

# Metodología de diseño y compilación de un corpus representativo de seguros turísticos

Miriam Seghiri  
Universidad de Málaga  
*seghiri@uma.es*

## 1. Introduction

En el presente trabajo<sup>1</sup> se presenta una metodología protocolizada para la compilación de corpus virtuales, que se ilustrará mediante la creación de un corpus virtual bilingüe (inglés-español) de pólizas de seguros turísticos, que hemos denominado *Segurtur*, creado exclusivamente a partir de recursos electrónicos en red. Una vez demostrada la representatividad de *Segurtur*, mediante la aplicación informática *ReCor*<sup>2</sup>, el corpus compilado estaría en condiciones de ser utilizado para la realización de estudios lingüísticos y traductológicos sobre la contratación de los seguros turísticos. La elección se justifica por la enorme demanda de este tipo de traducciones, tanto directa como inversa, en nuestro país (cfr. ACT, 2005), ya que la industria turística española representa uno de los pilares fundamentales de nuestra economía, pues tal y como apuntan Alcaraz Varó et al. (2006) «El turismo de masas es uno de los fenómenos más novedosos desde la segunda mitad del siglo xx. En España es la industria número uno, la fuente principal generadora de riqueza y de puestos de trabajo». A ello se suma el hecho de que el consumidor europeo está en posición de exigir la traducción de este tipo de documentos, al amparo de las Directivas

---

<sup>1</sup> Este trabajo se enmarca en el seno del proyecto de investigación autonómico de excelencia *La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos* (HUM-892, 2006-2009), así como en el proyecto I+D nacional *Espacio único de sistemas de información ontológica y tesauros sobre el medio ambiente: Ecoturismo* (FFI2008-06080-C03-03) y en el grupo de investigación *Lexicografía y Traducción* (Hum 106-J.A).

<sup>2</sup> La aplicación informática *ReCor*, diseñada por las doctoras Gloria Corpas Pastor y Miriam Seghiri, de la Universidad de Málaga, supone una solución eficaz para determinar *a posteriori*, por primera vez, de forma objetiva y cuantificable, el tamaño mínimo que debe alcanzar un corpus para que sea considerado representativo en términos estadísticos. Asimismo, *ReCor* ha recibido el *Premio de Investigación en Tecnologías de la Traducción* (2007) concedido por el Observatorio de Tecnologías de la Traducción. Para más información, véase <<http://www.uem.es/web/ott/>>.

europas en materia asegurativa —*Directiva 92/49/CEE*<sup>3</sup> y *Directiva 92/96/CEE*<sup>4</sup>—, y sus correspondientes transposiciones nacionales<sup>5</sup>, las cuales reconocen al tomador del seguro el derecho a que la póliza se redacte, además de en una lengua oficial del territorio del Estado miembro donde se formalice, en otra lengua oficial en Europa que él desee.

## 2. El corpus virtual para la práctica traductora

Las ventajas de utilizar los corpus en la práctica traductora han sido señaladas por un gran número de investigadores (cfr. Laviosa, 1998; Bowker, 2002; Bowker y Pearson, 2002; Zanettin et al. 2003, entre otros). De hecho, algunos autores (cfr. Pearson, 1998; Bernardini y Zanettin, 2000; Corpas Pastor, 2001 y 2004; Zanettin, 2002*a* y *b*; Sánchez-Gijón, 2003*a* y *b*; Seghiri, 2006, 2008 y en prensa *a*, *b* y *c*) van más allá y señalan un tipo concreto de corpus, el *virtual*, como uno de los principales valedores del traductor a la hora de afrontar a una traducción especializada. Un *corpus virtual* puede definirse como aquel que se crea exclusivamente a partir de recursos electrónicos para la realización puntual de un determinado encargo de traducción o para la documentación de un bloque textual, y cuyo principal objetivo es satisfacer las principales necesidades del traductor de forma *económica* (a través de textos que se encuentran disponibles de forma gratuita en la red Internet), *rápida* (dado que los documentos pueden descargarse en cuestión de segundos con un “click” de ratón) y *fiable* (pues permite observar, entre otros, la macro- y microestructura textual así como la terminología propia del campo de especialidad *in vivo*). Estos corpus, así pues, tienen que ser creados por el propio traductor pero para ello, previamente, deberá considerar una serie de parámetros para su diseño, que pasamos a describir a continuación.

## 3. Creación de un corpus virtual

---

<sup>3</sup> <[http://eur-lex.europa.eu/smartapi/cgi/sga\\_doc?smartapi!celexplus!prod!DocNumber&lg=es&type\\_doc=Directive&an\\_doc=92&nu\\_doc=49](http://eur-lex.europa.eu/smartapi/cgi/sga_doc?smartapi!celexplus!prod!DocNumber&lg=es&type_doc=Directive&an_doc=92&nu_doc=49)>.

<sup>4</sup> <<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1992L0096:19950807:ES:PDF>>.

<sup>5</sup> Estas trasposiciones, con el principal objetivo de proteger a los consumidores y de favorecer la pluralidad lingüística de Europa, se han visto plasmadas, en el caso de España, en la *Ley 18/1997, de 13 de mayo, de modificaciones del artículo 8 de la Ley de Contrato de Seguro, para garantizar la plena utilización de todas las lenguas oficiales en la redacción de los contratos*, (BOE, de 14 mayo de 1997); en Reino Unido, en el *Statutory Instrument 2004, n.º 353. Insurers (Reorganisation and Winding Up) Regulations 2004*; y, finalmente, en la República de Irlanda, en la *Insurance Act, 2000*.

El protocolo de compilación de corpus virtuales consistente en cuatro fases bien diferenciadas, a saber, (1) búsqueda y acceso a la información, (2) descarga, (3) normalización y (4) almacenamiento. Para ello, sin embargo, será preciso determinar previamente los parámetros de diseño que regirán la creación del corpus virtual. Esta metodología, válida para la compilación de cualquier corpus de especialidad y en cualquier lengua, se ilustrará a través de la compilación, como indicamos con anterioridad, de un corpus de pólizas aseguramiento turístico (español-inglés).

### 3.1. Criterios de diseño

Antes de iniciar el proceso de compilación de un corpus, es imprescindible que el traductor establezca unos criterios claros de diseño. En nuestro caso, pretendemos crear un corpus de pólizas de aseguramiento turístico, compilado exclusivamente a partir de textos disponibles en la red y, por consiguiente, será un corpus *virtual*. Este corpus *bilingüe* —integrado dos subcorpus, uno de pólizas de aseguramiento turístico en inglés y otro en español— estará delimitado diatópicamente a aquellos documentos redactados en España y Reino Unido, habida cuenta del vasto número de países en los que el español y el inglés son lenguas oficiales. También será *comparable*, dado que se compondrá de pólizas originales, no traducidas, para la contratación turística, a la par que *textual*, pues lo integrarán documentos completos. A su vez, estaremos frente a un corpus muy *homogéneo*, que recogerá exclusivamente pólizas para el aseguramiento turístico, aunque, en principio, podemos presuponer que será *desequilibrado* en cuanto a número de documentos, ya que no hay que olvidar que el inglés es el idioma dominante en la red Internet<sup>6</sup> y, por consiguiente, es probable que se encuentren más documentos en esta lengua que en español. No obstante, aunque *desequilibrado*, será un corpus *representativo* a nivel *cualitativo* —pues la calidad viene garantizada los parámetros de diseño— así como a nivel *cuantitativo*, a través de la aplicación del programa diseñado para tal fin, *ReCor*.

### 3.2. Metodología protocolizada de compilación

---

<sup>6</sup> A este respecto, remitimos al reciente estudio sobre la presencia de las lenguas en la red Internet realizado en diciembre de 2009 por Internet World Stats (<http://www.internetworldstats.com/stats7.htm>).

Una vez establecidos los parámetros de diseño preliminares, el traductor habrá de seguir un protocolo de creación del corpus, si quiere asegurar la calidad de la muestra. De este modo, el protocolo comprenderá cuatro fases, que pasamos a describir a continuación.

### 3.2.1. Búsqueda y acceso de la documentación

Una vez diseñado el tipo de corpus virtual que se pretende compilar, la primera fase del protocolo consiste en localizar y acceder a la información disponible en Internet. Son principalmente dos los tipos de búsquedas que puede llevar a cabo el traductor en la red, a saber, la *búsqueda institucional*, que es aquella que se realiza en sitios web específicos de instituciones, organizaciones, asociaciones y empresas internacionales o nacionales de reconocido prestigio, y la *búsqueda por palabra clave*, mediante buscadores.

De este modo, iniciamos nuestra incursión en la red a partir de una *búsqueda institucional*<sup>7</sup>, que ha sido una de las más fructíferas para la alimentación del corpus, no sólo por la gran cantidad de documentos que albergan hoy en día este tipo de instituciones, organizaciones, asociaciones o empresas, entre otras, en la red, sino por la alta calidad y grado de fiabilidad que se les presupone de antemano, ya que los redactores son especialistas en la materia. Esta búsqueda de pólizas se realizará principalmente, aunque no de forma exclusiva, en las sedes de las grandes aseguradoras como las internacionales *Mondial Assistance*<sup>8</sup> o *Europ Assistance*<sup>9</sup>; las británicas *AT Bell Insurance Brokers Ltd*<sup>10</sup>, *Royal and Sun Alliance*<sup>11</sup> o *Lloyds of London*<sup>12</sup>; o las españolas *Allianz*<sup>13</sup>, *MAPFRE*<sup>14</sup> u *Ocaso*<sup>15</sup>, por citar sólo algunas de las más representativas.

---

<sup>7</sup> En gran número de ocasiones, cuando se agotan el nombre de entidades conocidas en esta materia, la búsqueda institucional requiere de otra por palabra clave, por lo general en motores de búsqueda. Normalmente, la introducción de descriptores en *Google* <<http://google.com>> utilizando las técnicas booleanas propias de este buscador como, por ejemplo, *organismo OR turismo*, *organismo AND turismo* u “*organismo turístico*”, en el caso de que quisiéramos incrementar el número de nombres que integran este tipo de instituciones turísticas, con vistas a visitar sus sedes web para extraer información susceptible de ser incluida en el corpus de pólizas de seguros turísticos.

<sup>8</sup> <<http://www.mondial-assistance.com/en/aboutus/homepage.htm>>.

<sup>9</sup> <<http://www.europassistance.com>>.

<sup>10</sup> <<http://www.atbell.co.uk>>.

<sup>11</sup> <<http://www.royalsunalliance.com/royalsun>>.

<sup>12</sup> <<http://www.lloyds.com>>.

<sup>13</sup> <<http://www.allianz.es>>.

<sup>14</sup> <<http://www.mapfre.com/pmapfre/es/index.html>>.

<sup>15</sup> <<http://www.ocado.es>>.

Otra opción complementaria a la anterior es llevar a cabo una *búsqueda por palabra clave*. Para ello, es posible recurrir a un buscador genérico, como por ejemplo *Google* que, según un gran número de analistas, es el mejor por calidad en el resultado de las búsquedas (cfr. Radev et al., 2005: 580), puesto que permite aplicar técnicas de búsqueda booleanas y delimitar diatópicamente la búsqueda a partir la opción de *búsqueda avanzada*<sup>16</sup>. Aquí, el establecimiento de *descriptores* es esencial. A continuación ilustramos (cfr. Tabla 1)<sup>17</sup>, a modo de ejemplo, cómo se realizaron las búsquedas para la localización de las pólizas que conforman el corpus:

Lengua	Delimitación diatópica	Tipo textual	Descriptores	Ecuación de búsqueda
<i>Español</i>	<i>España</i>	<i>Póliza</i>	Póliza, seguro turístico, asistencia en viaje <sup>18</sup>	“seguro turístico” OR “asistencia en viaje” AND póliza
<i>Inglés</i>	<i>Reino Unido</i>	<i>Póliza (Policy)</i>	Policy, travel insurance	policy AND “travel insurance”

*Tabla 1: Descriptores para la localización de pólizas de aseguramiento turístico (español-inglés).*

<sup>16</sup> Para la extracción de pólizas en español mediante la versión de *Google* para España (<<http://www.google.es>>), se selecciona la opción «páginas de España», con el fin de filtrar todos aquellos documentos provenientes de otros países hispanohablantes. El mismo procedimiento se puede seguir para la búsqueda de información en inglés, es decir, el usuario se dirige a las versiones de este buscador para Reino Unido (<<http://www.google.co.uk>>), con el fin de evitar la presencia de documentos provenientes de otros países.

<sup>17</sup> En esta tabla hemos presentado los descriptores que nos han permitido extraer mayor número de documentos para el tipo textual que se rastreaba en dos lenguas determinadas, en este caso español e inglés, aunque, en este sentido, conviene señalar que existen infinitos criterios de búsqueda y que aquí sólo pretendemos recoger una muestra a modo ilustrativo.

<sup>18</sup> El término *seguro turístico* cuenta con una larga tradición en España desde la aprobación, en 1964, por parte de la Presidencia de Gobierno, del *Decreto 3404/1964*, de 22 de octubre, por el que se establece el «seguro turístico». Sin embargo, la *Orden Ministerial de 27 de enero de 1988 por la que se califica la cobertura de las prestaciones de asistencia en viaje como operación de seguro privado*<sup>18</sup> transpuso a nuestro ordenamiento el texto de la *Directiva 84/641/CEE* del Consejo, de 10 de diciembre de 1984, por la que se modifica, en lo que se refiere en particular a la asistencia turística, la *Directiva 73/239*. La orden ministerial transpuesta empleaba el término «travel assistance», que fue traducido al español por el calco neológico oficialmente aceptado de «asistencia en viaje». No obstante, nos referiremos normalmente a éste como *seguro turístico*, pues coincidimos con la postura adoptada por Auriolés (cfr. Auriolés Martín (2005 [2002]) y Auriolés Martín et al. (2004), pues entendemos que es más acertado que el calco español del original inglés *asistencia en viaje*, dado que la asistencia en viaje es tan sólo una de las posibles coberturas que ofrece el seguro turístico, como pueden ser otras del tipo anulación del viaje o asistencia médica, por citar algunas de las más frecuentes en esta contratación. Para una visión más amplia al respecto, véase Seghiri (2006 y 2008 [2007]).

El principal inconveniente de la búsqueda por palabra clave es la elección de los descriptores más precisos, puesto que, de lo contrario, la información no pertinente tendrá una presencia elevada. Corresponde, así pues, al traductor filtrar todo ese ruido documental en cada una de las páginas susceptibles de ser incluidas en el corpus.

### 3.2.2. Descarga de datos

A la fase de búsqueda y acceso de los documentos le sigue una *descarga de datos*. Normalmente, esta fase se realiza de forma manual (Ctrl+S), aunque es posible automatizar la tarea cuando se trata de un grupo de páginas a través de programas que permiten la descarga en grupo de páginas o “en lotes”, como *GNU Wget*<sup>19</sup> o *GetBot*<sup>20</sup>.

### 2.2.3. Normalización

Se puede observar que en Internet, en el caso de las pólizas, existe una predilección por los formatos HTML (.html) y, sobre todo, PDF (.pdf). El primero no supone grandes problemas de conversión, pues puede copiarse la información y ser pegada en un documento de texto sin mayores dificultades. Por lo que se refiere a los documentos en .pdf, *Google* permite, en la mayoría de las ocasiones, visualizarlos como .html a través de la opción “Versión en HTML”, por lo que se puede proceder del mismo modo. Cuando esto no es posible, se puede recurrir a programas de conversión, como *Solid Converter*<sup>21</sup>, *PDFtoWord*<sup>22</sup> o *ABBYY FineReader*<sup>23</sup>. Se cierra, así, esta tercera fase, que sigue a la descarga, donde todos los documentos presentarán un formato ASCII o de texto plano, es decir, sin macar .html o código de otro tipo, siguiendo, en este punto, la *clean-text policy* expuesta por Sinclair (1991: 21).

### 2.2.4. Almacenamiento

---

<sup>19</sup> Este *software* libre, listo para descarga y acompañado de su manual de instrucciones se encuentra en la siguiente dirección URL: <<http://www.gnu.org/software/wget/>>.

<sup>20</sup> GetBot es un programa gratuito que se encuentra disponible para su descarga en la siguiente dirección URL: <<http://descargar.mp3.es/lv/group/view/kl42345/GetBot.htm>>.

<sup>21</sup> Puede descargarse la versión de prueba de *Solid Converter* gratuitamente en <<http://www.solidpdf.com>>. Esta versión, puesto que es de prueba, cuenta con una serie de limitaciones: funciona durante quince días y permite convertir un máximo de diez páginas por documento, aunque es posible convertir un texto en varias operaciones especificando un intervalo de páginas distinto cada vez.

<sup>22</sup> Puede descargarse la versión limitada de *PDFtoWord* en la dirección URL:<<http://pdf-to-word.softonic.com>>.

<sup>23</sup> Puede descargarse una versión de prueba en la dirección URL <<http://finereader.abbyy.com>>.

La última fase consiste en el *almacenamiento* de datos. De este modo, los documentos descargados se guardan, correctamente identificados y relacionados. Una forma posible de hacerlo es mediante subcarpetas en función de la lengua (inglés-español) y el formato, bien original (en el caso de que se deseen guardar) o bien en texto plano (.txt).

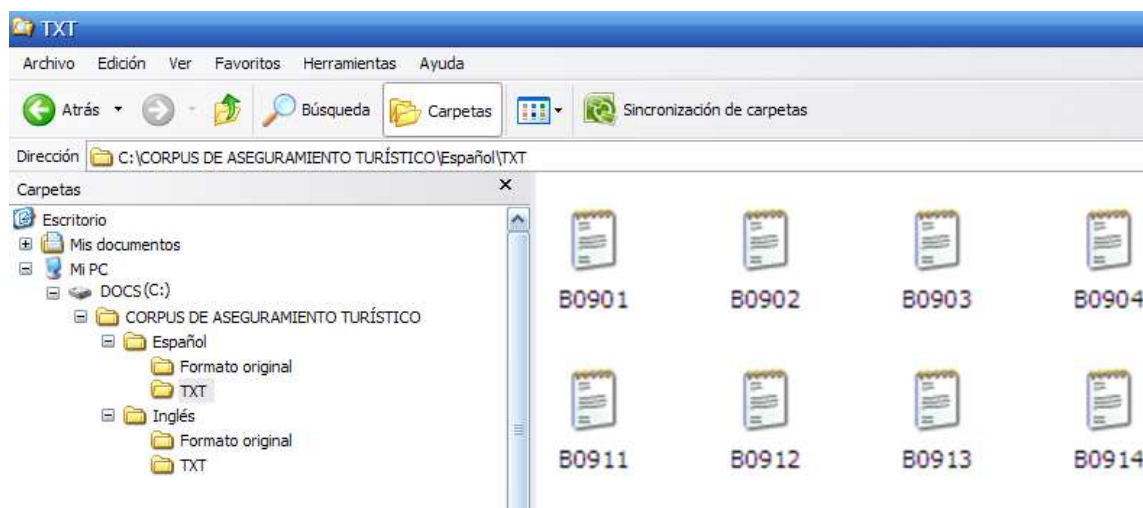


Ilustración 1: Almacenamiento del corpus de aseguramiento turístico

Una vez realizados los cuatro pasos —búsqueda y acceso de la documentación, descarga de datos, normalización y almacenamiento— el resultado ha sido la creación de un corpus virtual de pólizas de aseguramiento turístico, denominado *Segurtur*, integrado por un subcorpus de español, con 92 textos (901.869 palabras), y un subcorpus en inglés, de 158 documentos (1.902.198 palabras). Ahora que ya hemos compilado un corpus fiable y de calidad, sólo queda verificar que se ha alcanzado el tamaño mínimo de palabras y documentos para que la colección compilada sea considerada totalmente *representativa* de este campo de especialidad.

#### 4. Determinación de la representatividad

Con objeto de poder estimar la representatividad de los corpus en función de su tamaño vamos a utilizar *ReCor*, una aplicación informática se caracteriza, ante todo, por la sencillez de su interfaz de usuario (cfr. Ilustración 2), frente a la carga eminentemente matemática y de formulación que abundan en este tipo de trabajos.

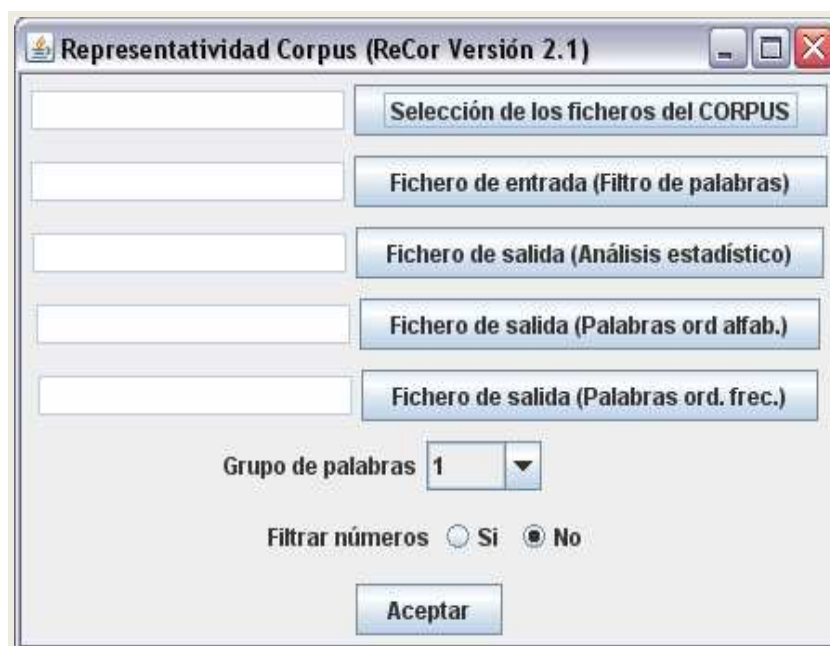


Ilustración 2: Interfaz de ReCor (versión 2.1)

En este apartado mostraremos el programa *ReCor* (versión 2.1)<sup>24</sup> en funcionamiento que, como veremos, resulta de muy fácil manejo. En primer lugar, procedemos a la selección de los archivos que conforman el subcorpus de pólizas de aseguramiento turístico en español mediante el botón «Selección de los ficheros del corpus». Una vez se han elegido los archivos que integran el corpus en español, podremos incorporar, si se desea, un «filtro de palabras». En nuestro caso, hemos incluido un filtro que contiene numeración romana. Además, el programa genera tres ficheros de salida (*Análisis estadístico*, *Palabras ord. alf.* y *Palabras ord. frec.*) que se crearán por defecto en la ubicación que determine la aplicación. Si se desea otra localización de los archivos de salida generados, puede indicarse una nueva ruta. El primero, «Análisis estadístico», recoge los resultados de dos análisis distintos; de un lado, los ficheros ordenados alfabéticamente por nombre; de otro, para los ficheros ordenados en orden aleatorio. El documento aparecerá estructurado en

---

<sup>24</sup> Para una descripción detallada del programa y de sus diferentes versiones, véanse Seghiri (2006) y Corpas y Seghiri (2007 a, 2007 b y en prensa). Asimismo, este programa ha sido validado mediante la Ley de Heaps (cfr. Seghiri, 2006) y se ha utilizado la determinar la representatividad de sendos corpus como los utilizados en los trabajos de Seghiri (2006), Toledo Báez (2009) o Castillo Rodríguez (2010).



cinco columnas, a saber, muestra de *types*, *tokens*, cociente entre palabras distintas y totales (*types/tokes*), número de palabras con una aparición (V1) y número de palabras con dos apariciones (V2). El segundo, «Palabras ord. alfa.», generará dos columnas en la que aparecerán las palabras ordenadas por orden alfabético, de una parte, y sus correspondientes ocurrencias, de otra. En tercer lugar, «Palabras ord. frec.», presenta la misma información que el fichero de salida anterior, pero esta vez las palabras se ordenan en función de su frecuencia, es decir, por rango.

Por último, procederemos a especificar «Grupo de palabras», esto es, los n-gramas. En este punto escogemos «1» (cfr. Ilustración 3). Asimismo, indicaremos «sí» en la opción «Filtrar números».

Una vez se han seguido los pasos descritos más arriba, la aplicación está lista para realizar el análisis, cuyo resultado se expresa en forma de representaciones gráficas y ficheros de salida en .txt con datos estadísticos exportables a tablas y tablas en Excel. Para generar las representaciones gráficas A y B, pulsamos «Aceptar». *ReCor* creará, además de los ficheros de salida, las representaciones gráficas A y B, que serán las que nos permitan determinar si, efectivamente, nuestra colección es representativa (cfr. Ilustración 3).

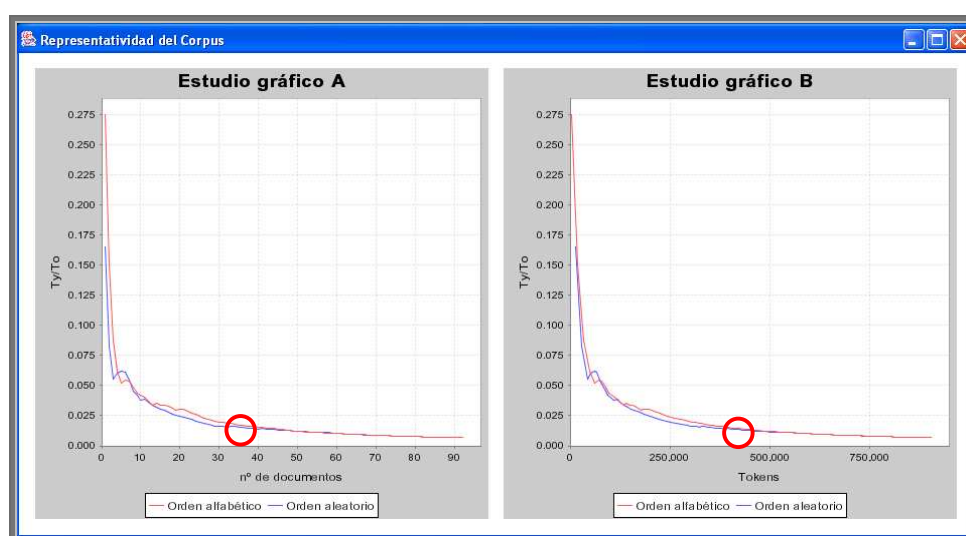


Ilustración 3: Representatividad del subcorpus en español de pólizas de aseguramiento turísticos (1-grama)

A partir de los datos arrojados por *ReCor*, podemos concluir que el subcorpus español de contratación de seguros turísticos (cfr. Ilustración 3) es representativo con 35 documentos y 400.000 palabras.

Seguimos los mismos pasos para gestionar el subcorpus de pólizas en lengua inglesa, y *ReCor* revela los siguientes resultados (cfr. Ilustración 4):

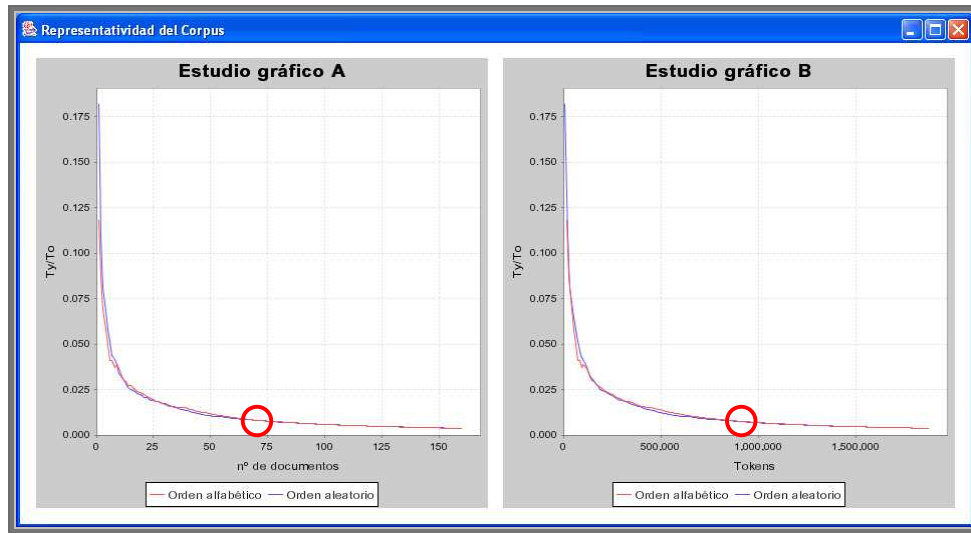


Ilustración 4: Representatividad del subcorpus en inglés de pólizas de aseguramiento turísticos (1-grama)

*ReCor* muestra que el subcorpus en lengua inglesa es representativo a partir de los 70 documentos y 900.000 palabras. Por consiguiente, si comparamos los resultados generados por el programa para cada uno de los subcorpus, puede inferirse que las pólizas en español tienden a ser más homogéneas que aquellas en inglés, es decir, se puede predecir que las pólizas en español presentarán una super-, macro- y microestructura muy similar entre ellas además de utilizar un abanico terminológico más reducido que en lengua inglesa.

Si deseamos ver los resultados para dos o más gramas, repetiremos los pasos anteriormente expuestos y especificaremos la cifra en «Grupo de palabras». Esta opción, posibilita el análisis por grupos de palabras (hasta 10), esto es, permitirá comprobar si el

texto es, a su vez, *representativo* con unidades superiores, lo que permitiría realizar estudios a nivel colocacional y fraseológico<sup>25</sup>.

## 5. Conclusiones

A día de hoy, no cabe duda de que el corpus virtual supone un recurso documental por excelencia en la práctica de la traducción especializada. Sin embargo, la metodología para su compilación a veces no resulta demasiado clara y la disponibilidad de los documentos en red viene siendo, en la mayoría de las ocasiones, el criterio que determina el tamaño final de la colección textual. Así, para que dicha colección pueda ser denominada *corpus* tendrá que cumplir con unos parámetros de diseño claros, que hemos descrito en primer lugar, seguido de un protocolo de compilación específico dividido en cuatro fases bien diferenciadas —(1) búsqueda y acceso a la información, (2) descarga, (3) normalización y (4) almacenamiento— de forma que quede asegurada la representatividad cualitativa de la muestra.

Para comprobar si el corpus compilado es representativo a nivel cuantitativo, se ha utilizado la aplicación informática *ReCor*, que permite, por primera vez, determinar el tamaño mínimo de un corpus.

El resultado obtenido ha sido un corpus virtual y bilingüe (español-inglés) de pólizas de seguro turístico, de ahí su nombre *Segurtur*, que, una vez demostrada su representatividad, se encuentra disponible para ser utilizado en cualquier tipo de estudio de corte monolingüe y monocultural, así como de traducción, de comparación y de mediación interlingüística e intercultural. Asimismo, el corpus virtual compilado puede resultar de gran utilidad para la docencia de la traducción especializada (jurídico-turística), pues ayuda a los procesos autónomos de enseñanza-aprendizaje, así como para la revisión y validación de encargos de traducción, además de otros posibles usos aún por descubrir.

## 6. Bibliografía

ACT. 2005. *Primer estudio de mercado de los servicios de traducción profesional en España de la Asociación de Empresas de Traducción (ACT)*. Madrid: ACT.

---

<sup>25</sup> En este sentido, se ha verificado la representatividad colocacional y fraseológica en Corpas y Seghiri (2007 a, 2007 b y 2008).

- Alcaraz, Enrique, Brian Hughes, Miguel Ángel Campos Pardillo, Víctor Manuel Pina y María Amparo Alesón Carbonell. 2006. *Diccionario de términos de turismo y ocio. Inglés-Español /Spanish-English*. Editorial Ariel: Barcelona.
- Bernardini, Silvia y Federico Zanettin (eds.). 2000. *I corpora nella didattica della traduzione. Corpus Use and Learning to Translate*. Bologna: CLUEB.
- Bowker, Lynne. 2002. *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- Bowker, Lynne y Jennifer Pearson. 2002. *Working with Specialized Language: A practical guide to using corpora*. Londres: Routledge.
- Castillo Rodríguez, Cristina M. 2010. *El trabajo terminográfico en traducción: explotación de un corpus multilingüe de turismo de salud y belleza (español-francés/inglés/italiano)*. Málaga: Universidad de Málaga.
- Corpas Pastor, Gloria. 2001. «Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada». *TRANS. Revista de Traductología*. N.º 5. 155-184.
- Corpas Pastor, Gloria. 2004. «Localización de recursos y compilación de corpus vía Internet: Aplicaciones para la didáctica de la traducción médica especializada». En Consuelo Gonzalo García y Valentín García Yebra (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros. 223-257.
- Corpas Pastor, Gloria y Míriam Seghiri Domínguez. 2007a. «Specialized Corpora for Translators: A Quantitative Method to Determine Representativeness». *Translation Journal*, 11 (3). <<http://accurapid.com/journal/41corpus.htm>>.
- Corpas Pastor, Gloria y Míriam Seghiri Domínguez. 2007b. «Determinación del umbral de representatividad de un corpus mediante el algoritmo N-Cor». *SEPLN: Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural*, 39. Sevilla: Universidad Hispalense. 165-172. <<http://www.sepln.org/revistaSEPLN/revista/39/20.pdf>>.
- Corpas Pastor, Gloria y Míriam Seghiri Domínguez. En prensa. *El concepto de*

*representatividad en lingüística de corpus: Aproximaciones teóricas y consecuencias para la traducción*. Málaga: Servicio de Publicaciones de la Universidad de Málaga.

Laviosa, Sara (ed.). 1998. *L'approche basée sur le corpus / The Corpus-based Approach, Meta*. Número monográfico 43 (4). Montréal: Presses de l'Université de Montréal.

Pearson, Jennifer. 1998. *Terms in Context, Studies in Corpus Linguistics*, 1. Amsterdam y Filadelfia: John Benjamins Publishing.

Radev, Dragomir, Weiguo Fan, Hong Qi, Harris Wu y Amardeep Grewal. 2005. «Probabilistic question answering on the web». *Journal of the American Society for Information Science and Technology (JASIST)*. 56 (6). 571-583.

Sánchez-Gijón, Pilar. 2003a. «És la web pública la nova biblioteca del traductor?». *Tradumàtica: Traducció i tecnologies de la informació i la comunicació*. Noviembre de 2003. N.º 2. 1-7 <<http://www.bib.uab.es/pub/tradumatica/15787559n2a7.pdf>> [Consulta: 30/04/2006].

Sánchez-Gijón, Pilar. 2003b. *Els documents digitals especialitzats: utilització de la lingüística de corpus com a front de recursos per a la traducció*. Tesis doctoral. Barcelona: Universidad Autónoma de Barcelona.

Seghiri, Miriam. 2006. *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Málaga: Servicio de Publicaciones de la Universidad de Málaga

Seghiri, Miriam. 2008 [2007]. «La traducción al español del término *travel insurance*». *PUNTOYCOMA, Boletín de los traductores españoles de las instituciones de la Unión Europea*. Bruselas: Comisión Europea. <[http://ec.europa.eu/translation/bulletins/puntoycoma/107/pyc1071\\_es.htm](http://ec.europa.eu/translation/bulletins/puntoycoma/107/pyc1071_es.htm)>.

Seghiri, Miriam. 2008. «Creating virtual corpora step by step». *Researching and Teaching specialized languages: New contexts, new challenges. VII Annual Conference of the European Association of Languages for Specific Purposes (AELFE)*. Murcia: Universidad de Murcia. 435-449.

- Seghiri, Miriam. En prensa *a*. «Virtual Corpus: A systematic Methodology for Compilation». *American Association for Applied Linguistics Annual Conference (AAAL 2008)*. Washington, D.C. (EE.UU.): Georgetown University.
- Seghiri, Miriam. En prensa *b*. «Using corpora in translation training: a step-by-step approach». *The American Translation and Interpreting Studies Association Conference (ATISA 2008)*. El Paso (EE.UU.): University of Texas at El Paso (EE.UU.)
- Seghiri, Miriam. En prensa *c*. «El corpus comparable para la didáctica de la traducción jurídica inversa (español-inglés)». *IV Congreso de la Asociación Ibérica de Estudios de Traducción e Interpretación: "Traducir en la frontera"*. Vigo: Universidad de Vigo.
- Sinclair, John M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Toledo Báez, M.<sup>a</sup> Cristina. 2009. *El resumen automático como recurso documental para la traducción de artículos de investigación del ámbito jurídico-tecnológico (español-inglés-francés)*. Tesis doctoral. Málaga: Universidad de Málaga.
- Zanettin, Federico. 2002*a*. «DIY Corpora: The WWW and the Translator». En Belinda Maia; Jonathan Haller y Margherita Urlrych (eds.). *Training the Language Services Provider for the New Millennium*. Oporto: Faculdade de Letras, Universidade do Porto. S. pag. <<http://www.federicozanettin.net/DIYcorpora.htm>> [Consulta: 30/04/2006].
- Zanettin, Federico. 2002*b*. «CEXI. Designing an English Italian Translational Corpus». En Bernhard Ketteman y Georg Marko (eds.). *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi. 329-343.