

Fuentes de información electrónicas para la compilación de un corpus virtual bilingüe de seguros turísticos

**Gloria CORPAS PASTOR
Míriam SEGHIRI DOMÍNGUEZ
Universidad de Málaga**

Como citar este artículo:

CORPAS PASTOR, Gloria y SEGHIRI DOMÍNGUEZ, Míriam (2008) «Fuentes de información electrónicas para la compilación de un corpus virtual bilingüe de seguros turísticos», en PEGENAUTE, L.; DECESARIS, J.; TRICÁS, M. y BERNAL, E. [eds.] *Actas del III Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. La traducción del futuro: mediación lingüística y cultural en el siglo XXI. Barcelona 22-24 de marzo de 2007*. Barcelona: PPU. Vol. n.º 2, pp. 297-308. ISBN 978-84-477-1027-0. Versión electrónica disponible en la web de la AIETI: http://www.aieti.eu/pubs/actas/III/AIETI_3_GCP_MSD_Fuentes.pdf.

Fuentes de información electrónicas para la compilación de un corpus virtual bilingüe de seguros turísticos*

Gloria Corpas Pastor
Míriam Seghiri Domínguez
Universidad de Málaga

1. Introducción

La importancia de la Documentación para el traductor queda patente ya desde la propia concepción de las materias troncales que deben figurar en los planes de estudios de la licenciatura en Traducción e Interpretación. La formación en Documentación¹ se considera, pues, imprescindible para un traductor, ya que una adecuada y concienzuda labor de documentación será la única que le permita traducir adecuadamente un texto especializado.² Las fuentes de las que puede beber información el traductor son muy variadas aunque, en este trabajo, nos vamos a centrar en aquellas a las que puede acceder a través de la red Internet puesto que, en la telaraña mundial, el traductor cuenta, en la actualidad, con gran cantidad de información sobre cualquier dominio de especialidad, a su alcance en cuestión de segundos, y en cualquier idioma, en una profesión donde normalmente se trabaja contrarreloj.

De este modo, la documentación electrónica se perfila como vital, hoy día, para la práctica, la docencia y la investigación en Traducción. Mucho se ha reflexionado en el seno de esta área de conocimiento sobre el tema³, aunque aún se está lejos de llegar a un consenso, una situación preocupante puesto que, para la compilación de un corpus, la cuestión de la documentación, en general, y de la localización y explotación de fuentes de información electrónica, en particular, reviste una importancia crucial. Esto es así porque, normalmente, estos criterios han sido demasiado generales y, en ocasiones, olvidan dos factores a tener en cuenta: por un lado, por la propia naturaleza de las lenguas objeto de estudio y, por otro, la propia tipología textual, en este caso, de nuestro corpus bilingüe (español-inglés) de seguros turísticos.

2. Características del corpus de seguros turísticos

La utilización de los corpus virtuales, de este modo, supone un gran avance para el manejo y extracción de información de las distintas fuentes documentales que en la actualidad pone a nuestro alcance la red Internet. Así, nos ocuparemos, específicamente, de la localización de recursos documentales para la creación de un corpus bilingüe

* El presente trabajo ha sido realizado en el seno del proyecto La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos (Ref. nº HUM-892, 2006-2009. Proyecto de Excelencia, Junta de Andalucía).

¹ Remitimos, en este sentido, al volumen recopilatorio de Félix Fernández y Ortega Arjonilla (1998) dedicado a las herramientas de apoyo a la traducción: terminología, documentación y recursos informáticos, así como a la traducción jurídica.

² Hurtado Albir (2001: 59-62) defiende la expresión «traducción de textos especializados» frente a «traducción especializada», ya que para la autora cualquier tipo de traducción es de algún modo especializada y precisará también, por consiguiente, de habilidades y conocimientos específicos.

³ Diversos documentalistas, así, se han ocupado de este tema y como resultado de sus reflexiones tenemos hoy día una multiplicidad de criterios y modelos generales no contrastados ni uniformizados. Véanse, por ejemplo, las propuestas por Pinto Molina y Cordón García (1999), Austermühl (2001), Rabadán y Fernández Nistal (2002), Corpas Pastor (2003), Palomares Perraut y Gómez Camarero (2003), Gonzalo García (2004 [2003] y 2004), Gonzalo García y Fraile Vicente (2004) o Gonzalo García y García Yebra (2004).

(español-inglés) de aseguramiento turístico que pretendemos compilar de forma virtual a partir, únicamente, de recursos electrónicos jurídico-turísticos disponibles en red. Este corpus, por su diseño —es bilingüe⁴, comparable⁵, paralelo⁶, textual⁷, desequilibrado⁸ y especializado, pues los textos que integran el corpus de seguros turísticos son específicamente legislación (comunitaria, nacional y autonómica) y los elementos formales del contrato, principalmente, solicitudes de seguro, propuestas, cartas de garantía y pólizas⁹—, responde a los parámetros de creación de corpus, por lo que estaría en condiciones de ser utilizado de forma independiente para la realización de estudios lingüísticos y traductológicos sobre la legislación y los elementos formales de este tipo contractual.

A este respecto, téngase en cuenta que en España existe una enorme demanda de este tipo de traducciones, tanto directa como inversa (*cf.* ACT, 2005). No en vano, la industria turística española se configura como uno de los principales motores de nuestra economía¹⁰. A esta situación, que podría ser pasajera, se añade el derecho del consumidor europeo a exigir la traducción de este tipo de documentos, al amparo de las directivas europeas¹¹ en materia asegurativa y sus correspondientes transposiciones

⁴Aunque es un corpus bilingüe (inglés-español), se encuentra delimitado diatópicamente. De este modo, los textos que integran el corpus de seguros turísticos son legislación y elementos formales del contrato que hayan sido redactados en España, República de Irlanda y Reino Unido (Escocia, Gales, Inglaterra e Irlanda del Norte).

⁵ Se trata de un corpus comparable pues está integrado por dos subcorpus, uno en español y otro en inglés, de textos originales para la contratación turística, concretamente, elementos formales del contrato y legislación.

⁶ Este corpus es también paralelo ya que cuenta con un subcorpus de textos traducidos, de menor tamaño que el comparable, debido a que los textos traducidos cuentan con una menor presencia en red.

⁷ El corpus de seguros turísticos compilado incluye documentos completos ya que este tipo de corpus es el que permite llevar a cabo investigaciones lingüísticas léxicas y de análisis del discurso, a la par que posibilita la creación de un subcorpus, o un componente, a partir de la selección de fragmentos más pequeños (Sinclair 1991). De hecho, Sinclair (1991) y Alvar Ezquerro *et al.* (1994) han puesto de manifiesto la necesidad de incluir textos enteros porque, de este modo, se elimina la discusión en torno a la representatividad de las distintas partes de un texto así como a la validez de las técnicas de muestreo.

⁸ Estamos ante un corpus eminentemente desequilibrado dada la evidente hegemonía del inglés en el ciberespacio frente al resto de lenguas; a ello, se suma el hecho de que en el diseño de este tipo de corpus virtuales compilados ad hoc «no prima tanto la cantidad como la calidad» (Corpas Pastor 2004: 236). A su vez, aunque desequilibrado, será limitado en su contenido y tremendamente homogéneo, puesto que atiende a un objetivo de creación determinado y puntual: la documentación de un bloque textual, la contratación de seguros turísticos, y que, en este sentido, también podrá ser reutilizable.

⁹ También existen, por un lado, el duplicado de la póliza, tramitado a petición del tomador, del asegurado o del beneficiario, siempre por escrito, a la aseguradora, quien tiene la obligación de realizar un duplicado o copia de la póliza en caso de extravío, que será idéntico y tendrá la misma validez que el documento original; por otro, el boletín de adhesión —un documento acreditativo del aseguramiento—, que no se ha incluido puesto que atañe exclusivamente a los seguros de vida.

¹⁰ El turismo mueve un volumen de negocio astronómico en la economía internacional, donde Europa ocupa una posición privilegiada que la sitúa a la cabeza a escala mundial. Así, en 2006, Europa generó en este sector 6.466,2 billones de dólares, lo que supone el 10,3 % del producto interior bruto (PIB) mundial —con una perspectiva de que alcance el 11 % en el año 2011—, y el 8,7 % del total de empleos (*cf.* WTTC, 2006a). Asimismo, remitimos a los estudios de WTTC sobre Reino Unido (2006b), Irlanda (2006c), y España (2006d) para una visión más detallada en torno a las cifras que mueve el sector en estos países.

¹¹ Nos referimos a la Directiva 92/49/CEE del Consejo de 18 de junio de 1992 sobre coordinación de las disposiciones legales, reglamentarias y administrativas relativas al seguro directo distinto del seguro de vida y por la que se modifican las Directivas 73/239/CEE y 88/357/CEE (tercera Directiva de seguros distintos del seguro de vida) y a la Directiva 92/96/CEE del Consejo, de 10 de noviembre de 1992, sobre coordinación de las disposiciones legales, reglamentarias y administrativas relativas al seguro directo de

nacionales¹², pues éstas reconocen al tomador del seguro el derecho a que el «contrato»¹³ se redacte, además de en una lengua oficial del territorio del Estado miembro donde se formalice, en otra lengua que él elija. Otras Directivas posteriores, como la 2002/92/CE¹⁴, también han contribuido a la demanda de traducciones de la totalidad de los documentos formales del contrato.

3. Fuentes de información electrónicas para la compilación del corpus

De esta forma, una vez decidido que el corpus será necesariamente electrónico, se nos plantea el problema del acceso a los documentos; es decir, dónde localizar esta información. Para este tipo textual, el seguro turístico, las principales fuentes de información¹⁵ han sido la *institucional* y la *telemática*. Las fuentes institucionales, por un lado, son aquellas que ofrecen información específica de instituciones, asociaciones u organizaciones internacionales, por ejemplo, o generada por ellas. Por otro, las fuentes telemáticas, son aquellas que permiten realizar búsquedas a través de los denominados *sistemas de búsqueda temáticos* —integrados principalmente por directorios— y, *sistemas de búsqueda generales* —como los buscadores, los meta- y multibuscadores— en los que se precisará del establecimiento de descriptores para la localización de la información.

3.1. Fuentes institucionales

Estas fuentes¹⁶, como ya apuntamos más arriba, tienen como principal cometido contener información creada por instituciones, organismos, entidades o colectivos. De hecho, los sitios creados por estas instituciones u organismos son habitualmente de gran fiabilidad; esto es así porque los redactores son expertos en la materia, de modo que la información contenida es correcta y la terminología utilizada es la normalizada para el

vida, y por la que se modifican las directivas 79/267/CEE y 90/619/CEE (tercera directiva de seguros de vida).

¹² 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 .

¹³ La póliza (*policy*, en inglés) es el documento en que se plasma el contrato de seguro, así como en el que se establecen las obligaciones y derechos tanto de la aseguradora como del asegurado, donde se describen las personas u objetos que se aseguran y se establecen las garantías e indemnizaciones en caso de siniestro, y que supone, además, la formalización y culminación de todo el proceso de contratación, de ahí que en muchos casos la póliza del seguro sea denominada contrato (*contract*) (cf. Ley 50/1980; Insurance Act, 2000; The Financial Services and Markets Act 2000).

¹⁴ Nos referimos específicamente a la Directiva 2002/92/CE del Parlamento Europeo y del Consejo, de 9 de diciembre de 2002, sobre la mediación en los seguros. Esta directiva, en su artículo 13, denominado «Modalidades de transmisión de la información», contempla que información proporcionada a los clientes deberá realizarse: «a) en papel o en otro soporte duradero disponible y accesible para el cliente; b) de forma clara y precisa, comprensible para el cliente; c) en una lengua oficial del Estado miembro del compromiso o en cualquier otra lengua acordada por las partes».

¹⁵ En este sentido, seguimos la taxonomía propuesta por Seghiri Domínguez (2006) para la clasificación de las fuentes de información en línea.

¹⁶ 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://www.google.com>), por ejemplo, utilizando las técnicas booleanas propias de este buscador del tipo 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 subjetiva de ser incluida en el corpus de seguros turísticos.

ámbito correspondiente o, a falta de ella, al menos de carácter oficial y acuñada según los principios de formación habituales para el campo temático en cuestión.

De hecho, este tipo de búsqueda ha sido crucial para localizar la gran parte de documentos que integran nuestro corpus. Así, para la localización de la legislación que conformará nuestro corpus, nos hemos servido, principalmente, de los siguientes sitios y páginas web:

Por lo que se refiere organismos e instituciones de carácter oficial, hemos extraído información legislativa de las sedes de la *ABI (Association of British Insurers)*, la *ABTA (Association of British Travel Agents)* o la *FSA (Financial Services Authority)*¹⁷, para el Reino Unido; y de *Mesa del Turismo*¹⁸, de la que destacamos el enlace destinado a la «legislación general», que incluye leyes de ordenación y leyes de disciplina turística; «sectorial», en el caso de España. Ocupa un lugar destacado, asimismo, el de la *OMT (Organización Mundial del Turismo)*¹⁹, que alberga uno de los principales recursos documentales en materia legislativa, *Lextour*²⁰. Se trata de la base de datos de legislación turística de la OMT, que enlaza a sitios web, bases de datos y servidores externos de legislación turística, todos ellos emitidos por parlamentos, organismos gubernamentales, universidades, asociaciones profesionales. También hemos accedido a otras bases de datos para la obtención de legislación comunitaria, como de la archiconocida *WestLaw*²¹ y, sobre todo, de *Eur-Lex*²², el portal del derecho de la Unión Europea que constituye, al día de hoy, la mayor base de datos documental del derecho de la Unión.

Para aproximarnos a las normativas nacionales, nos hemos servido, asimismo, de las bases de datos *HMSOnline (Her Majesty's Stationery Office)*²³, para Reino Unido, o de aquella incluida en *BAILII (British and Irish Legal Information Institute)*²⁴, que da acceso a las legislaciones de Reino Unido e Irlanda, además de a la de otros países comunitarios y extracomunitarios. Dentro de las bases de datos turísticas destaca, además, *Leisure Tourism Database*²⁵, de la Universidad de Southampton. En el caso de España, las principales bases de datos para la obtención de legislación en torno al aseguramiento turístico han sido las ubicadas en portales especializados como *JurisWeb*²⁶ o *Aranzadi*²⁷, además de *Legistur*²⁸, base de datos del Instituto de Estudios Turísticos del Ministerio de Turismo, Industria y Comercio, que cuenta con disposiciones legales en materia turística de la Administración del Estado, Comunidades Autónomas y Unión Europea.

Los documentos que conforman la contratación en materia de aseguración turística se ubican, prácticamente en su totalidad, en las sedes de las grandes aseguradoras;

¹⁷ <http://www.fsa.gov.uk/consumer/>. Todas las URLs incluidas en este trabajo estaban operativas el día 15 de enero de 2007.

¹⁸ <http://www.mesadelturismo.com>.

¹⁹ <http://www.world-tourism.org>.

²⁰ <http://www.world-tourism.org/doc/S/lextour.htm>.

²¹ <http://web2.westlaw.com/signon/default.wl?bhcp=1>.

²² <http://europa.eu.int/eur-lex/es/index.html>.

²³ <http://www.opsi.gov.uk>.

²⁴ <http://www.bailii.org>.

²⁵ <http://www.solent.ac.uk/library/databases/database.asp?database=lrta>.

²⁶ <http://www.jurisweb.com>.

²⁷ <http://www.aranzadi.es/nueva/web/index.htm>.

²⁸ http://www.iet.tourspain.es/paginas/cdte/legistur.aspx?option=doc&option2=BD_ES&idioma=es-ES.

también, aunque con menos frecuencia, en la infinidad de agencias de viajes que ofrecen sus servicios en red y cuelgan los documentos contractuales que han contratado con estas aseguradoras, poniéndolos a disposición de sus clientes. De este modo, aseguradoras internacionales como AIG Europe²⁹, Mondial Assistance³⁰, Europ Assistance³¹ o Winterthur³², británicas e irlandesas como AT Bell Insurance Brokers Ltd³³, AXA (UK)³⁴, Norwich Union³⁵, Royal and Sun Alliance³⁶, Lloyds of London³⁷ o AIG Europe (UK) Limited³⁸, o españolas, como Allianz³⁹, ARAG⁴⁰, ASEGMAP⁴¹, Europea⁴², MAPFRE⁴³ u Ocaso⁴⁴, por citar sólo algunas de las más representativas, han supuesto nuestras principales fuentes documentales.

3.2. Fuentes telemáticas

Estas fuentes en la red son aquellas creadas con el objetivo de que localicen información y recursos que se encuentran también en ella y están integradas, como ya apuntamos, por los sistemas de búsqueda temáticos —principalmente por directorios— y sistemas de búsqueda generales —buscadores, meta- y multibuscadores— en los que se precisa el establecimiento de descriptores para la localización de la información.

3.2.1. Sistemas de búsqueda temática

Otro paso para la localización de documentos que alimenten el corpus es el de proceder a la *búsqueda temática*⁴⁵ a través de los conocidos *directorios* o *índices*. El problema para la localización de la información, en este caso, viene dada por las estructuraciones propias que presentan cada uno de los directorios visitados, hecho que, en ocasiones, ralentizó la extracción de documentación. De este modo, en la extracción de legislación comunitaria, nacional y autonómica, se perfilaron como excelentes recursos los índices geográficos de carácter vertical como *AccessToLaw*⁴⁶, directorio jurídico muy completo y con una estructuración clara. También supusieron un caudal documental importante los directorios especializados, sobre todo aquellos que presentan los recursos evaluados o comentados como, para la compilación del subcorpus en español, el enlace «Dret» de

²⁹ <http://www.aigeuropa.com>.

³⁰ <http://www.mondial-assistance.com/en/aboutus/homepage.htm>.

³¹ <http://www.europassistance.com>.

³² <http://www.winterthur.com>.

³³ <http://www.atbell.co.uk>.

³⁴ <http://www.axa.co.uk>.

³⁵ <http://www.norwichunion.com>.

³⁶ <http://www.royalsunalliance.com/royalsun>.

³⁷ <http://www.lloyds.com>.

³⁸ <http://www.aigeuropa.co.uk>.

³⁹ <http://www.allianz.es>.

⁴⁰ <http://www.arag.es>.

⁴¹ <http://www.asegmap.org/counter/counter.php>.

⁴² <http://www.europeadeseguros.com/EuropeaSeguros/index.htm>.

⁴³ <http://www.mapfre.com/pmmapfre/es/index.html>.

⁴⁴ <http://www.ocaso.es>.

⁴⁵ Al igual que la institucional, la búsqueda temática puede ir acompañada de una por palabra clave en el caso de que desconozcamos o agotemos el nombre de árboles temáticos especializados en la materia objeto de búsqueda. De este modo, para la localización, por ejemplo, de directorios jurídicos, nos dirigimos normalmente a Google y, mediante descriptores combinados con operaciones booleanas, dieron lugar a ecuaciones de búsquedas tan fructíferas como «directorio jurídico» o «directorio AND jurídico».

⁴⁶ <http://www.accesstolaw.com/site/default.asp>.

los índices de *Universitat de Barcelona*⁴⁷ y de la *Universitat Autònoma de Barcelona*⁴⁸. Este último directorio, además, cuenta con el enlace «Centre de Documentació Europea», presentado por temas, del que nos servimos de la información contenida en «Turisme» y «Dret comunitari». Para el subcorpus en inglés, recurrimos al directorio *The Argus Clearinghouse*⁴⁹ y *Search the Law*⁵⁰; de este último destacamos el bloque destinado a viajes, «Travel».

3.2.2. Sistemas de búsqueda general

La búsqueda temática por índices o directorios resulta, de esta forma, fructífera para la extracción de legislación pero no de documentos contractuales. Para ello, se necesitó dar un paso más, concretamente la realización de búsqueda por palabras clave en motores de búsqueda —como el archiconocido *Google*—. Además de los motores de búsqueda o buscadores, la red Internet pone a nuestra disposición otros sistemas como los meta- y multibuscadores. A pesar de que una parte de los internautas puede considerar sinónimos ambos términos, no lo son: el primero, el metabuscador, supone una evolución del segundo. Los multibuscadores —como *Botfeeder*⁵¹, desarrollado por la compañía SearchApps o *Multibuscador.com*⁵², que realiza búsquedas de manera simultánea en más de diez buscadores españoles e internacionales— podrían definirse como buscadores de buscadores, pues esa es su única tarea. Los *metabuscadores*⁵³ —como *Killerinfo*⁵⁴, que interroga a más de 250 buscadores; *Vivísimo*⁵⁵, capaz de organizar la información en *clusters*; o *Ithaki*⁵⁶, versión internacional del español *Todalanet*⁵⁷, capaz de aunar algunos de los buscadores más poderosos de Internet como *Voilà*, *Teoma*, *Google*, *Fast/Lycos*, *Altavista*, *Northlight*, *Hotbot*, *Webcrawler*, *WiseNut*, *Directhit* o *Euroseek*— permiten, además, evaluar la relevancia de cada página encontrada al mostrar en los resultados la posición que dicha página ocupa en el buscador en el que se ha localizado.

El principal inconveniente de estos multi- y metabuscadores estriba, sin embargo, en que presentan un gran número de respuestas obtenidas, generando, por lo general, bastante ruido documental, y, por ello, son más efectivos para un primer acercamiento a la información que buscamos. Por ello, para la localización de elementos formales del contrato, nos dirigimos a *Google*, según un gran número de analistas el mejor por calidad en el resultado de las búsquedas⁵⁸ (Lorán 2005: 2).

⁴⁷ <http://www.bib.ub.es/bub/internet.htm>.

⁴⁸ <http://www.bib.uab.es/internet.htm>.

⁴⁹ <http://www.clearinghouse.net>.

⁵⁰ <http://www.search-the-law.com>.

⁵¹ <http://www.botfeeder.com>.

⁵² <http://www.multibuscador.com>.

⁵³ También llamados, aunque menos frecuente, megabuscador, superbuscador y gigabuscador.

⁵⁴ <http://www.killerinfo.com/>.

⁵⁵ <http://www.fazzle.com/log.jsp>.

⁵⁶ <http://www.ithaki.net/indexcau.htm>.

⁵⁷ <http://es.todalanet.net/>.

⁵⁸ Esto es así porque la gran parte de los motores de búsqueda entrega sus resultados con base en la frecuencia con que las palabras clave aparecen en un sitio web determinado; Google, no obstante, funciona de forma diferente: utiliza una tecnología conocida como PageRank, que realiza una valoración de la importancia de los sitios web mediante una ecuación que es capaz de detectar los vínculos con los que cuenta una página (los llamados ganglios y autoridades), esto es, cuantos más enlaces dirijan hacia una página determinada mayor será su importancia y ocupará, por consiguiente, un lugar preferente a la hora de que Google presente sus resultados. Concretamente, estudios llevados a cabo acerca de la eficacia

Como la legislación que afecta a la contratación de seguros turísticos ya se había visto completada con las búsquedas anteriores, la localización de documentos en sistemas de búsqueda generales mediante *palabras clave* demostró ser la más adecuada —junto con las visitas a sedes web de aseguradoras— y, sobre todo, la más rápida para rescatar los documentos que integran la contratación. En este sentido, los motores de búsqueda dan mejores resultados si sabemos utilizar los recursos que ofrecen —operadores booleanos, truncamientos, búsquedas por frase, entre otros— así como cuanto más definida sea nuestra búsqueda. En este sentido, el establecimiento de *descriptores* es esencial y para cuya selección hemos seguido nuestra propia intuición⁵⁹. De este modo, aunque para un primer acercamiento hayamos elegido los términos de búsqueda de forma intuitiva, una vez localizados los primeros documentos, puede accederse a las palabras clave del documento mediante el análisis del código fuente HTML. De dicho análisis, podemos afirmar que la elección de estas palabras es subjetiva y dependerá de aquellas que haya asignado el autor al documento para que éste sea rescatado por los distintos motores de búsqueda. A continuación ilustramos (*cf.* tablas 1 y 2⁶⁰), a modo de ejemplo, cómo se realizaron las búsquedas para la localización de los textos que conforman el corpus. Para ello, se tomaron como descriptores los tipos textuales y el ámbito de seguro en el que deseábamos localizar la información —seguro turístico—, y se aplican técnicas de búsqueda booleanas a partir del sencillo manejo que ofrece, por ejemplo, la búsqueda avanzada de *Google*⁶¹:

de los buscadores revelan que «Google is the best engine as the potential answers source for web-based question answering» (Radev *et al.* 2005: 580).

⁵⁹ Almahano Güeto (2002: 287-306) aboga por la intuición para la búsqueda de documentación mediante palabra clave para la compilación de un corpus ad hoc especializado.

⁶⁰ 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 y se han manejado infinitos criterios de búsqueda y que aquí sólo pretendemos recoger una muestra a modo ilustrativo.

⁶¹ Para la extracción de documentos contractuales en español nos dirigimos a la versión de Google para España (<http://www.google.es>), donde seleccionamos la opción «páginas de España», con el fin de filtrar todos aquellos documentos provenientes de otros países hispanohablantes. El mismo procedimiento fue el seguido para la búsqueda de información en inglés, es decir, nos dirigimos a las versiones de este buscador para Reino Unido (<http://www.google.co.uk>) e Irlanda (<http://www.google.ie>) y marcamos, respectivamente, «pages from the UK» y «pages from Ireland», con el fin de evitar la presencia de documentos provenientes de otros países. En algunas ocasiones, este filtrado no será suficiente por lo que, además de la realización de búsquedas por países, en el caso de que dudemos sobre el origen de la página localizada por Google, el dominio también nos puede ayudar a corroborar su procedencia. De este modo, serán objeto de nuestro interés los dominios .es, para España; .ie, para Irlanda y .uk, para Reino Unido, mientras que otros dominios indicativos de países como, por ejemplo, para páginas en español, .ar (Argentina) o .mx (México); o en inglés, como .au (Australia) o .us (Estados Unidos), serán automáticamente descartados, debido a que no son pertinentes para nuestro corpus.

Tipo textual	Descriptor	Ecuación de búsqueda
<i>Póliza</i>	Póliza, seguro turístico, asistencia en viaje ⁶² póliza AND «seguro turístico» póliza AND «asistencia en viaje»	
<i>Solicitud</i>	Solicitud de póliza	solicitud AND póliza AND «seguro turístico» solicitud AND póliza AND «asistencia en viaje»
<i>Propuesta</i>	Propuesta, proposición	póliza AND propuesta OR proposición «seguro turístico» póliza AND propuesta OR proposición «asistencia en viajes»
<i>Carta de Garantía</i>	Carta de garantía	«carta de garantía» AND «asistencia en viaje» «carta de garantía» AND «seguro turístico»

Tabla 1: Descriptores para la localización de elementos formales del contrato de seguro turístico (español).

Tipo textual	Descriptor	Ecuación de búsqueda
Policy	Policy, travel insurance	policy AND «travel insurance»
Quote	Quote	Quote AND policy AND «travel insurance»
Proposal Form	Proposal Form	«proposal form» AND policy AND «travel insurance»
Certificate of Insurance	Certificate of Insurance Insurance Certificate	«certificate of insurance» OR «insurance certificate» AND policy

Tabla 2: Descriptores para la localización de elementos formales del contrato de seguro turístico (inglés).

El principal inconveniente de las palabras clave es la elección de los descriptores más precisos para la búsqueda que se pretende realizar, puesto que, de lo contrario, la información no pertinente tendrá una presencia elevada. Nos corresponderá a nosotros filtrar todo ese ruido documental en cada una de las páginas susceptibles de ser incluidas en el corpus.

4. Corolario

El resultado final ha sido la creación rápida y sencilla de un corpus virtual bilingüe integrado, hoy día, por un subcorpus de español, con 259 textos⁶³ (1.837.869 palabras),

⁶² El término seguro turístico cuenta con una larga tradición en nuestro país, 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 privado¹ 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 (*cf.* 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 la clasificación trilingüe (español-inglés-italiano) de pólizas de seguro turístico en función de su cobertura propuesta por Seghiri Domínguez (2006: 279-281).

⁶³ Por lo que se refiere a los documentos legislativos que integran el corpus (17 textos en inglés y 2 textos en español), hemos de decir que los seguros turísticos no cuentan con legislación sustantiva, pero se ven afectados por las regulaciones de los seguros distintos del de vida mediante diversas directivas comunitarias como: 73/239/CEE, 73/240/CEE, 76/580/CEE, 78/473/CEE, 84/641/CEE, 87/343/CEE, 87/344/CEE, 88/357/CEE, 90/618/CEE, 92/49/CEE, 95/26/CEE, 2000/26/CE, 2000/64/CE y 2002/13/CE. Además, en España, los contratos de seguros turísticos se rigen en la actualidad por la Ley 50/1980, de 8 de octubre, de Contrato de Seguro, así como por la Ley 30/1995, de 8 de noviembre, de ordenación y supervisión de los Seguros Privados. En Irlanda, se regulan a través del Insurance Act, 2000, así como a partir del European Communities (Non-Life Insurance) Framework Regulations, 1994 (S.I. No. 359 of 1994). En Reino Unido, se les aplica el The Financial Services and Markets Act 2000 (Statutory Instrument 2003 N.º 1476), específicamente el Amendment, N.º. 2, Order 2003. Por otro lado, el

y un subcorpus en inglés, de 302 documentos (3.202.118 palabras). Todo ello, gracias a la selección de las fuentes de información en línea pertinentes para este tipo textual. En este punto, el corpus compilado abre la posibilidad un sinfín de posibilidades que van desde la creación de glosarios hasta el estudio del discurso jurídico propio del sector del aseguramiento turístico, la legislación nacional y comunitaria aplicable, sus manifestaciones contractuales correspondientes, y sus realizaciones y formas textuales en las dos lenguas implicadas, todo ello desde una perspectiva monolingüe y monocultural, así como desde el punto de vista de la traducción, la comparación y mediación interlingüística e intercultural, además del derecho comparado.

Referencias bibliográficas

- ACT (2005). *Primer estudio de mercadode los servicios de traducción profesional en España de la Asociación de Empresas de Traducción (ACT)*. Madrid: ACT.
- Almahano Güeto, I. (2002). *El contrato de viaje combinado en alemán y español: las condiciones generales. Un estudio basado en corpus*. Málaga: Universidad de Málaga. [Tesis doctoral.]
- Alvar Ezquerro, M., M. J. Blanco Rodríguez y F. Pérez Lagos (1994). «Diseño de un corpus español en el marco de un corpus europeo». En M. Alvar Ezquerro y J. A. Villena Ponsoda (coords.). *Estudios para un corpus del español*. Málaga: Universidad de Málaga. 9-30.
- Auriolés Martín, A. 2005 [2002]. *Introducción al Derecho Turístico (Derecho Privado del Turismo)*. 2.^a ed. Tecnos: Madrid.
- Auriolés Martín, A., P. Guillermina, B. Velasco y M. B. González Fernández (2004). *Contratación Turística*. Documento Técnico. [BFF2003-04616 MCYT/TI-DT-2004-1]. En línea en <http://turicor.com/privada/documentos/TI-DT-2004-1.pdf> [Consulta: 15/01/2007].
- Austermühl, F. (2001). *Electronic Tools for Translators*. Manchester: St. Jerome.
- Corpas Pastor, G. (ed.) (2003). *Recursos documentales y técnicos para la traducción del discurso jurídico (español, alemán, inglés, italiano, árabe)*. Granada: Comares.
- Corpas Pastor, G. (2004). «Localización de recursos y compilación de corpus vía Internet: Aplicaciones para la didáctica de la traducción médica especializada». En C. Gonzalo García y V. García Yebra (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros. 223-257.
- Félix Fernández, L. y E. Ortega Arjonilla (coords.) (1998). II Estudios sobre traducción e interpretación: Actas de las I Jornadas Internacionales de Traducción e Interpretación de la Universidad de Málaga, Málaga, 17-20 de abril de 1997. Tomo III. Málaga: Universidad de Málaga.
- Gonzalo García, C. y E. Fraile Vicente (2004). «Selección y evaluación de recursos lingüísticos en Internet para el traductor especializado». En C. Gonzalo García y V. García Yebra (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros, S.L: 337-372.
- Gonzalo García, C. y V. García Yebra (eds.) (2004). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros.
- Gonzalo García, C. 2004 [2003]. «DocuTradSo en línea: Fuentes de información para la actividad traductora». En línea en <http://www.uva.es/DocuTradSo> [Consulta: 15/01/2007].
- Gonzalo García, C. (2004). «Fuentes de información en línea para la traducción especializada», en C. Gonzalo García y V. García Yebra (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros. 275-336.

componente de pólizas en español cuenta con 101 documentos (1.000.067 palabras) mientras que el componente en inglés consta de 176 (1.903.661 palabras). El resto del corpus lo integran el resto de elementos formales del contrato.

- Hurtado Albir, A. (2001). *Traducción y traductología: introducción a la traductología*. Madrid: Cátedra.
- Lorán, R. 2005. «Los buscadores son cada vez más útiles e imprescindibles». En línea en <http://bibliotecavirtualut.suagm.edu/Publicaciones%20profesores/Roberto%20Lor%C3%A1n/Los%20buscadores%20son%20cada%20vez%20m%C3%A1s%20%C3%BAtiles%20e%20imprescindibles.pdf> [Consulta: 15/01/2007].
- Palomares Perraut, R. y C. Gómez Camarero (2003). «Evaluación de recursos electrónicos de información jurídica: criterios y modelos». En G. Corpas Pastor (ed.). *Recursos documentales y terminológicos para el estudio del discurso jurídico (español, alemán, árabe, inglés e italiano)*. Granada: Comares. 101-120.
- Pinto Molina, M. y J. A. Cordón García (eds.) (1999). *Técnicas documentales aplicadas a la traducción*. Madrid: Síntesis. 179-193.
- Rabadán, R. y P. Fernández Nistal (2002). *La traducción inglés-español: fundamentos, herramientas, aplicaciones*. León: Universidad de León.
- Radev, D; W. Fan; H. Qi; H. Wu y A. Grewal (2005). «Probabilistic question answering on the web». *Journal of the American Society for Information Science and Technology (JASIST)* 56 (6). 571-583.
- Seghiri Domínguez, M. (2006). *Compilación de un corpus trilingüe de seguros turísticos (español-inglés-italiano): aspectos de evaluación, catalogación, diseño y representatividad*. Málaga: Universidad de Málaga. [Tesis doctoral.]
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- WTTC (2006a). *World Travel and Tourism climbing to new heights. The 2006 Travel & Tourism Economic Research*. Londres: World Travel & Tourism Council.
- WTTC (2006b). *United Kingdom Travel and Tourism climbing to new heights. The 2006 Travel & Tourism Economic Research*. Londres: World Travel & Tourism Council.
- WTTC (2006c). *Ireland Travel and Tourism climbing to new heights. The 2006 Travel & Tourism Economic Research*. Londres: World Travel & Tourism Council.
- WTTC (2006d). *Spain Travel and Tourism climbing to new heights. The 2006 Travel & Tourism Economic Research*. Londres: World Travel & Tourism Council.