre tres aspectos fundamentales del *Corpus-based Translation Studies*, de ahora en adelante CBT:

Its theoretical links with target-oriented approaches, the unique methodology it employs and the potential of this methodology for investigating the distinctive linguistic nature of translation, conceived as a distinctive communicative event, shaped by its own goals, pressures and contexts of production (Baker, 1994:175; *apud* Laviosa, 2002: 23).

En lo que se refiere a la aplicación del CBT como metodología, Baker subraya la importancia en la elaboración de unos criterios e hipótesis que definan al corpus acorde con las necesidades específicas del estudio.

Así pues, hoy por hoy, podemos afirmar que los estudios de traducción basados en corpus constituyen un área de investigación empírica en traductología:

Within a wider empirical paradigm, corpus based studies for translations have given rise to a coherent and distinct body of research in both branches of the discipline, pure, and applied (Laviosa, 2002:1).

Desde un punto de vista estricto, «a corpus is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language» (Sinclair, 1991: 171), es decir, que un «corpus ha de ser representativo del estado de la lengua o de la selección de lengua específicamente seleccionada» (Corpas, 2001: 156).

Tras el advenimiento de la informática a finales de la década de los ochenta, esta definición ha tomado connotaciones más específicas dentro

de la lingüística de corpus, introduciéndose el formato electrónico como requisito a la hora de analizar un corpus (Baker, 1996; Leech, 1992).

En esta misma línea encontramos las definiciones de Bowker y Pearson (2002: 9) en la que un corpus «a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria» y de Hunston (2002:2) «collections of texts (or parts of texts) that are stored and accessed electronically». Como podemos observar, en la actualidad, el formato digital es una necesidad que se hace patente en las propias definiciones de corpus (Johansson, 1998; McEnery & Wilson, 1996; Baker, 1995).

Según esto, deducimos que existen varias características que un corpus ha de poseer para poder ser considerado como tal (Bowker y Pearson, 2002):

-Real: los textos o los fragmentos de textos escogidos para el estudio han de ser textos naturales, entendemos como naturales: «an instance of language in use, either spoken or written: a piece of language behavior which has occurred naturally, without the intervention of the linguist» (Stubbs 1996: 4), este requisito se encuadra dentro de la herencia firhiana, en cuanto al estudio de la lengua en su uso (Firth, 1957).

-Electrónico: han de estar en formato electrónico para poder procesarlos posteriormente en un ordenador. Existen herramientas, de las que más adelante hablaremos, que nos permiten acceder a la información del corpus de forma rápida y eficaz. Si el corpus no estuviera en formato electrónico, tratándose de un texto escrito, tendríamos que leer todo el texto para poder extraer información, sin embargo, gracias a estas herramientas, podremos obtener la información de manera inmediata. Por otro lado, esto nos da la posibilidad de disponer de un corpus mucho más amplio ya que las herramientas informáticas pueden procesar más información en menos tiempo.

-Amplio: con respecto a esta característica no existen reglas específicas, todo dependerá de los datos que esperemos obtener a partir de nuestro corpus. Bowker y Pearson (2002:10) definen esta característica como «a greater number of texts than you would be able to easily collect and read in printed form».

-Criterios específicos: un corpus no es solo una colección de textos, sino que se trata de una colección de textos seleccionados según una serie de criterios que correspondan con una muestra representativa de la lengua y/o lenguaje dependiendo de aquello que se desee investigar.

-Cantidad: cuando nos referimos a la cantidad, estamos hablando de extensión del corpus en número de palabras, este criterio es igual que el del corpus amplio anteriormente citado.

-Calidad: con este requisito nos referimos a la representatividad y al equilibrio del corpus, está íntimamente ligado a la cantidad (*balanced corpus*). De acuerdo con Corpas Pastor (2001) la calidad se mide por los criterios de inclusión de los documentos que los componen (Alvar Ezquerro y Corpas Pastor, 1994), con el fin de garantizar la homogeneidad de la colección dentro de la heterogeneidad de los textos incluidos.

-Simplicidad de codificación, con ello se refiere a que el texto ha de estar codificado en caracteres ASCII, es decir, en formato de texto plano.

-Documentación: con ello nos referimos a la fuente exacta de las fuentes documentales utilizadas.

1.1. Tipos de corpus

A continuación pasaremos a realizar una revisión de los diferentes tipos de corpus para poder encuadrar el corpus que propondremos en el caso práctico. Huelga decir que no existe acuerdo en torno a la denominación y clasificación de los diferentes tipos de corpus probablemente debido a que son varias las disciplinas que utilizan esta metodología (McEnery y Xiao, 2008).

Bowker y Pearson (2002) afirman que existen tantos tipos de corpus como tipos de investigaciones. Como podemos ver, el establecimiento de una definición sobre las tipologías existentes de corpora es un concepto amplio y esto nos lleva a realizar una revisión de las propuestas que giran en torno a esta noción para poder situar nuestra investigación.

Existen diversas publicaciones que hacen referencia a los tipos de corpus existentes. Para enmarcar nuestro corpus hemos recurrido a la clasificación que realiza Sinclair en el informe técnico elaborado para EAGLES (1994). En este informe encontramos con cuatro tipologías:

-Corpus de referencia: definido por Sinclair en el informe EAGLES (1996) como:

A reference corpus is one that is designed to provide comprehensive information about a language. It aims to be large enough to represent all the relevant varieties of the language, and the characteristic vocabulary, so that it can be used as a basis for reliable grammars, dictionaries, thesauri and other language reference materials.

-Corpus monitor: este corpus mantiene un volumen constante pero actualizado. Es decir, se va añadiendo contenido a la vez que se elimina aquel que ya está en desuso o poco actualizado.

-Corpus paralelo: definido como «a collection of texts, each of which is translated into one or more other languages than the original» (Sinclair; 1996).

-Corpus comparable: como «one which selects similar texts in more than one language or variety» (Sinclair; 1996).

Este informe se realizó dentro del ámbito de la Lingüística. Pero dentro del ámbito de la Traductología nos encontramos con dos clasificaciones que cabe destacar, la realizada por Corpas Pastor (2001), que parte de la propuesta realizada por Torruella y Llisteri (1999), completada con las propuestas Johansson (1998) y de Baker (1995) para los corpora multilingües, y la propuesta realizada por Laviosa (2002).

1.2. Aplicación del corpus en bases de datos para la traducción asistida por ordenador

Por regla general, los sistemas de traducción asistida no suelen utilizarse de manera aislada sino que suelen utilizarse junto con glosarios, diccionarios, bases de datos terminológicas y otros recursos documentales, primando hoy en día aquellos recursos que son electrónicos.

La lingüística de corpus se ocupa en este ámbito de la creación de estos recursos terminológicos utilizando herramientas que permiten la extracción terminológica. En el caso práctico propuesto utilizaremos tres herramientas que nos servirán para la extracción terminológica, búsqueda de equivalentes y gestión y recuperación de la terminología. En primer lugar, una vez descargado y tratado el corpus utilizaremos *AntConc*, se trata de un programa gestor de corpus, el cual nos permite extraer listados de palabras por frecuencias eliminando aquellos términos gramaticales o de uso muy general. *ParaConc* con el cual podremos realizar un análisis contrastivo del corpus paralelo bilingüe o multilingüe y podremos realizar una búsqueda de equivalentes. Y, por último, *Multiterm* de SDL, en el cual crearemos la base de datos terminológica y gracias al cual se podrá recuperar la información almacenada de manera aislada o conjuntamente con SDL *Trados* 2009/2011/2014 mientras se está traduciendo.

2. ENFOQUE PRÁCTICO: PROPUESTA DIDÁCTICA

En este caso nos centraremos principalmente en la compilación de un corpus *ad hoc* (Corpas, 2011) paralelo multilingüe extraído del Registro Público de Documentos del Parlamento Europeo. Este registro

Contiene las referencias de los documentos producidos o recibidos por la Institución a partir del 3 de diciembre de 2001, fecha de aplicación del Reglamento (CE) nº 1049/2001. En su mayor parte son directamente accesibles al público en versión electrónica. El acceso electrónico es gratuito y no requiere justificación particular (Parlamento Europeo, sine data).

Resulta relevante el hecho de que este recurso, a pesar de ser de gran utilidad para el traductor, no se cita habitualmente como tal, recurriendo normalmente a fuentes como Eurlex. Sin embargo, este registro incluye documentos de carácter técnico y administrativo, no solo legislativo, por lo que resulta bastante útil como obra de consulta en la traducción especializada, sobre todo porque se trata de un recurso multilingüe y se pueden obtener textos de forma paralela.

A continuación describiremos la propuesta didáctica que realizamos con los alumnos para la elaboración de bases de datos terminológicas con este recurso basándonos en la metodología propuesta por Corpas (2011) para la traducción inversa.

2.1. *Diseño de corpus*

Existen diferentes estudios que nos ofrecen pautas para diseñar un corpus (Aston y Burnard, 1998; Atkins et al. 1992; Kennedy, 1998; Sinclair, 1991), no obstante, nosotros nos centraremos en aquellas ofrecidas por Bowker y Pearson (2002, 45 y ss.), para luego aplicarlas directamente al ámbito de la traducción de textos especializados, en concreto en la elaboración de una base de datos terminológica en Multiterm.

-Tamaño del corpus: con el que nos referimos al número total de palabras y no al número de textos.

-Número de textos: Hasta ahora hemos hablado de tamaño del corpus en relación al número total de palabras, pero no es sólo la única cuestión que nos interesa saber a la hora de diseñar nuestro corpus, también nos interesa saber cuántos textos hemos de recopilar y cómo han de ser.

-Medio: En este caso nos referimos a si se trata de una transcripción de un texto oral o un texto escrito directamente, ya que el uso de la lengua es diferente.

-Tema: El tema constituye una de las partes fundamentales a la hora de compilar un corpus. Cabe destacar que existen temas que pertenecen a varios ámbitos, es decir, que son multidisciplinares y esto puede dificultar nuestra búsqueda. En nuestra propuesta limitaremos la búsqueda al ámbito de los presupuestos y las finanzas, al alumno se le propone trabajar con el ámbito que desee dentro de las políticas de la UE.

-Tipo de texto: Es necesario tener en cuenta el grado de especialización que buscamos cuando se compila un corpus, ya que no es lo mismo un texto escrito por un experto para una revista especializada que para un periódico, o un libro de texto, o un texto redactado por un estudiante.

-Autor y fecha de publicación: Con el fin de asegurar que nuestro corpus sea fiable, es necesario que la fuente, es decir, la web de la cual extraemos el texto tenga autoría y fecha de publicación actualizada, en el Registro Público de Documentos del Parlamento Europeo contamos con la fecha de

publicación, la autoría corresponde al Parlamento Europeo o a cualquier órgano del mismo o a los eurodiputados.

-Lengua: en nuestro caso se trata de un corpus multilingüe paralelo, el alumno elaborará la base de datos con sus lenguas de trabajo.

A continuación exponemos un cuadro con los criterios básicos de compilación del corpus (adaptación del cuadro propuesto por Bowker y Pearson, 2002: 54).

Diseño del corpus total

| | |
|----------------------|---|
| Tamaño | 100 000 <i>tokens aprox</i> |
| Número de textos | 60 aprox |
| Medio | Escrito |
| Tema | Presupuestos y finanzas |
| Autoría | Eurodiputados, Parlamento Europeo, órganos del Parlamento Europeo |
| Idioma | Español, inglés y francés |
| Fecha de publicación | Período 2010-2013 |

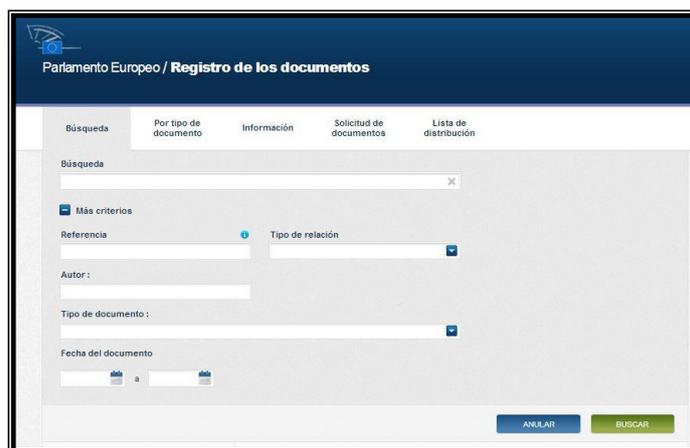
Elaboración propia a partir de la tabla presentada por Bowker y Pearson (2002: 54)

2.2. Descarga del corpus

Una vez diseñado el corpus y teniendo en cuenta las características que debía tener el mismo, nos dispondremos a recopilar los textos. En nuestro caso, debido a las características prediseñadas nos ceñiremos a la recopilación electrónica, en concreto a la búsqueda de textos en Internet, consultando el Registro Público del Parlamento Europeo¹.

¹ Véase: <http://www.europarl.europa.eu/RegistreWeb/search/simpleSearchHome.htm?language=ES>

Registro Público de documentos del Parlamento Europeo



Registro Público de los documentos del PE

Como se puede apreciar en la imagen, la página web dispone de un buscador avanzado, y gracias a él podremos seleccionar los tipos de documentos que nos interesan y el período en el que han sido creados. El resultado que obtendremos de la búsqueda, seleccionados estos dos criterios, será un listado de textos, del cual podremos descargar aquellos que nos interesen en las lenguas de estudio (inglés, francés y español). La descarga se puede realizar, dependiendo del texto en los siguientes formatos **.doc* o **.pdf*.

Estos textos los almacenaremos por tipo de documento, ya que el análisis se realizará siguiendo dos líneas, una línea centrada en el subgénero y otra centrada en la totalidad del corpus compilado.

2.3. Formateado del corpus

Una vez seleccionados y descargados los textos, procederemos a su formateado para poder procesarlos en las herramientas de análisis y comparación de textos.

Como hemos comentado, los textos descargados se encuentran dos formatos fundamentalmente **.doc* y **.pdf*. Estos textos están llenos de etiquetas que podrían causar problemas a la hora de analizar nuestro corpus con *AntConc* y con *ParaConc*. Para que esto no ocurra hemos de transformarlos en formato plano (**.txt*) en codificación ANSI. Para ello utilizaremos los programas *Pdf to Word Free* y *Zilla Word to Txt*.

Realizada esta conversión, prepararemos los textos para la alineación que realizaremos con *ParaConc* una vez extraídos los términos, ya que tendremos que asegurarnos que el programa de alineación haga corresponder cada segmento con su traducción, puesto que este tipo de programas (*ParaConc*) suelen dar por hecho de que los textos y sus traducciones tienen el mismo número de párrafos y de frases. Los textos emanados del Parlamento Europeo suelen seguir una estructura muy similar entre ellos y en la mayoría de las ocasiones suelen contener el mismo número de párrafos. No obstante, revisaremos cada uno de los párrafos para que las correspondencias que haga *ParaConc* no sean equivocadas. Dicha revisión se realizará con el programa *Notepad++*, que es un editor avanzado de texto plano.

Para su almacenamiento crearemos una carpeta llamada *Corpus_financieros*, dentro de ella crearemos tres carpetas correspondientes a cada uno de los idiomas estudiados (*EN* para inglés, *ES* para español y *FR* para francés) y dentro de cada una de ellas crearemos dos carpetas más una llamada *original* (en donde se guardarán los textos en formato **.pdf* y/o **.doc*) y otra llamada *texto* (en donde se guardarán los textos en formato **.txt*).

2.4. Elaboración de la ficha terminológica en Excel

Antes de proceder a la selección de la terminología hemos de diseñar la ficha terminológica y determinar qué campos (tanto índice como descriptivos) va a contener. En este caso, proponemos a los alumnos que se basen en el modelo de ficha establecido por el Centro de Terminología de Bruselas (CTB) para la elaboración de la ficha terminológica (*Cuadro 2*), aunque para simplificar el trabajo (ya que de lo que se trata es que el alumnado aprenda a realizar una base de datos terminológica de terminología especializada en un breve periodo de tiempo y que a su vez sea útil en futuras traducciones, hemos reducido los campos a los siguientes:

-Campos índice: entrada en español (Español), entrada en inglés (Inglés), entrada en francés (Francés).

-Campos descriptivos: contexto en español (*Contexto_Es*), contexto en inglés (*Contexto_En*), contexto en francés (*Contexto_Fr*), abreviatura en español (*Abrev_Es*), abreviatura en inglés (*Abrev_En*), abreviatura en francés (*Abrev_Fr*) y definición en español (*Definición*).

Para que la recopilación sea más sencilla y automatizada y su volcado más rápido en *Multiterm*, esta información la almacenaremos en un primer lugar en una hoja de cálculo de *Excel* para luego importarla en *Multiterm* (*Tabla 2*).

Modelo de ficha del Centro de Terminología de Bruselas

**MODELO DE FICHAS DEL
CENTRO DE TERMINOLOGÍA DE BRUSELAS (CTB)**

| | |
|---------------------------------|--|
| ES-EN-FR-AR | |
| CAMPO DE UTILIZACIÓN: | |
| SUBCAMPO DE UTILIZACIÓN: | |
| ÁMBITO DE APLICACIÓN: | |
| ENTRADA: | |
| DEFINICIÓN: | |
| CONTEXTO 1: | |
| FUENTE: | |
| CONTEXTO 2: | |
| FUENTE: | |
| VARIANTE: | |
| NOTA: | |
| TÉRMINO EQUIVALENTE EN INGLÉS: | |
| TÉRMINO EQUIVALENTE EN FRANCÉS: | |
| TÉRMINO EQUIVALENTE EN ÁRABE: | |

Centro de Terminología de Bruselas

Hoja de cálculo para recopilación de información

| | A | B | C | D | E | F | G | H | I | J |
|----|---------|-------------|----------|------------|--------|-------------|----------|---------|-------------|----------|
| 1 | Español | Contexto_Es | Abrev_Es | Definición | Inglés | Contexto_En | Abrev_En | Francés | Contexto_fr | Abrev_Fr |
| 2 | | | | | | | | | | |
| 3 | | | | | | | | | | |
| 4 | | | | | | | | | | |
| 5 | | | | | | | | | | |
| 6 | | | | | | | | | | |
| 7 | | | | | | | | | | |
| 8 | | | | | | | | | | |
| 9 | | | | | | | | | | |
| 10 | | | | | | | | | | |
| 11 | | | | | | | | | | |
| 12 | | | | | | | | | | |
| 13 | | | | | | | | | | |
| 14 | | | | | | | | | | |
| 15 | | | | | | | | | | |

2.5. Selección de la terminología y volcado de datos

Una vez seleccionado el corpus paralelo multilingüe y elaborada la hoja de cálculo, procederemos a extraer una lista de frecuencias mediante el programa de gestión de corpus *AntConc*. Este proceso lo realizaremos con el corpus en español únicamente, ya que luego, con el programa *ParaConc*² buscaremos los equivalentes. En la *Tabla 3* aparecen las 20 palabras (*tokens*) más usadas en orden decreciente de frecuencia, y en la *Tabla 4*, las 20 palabras más frecuentes tras haber descartado palabras gramaticales o de uso muy generalizado (Corpas, 2011).

Listado de frecuencias completo

| | | |
|---------------------------|------|----------|
| Total No. of Word Types: | | |
| 6 164 | | |
| Total No. of Word Tokens: | | |
| 95 517 | | |
| 1 | 9080 | de |
| 2 | 4186 | la |
| 3 | 2843 | el |
| 4 | 2732 | los |
| 5 | 2702 | en |
| 6 | 2662 | y |
| 7 | 2378 | a |
| 8 | 2115 | que |
| 9 | 1918 | las |
| 10 | 1704 | del |
| 11 | 1020 | se |
| 12 | 995 | con |
| 13 | 934 | al |
| 14 | 922 | Comisión |
| 15 | 867 | por |
| 16 | 843 | Artículo |
| 17 | 745 | para |
| 18 | 581 | o |
| 19 | 553 | un |
| 20 | 547 | artículo |

² En *ParaConc* se pueden extraer también listas bilingües de frecuencias, aunque debido a las omisiones o repeticiones, en ocasiones no coinciden en los diferentes idiomas, por eso proponemos una metodología diversa utilizando en primer lugar *AntConc*.

Listado de frecuencias tras extraer palabras gramaticales

| | | |
|------------------------------------|-----|-------------|
| Total No. of Word Types: 6021 | | |
| Total No. of Word Tokens: 48668 | | |
| 1 | 843 | Artículo |
| 2 | 448 | presupuesto |
| 3 | 409 | gestión |
| 4 | 316 | normas |
| 5 | 313 | ejercicio |
| 6 | 296 | Tribunal |
| 7 | 277 | apartado |
| 8 | 276 | Cuentas |
| 9 | 271 | Consejo |
| 10 | 268 | ejecución |
| 11 | 266 | créditos |
| 12 | 266 | Europeo |
| 13 | 247 | Reglamento |
| 14 | 202 | gastos |
| 15 | 191 | arreglo |
| 16 | 187 | relativas |
| 17 | 179 | cuentas |
| 18 | 177 | presente |
| 19 | 174 | conformidad |
| 20 | 169 | actos |
| 21 | 165 | ingresos |

De la última lista de frecuencias, el alumnado deberá seleccionar al menos 50 términos atendiendo no solo al criterio de la frecuencia, ya que ésta es únicamente un criterio orientativo, sino también al sentido común del alumno, ya que deberá seleccionar aquellos términos que considere más importantes o de mayor relevancia en el ámbito que está estudiando. Asimismo, deberá tener en cuenta los términos compuestos por más de una palabra, que correspondería a un segundo estadio de su investigación en el que deberá recurrir a su sentido común y a la función concordance de *AntConc*, sería el caso, por ejemplo, de *Tribunal*, que en la lista de

frecuencias aparece como tal, pero si buscamos en *concordance* observamos que se refiere al *Tribunal de Cuentas* y al *Tribunal de Justicia* (Cuadro 3), por lo que los términos que se introducirían en la base de datos serían *Tribunal de Cuentas* y *Tribunal de Justicia*. Estos términos se irán copiando en la hoja de cálculo de *Excel* explicada anteriormente.

Estudio de *Tribunal*

The screenshot displays the Concordance software interface. At the top, there are several tabs: Concordance, Concordance Plot, File View, Clusters, Collocates, Word List, and Keyword List. The main window shows a list of search results with columns for Hit and KWIC. The text in the KWIC column is a snippet from a document, mentioning the Tribunal de Cuentas Europeo and the Tribunal de Justicia. Below the list, there is a search control panel with the following options:

- Search Term: Words, Case, Regex
- Concordance Hits: 296
- Search Window Size: 50
- Buttons: Start, Stop, Sort
- Kwic Sort: Level 1 1R, Level 2 2R, Level 3 3R

Gracias a esta misma función de *concordance*, el alumno encontrará los contextos en español, que deberá de ir pegando en el apartado *Contexto_Es* de la hoja de cálculo, añadiendo tras el texto copiado la fuente de la cual lo ha extraído.

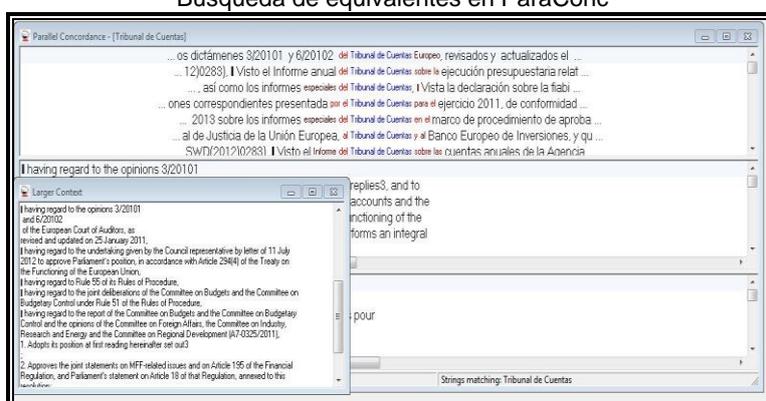
Una vez realizado esto, procederá a introducir los textos dentro del programa *ParaConc* (Cuadro 4) y utilizaremos la función *Search* (Ctrl+S) (Cuadro 5), para buscar los términos seleccionados en español y encontrar

sus equivalentes en inglés y en francés³ y sus respectivos contextos, así como las abreviaturas si las hubiere.

Introducción del corpus paralelo en ParaConc



Búsqueda de equivalentes en ParaConc



A la hora de copiar los contextos, siempre incluiremos la fuente, ya que en momentos determinados podría resultar útil a la hora de adquirir información contextual cuando se está traduciendo.

³ Para que la búsqueda sea lo más exacta posible se recomienda igualar en formato los textos (saltos de línea, párrafos, espacios, etc.) con el programa *Notepad++*, ya que si no *ParaConc* no alineará correctamente.

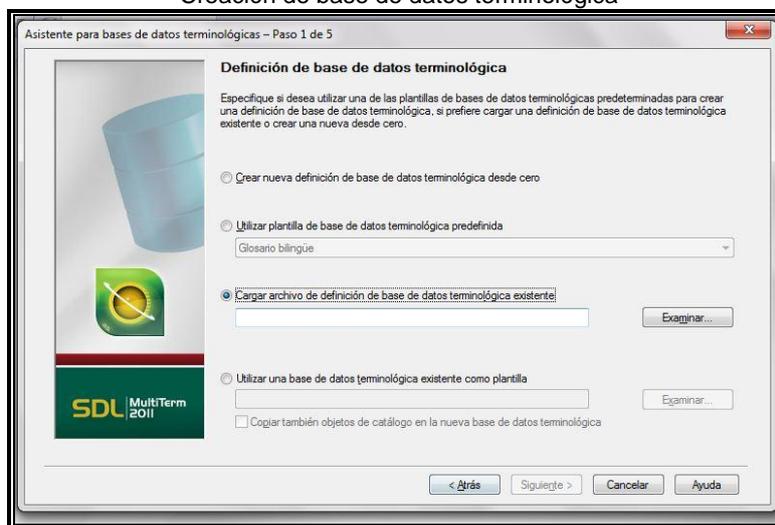
Por último, se rellenará el campo de la definición en español, para ello el alumno deberá recurrir a internet en busca de fuentes fiables que le proporcionen una definición correcta del término en cuestión.

2.6. Migración de los datos a Multiterm

Para la migración de los datos a *Multiterm* se recurrirá a dos herramientas básicas de *SDL*, por un lado a *Multiterm Convert*, que se encargará de modificar la hoja de cálculo en archivos de definición compatibles con *Multiterm*. Para ello será suficiente que el alumno siga el asistente que el programa proporciona, teniendo en cuenta cuáles son los campos índice y cuáles son los campos descriptivos y de cuál depende cada uno. En nuestro caso los campos índice son los idiomas y los descriptivos serán el contexto, la definición y la abreviatura, cada uno dependerá de su respectiva lengua.

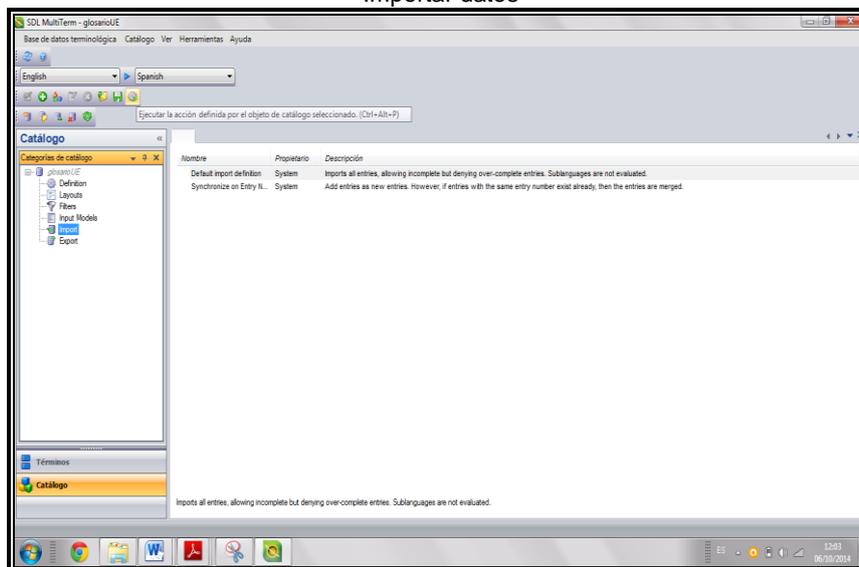
Finalizada la conversión, el alumno creará una base de datos en *Multiterm Desktop* basada en archivos, utilizando los ya creados por *Multiterm Convert*. En este caso, en vez de crear una base de datos desde cero, tendrá que cargar el archivo de definición existente (creado con *Multiterm Convert*) en el primer paso del asistente.

Creación de base de datos terminológica



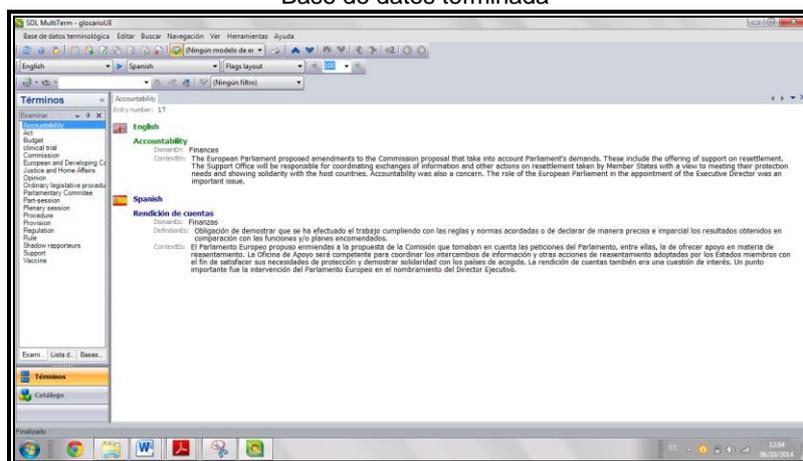
Una vez creada, se procederá a importar los datos cargados, mediante la función *catalogo>importar*.

Importar datos



Terminada la importación (*Cuadro 9*) el alumno podrá consultar la base de datos terminológica de forma aislada o integrada en *SDL Trados (2009/2011/2014)*. Asimismo, no solo podrá consultarla sino también agregar más términos ya sea desde *Multiterm* o desde *Trados* con la función *Agregar término*.

Base de datos terminada



CONCLUSIONES

La traducción especializada es una profesión que hoy en día está en auge y tal y como afirma Corpas (2011) sus perspectivas más inmediatas son la diversificación y ampliación de la oferta actual. Por ello, se le debe de ofrecer al alumnado del Grado de Traducción e Interpretación una formación integrada y transversal en la que no aprendan únicamente a traducir, sino también a ser eficientes, competitivos y a ofrecer un mayor número de servicios dentro del ámbito de la traducción.

El empleo de las herramientas informáticas ofrecen al traductor la posibilidad de ser más eficiente en su trabajo y, por lo tanto, más competitivo. Por ello, su trabajo no se entendería actualmente si un conocimiento profundo de este tipo de herramientas.

Un ejemplo de ello es la metodología aquí propuesta que ofrece, por un lado, la capacidad de aprender a documentarse en una fuente fiable y multilingüe como es el Registro Público de Documentos del Parlamento Europeo, la capacidad de extraer información de manera rápida y sencilla que se adapte tanto a necesidades concretas y puntuales (como la traducción económica, por ejemplo), como a unas necesidades más generales (como puede ser la traducción institucional y/o comunitaria), la capacidad de utilizar diversas herramientas informáticas y a generar/consultar glosarios o bases de datos terminológicas especializadas.

REFERENCIAS BIBLIOGRAFICAS

- BAKER, M., "Corpus-based Translation Studies. The Challenges that Lie Ahead". En: *Terminology, LSP and Translation*. Amsterdam: John Benjamins, 1996, pp. 175-86.
- BAKER, M. *et alii*, "Corpus Linguistics and South Asian languages: corpus creation and tool development". En: *Literary and Linguistic Computing* 19/4, 2004, pp. 509-524.
- BIBER, D.; CONRAD, S.; REPPEN, R., *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press, 1998.
- BOWKER, L.; PEARSON, J., *Working with Specialized Language*. Abingdon, UK: Taylor & Francis, 2002.
- BROWN, P.; LAI, J.; MERCER, R., "Aligning sentences in parallel corpora" En: *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Berkeley, CA., 1991, pp. 169-176.
- CORPAS PASTOR, G., *Un estudio paralelo de los sistemas fraseológicos del inglés y del español*. Tesis Doctoral. Madrid: Universidad Complutense de Madrid, 1995.
- , "Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada". *TRANS. Revista de Traductología*, 5, 2001, pp. 155-184.
- , "Traducir con corpus: de la teoría a la práctica". En *Texto, Terminología y Traducción*. Salamanca: Almar, 2002, pp.189- 226.
- , "Localización de recursos y compilación de corpus vía Internet: Aplicaciones para la didáctica de la traducción médica especializada". En: *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros, 2004, pp. 223-257.
- EAGLES, "Text corpora Working Group reading Guide". *EAGLES Document EAG-TCWG-FR-2*. 1996a
Disponible en línea: <http://www.ilc.cnr.it/EAGLES/corpintr/corpintr.html>
[Fecha de consulta: 17 de noviembre 2013].
- , "Preliminary Recommendations on Corpus Typology". *EAGLES Document EAG-TCWG-CTYP/P*. S. 1996b. Disponible en línea: <http://www.ilc.cnr.it/EAGLES96/corpusstyp/corpusstyp.html>
[Fecha de consulta: 17 de noviembre 2013].
- FIRTH, J., "A synopsis of Linguistic Theory, 1930- 1955". *Studies in Linguistic Analysis. Número especial de Philological Society*, 1957a, pp. 1-32.
- , *Papers in Linguistics*. London: Oxford University Press, 1957b.
- HUNSTON, S., *Corpora in applied linguistics*. Cambridge: Cambridge University Press, 2002.

- JOHANSSON, S., "On the Role of Corpora in Cross-Linguistic Research". En: *Corpora and Cross-Linguistic Research*. Amsterdam: Rodopi, 1998, pp. 3-24.
- KENNEDY, G., *An introduction to corpus linguistics*. Harlow: Longman, 1998.
- KENNY, D., *Lexis and Creativity in Translation. A Corpus-based Study*. Manchester: St. Jerome, 2001.
- KOHN, J., "What can (Corpus) Linguistics do for Translation?". En: *Translation Studies in Hungary. Budapest. Scolastica*, 1996, pp. 39-52.
- LAVIOSA, S. (ed.), "L'approche basée sur le corpus/The Corpus-based Approach". *Meta*, 43, 4. Montreal: Presses de l'Université de Montreal, 2001.
- , "How comparable can 'comparable corpora' be?". *Target*, 9, 2. 1997, pp. 289-319.
- LEECH, G., "Corpora and theories of linguistic performance". En: *Directions in Linguistics: Proceedings of Nobel Symposium 82* (Stockholm, 4-8 August 1991). Berlin: Mouton de Gruyter, 1991, pp. 105-122.
- MCENERY, A.. "Corpus linguistics". En: *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 2003, pp. 448-463.
- MCENERY, A.; WILSON, A., *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2001.
- MCENERY, A.; OAKES, M., "Sentence and word alignment in the Crater project: methods and assessment". En: *Proceedings of the Association for Computational Linguistics Workshop SIG-DAT Workshop*. Dublin, 1995, pp. 77-86.
- . "Sentence and word alignment in the Crater project: methods and assessment". En: *Using Corpora for Language Research*. London: Longman, 1996, pp. 211-231.
- MCENERY, A.; R. XIAO, "Domains, text types, aspect marking and English-Chinese translation". *Languages in Contrast*, 2, 2, 2002, pp. 211-229.
- MCENERY, T.; XIAO, R.; TONO, Y., *Corpus-Based Language Studies. An advanced resource book*. London/New York: Routledge, 2006.
- PÉREZ HERNÁNDEZ, M. C., "Explotación de los c rpora textuales informatizados para la creaci n de bases de datos terminol gicas basadas en el conocimiento". *Estudios de Ling stica Del Espa ol (ELiEs)*, 18, 2002.
Disponible en l nea: <http://elies.rediris.es/elies18/index.html>
[Fecha de consulta: 17 de noviembre 2013].
- SINCLAIR, J., "EAGLES preliminary recommendations on corpus typology". *EAGTCWG-CTYP/P*. Pisa: ILC-CNR, 1996.

- SINCLAIR, J., *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.
- STUBBS, M., "On inference theories and code theories: corpus evidence for semantic schemas". *Text* 21/3, 2001a, pp. 437-465.
- STUBBS, M., "Texts, corpora, and problems of interpretation: a response to Widdowson". *Applied Linguistics* 22/2, 2001b, pp. 140-172.
- STUBBS, M., *Words and Phrases*. Oxford: Blackwell, 2001c.
- SVARTVIK, J., "Introduction". En: *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, 1992.
- TORRUELLA, J.; LLISTERRI, J., "Diseño de corpus textuales y orales". En: *Filología e informática: Nuevas tecnologías en los estudios filológicos*. Barcelona: Seminario de Filología e Informática de la Universidad Autónoma de Barcelona y Ed. Milenio, 1999, pp. 45-77. Disponible en línea en: http://liceu.uab.es/%7Ejoaquim/publicacions/Torruella_Llisterri_99.pdf
[Fecha de consulta: 17 de noviembre 2013].
- TOURY, G., *Descriptive Translation Studies and beyond*. Amsterdam: John Benjamins, 1995.
- TYMOZCO, M., "Computerized corpora and the future of translation studies", *Meta*, 43, 1998, pp. 652-659.
- ZANETTIN, F., "Parallel corpora in translation studies: Issues in corpus design". En: *Intercultural Faultlines. Research methods in translation studies I: Textual and cognitive aspects*. Manchester: St Jerome, 2000, pp. 105 -18.