

EXTRACCIÓN DE GLOSARIOS BASADOS EN CORPUS Y SU APLICACIÓN A LA EVALUACIÓN DE LA CALIDAD EN LA INTERPRETACIÓN JUDICIAL

Luis Carlos Marín Navarro

lmarin@uma.es
Universidad de Málaga
Intérprete judicial y policial

Trabajo dedicado al gran profesor Emilio Ortega Arjonilla. Gracias por todo y por tanto.

Las denominadas “tecnologías de la traducción” son herramientas diseñadas específicamente para ayudar al traductor en su quehacer diario. Entre ellas se encuentra la compilación y gestión de corpus. La capacidad para manejarse con este tipo de tecnologías lingüísticas forma parte de las competencias profesionales del traductor.

(Corpas Pastor, 2012: 92)

1. Introducción

La cita anterior es el punto de partida del presente trabajo y sirve como uno de los ejes fundamentales para hacer eco de cómo las Tecnologías de la Información y la Comunicación (de ahora en adelante, TIC) y, más concretamente, las tecnologías de la traducción (e incluso la interpretación) se han ido abriendo paso en estos últimos años de forma radical. Numerosas investigaciones en este campo dan cuenta de la gran utilidad que estas poseen a la hora de traducir aunque, siguiendo a Muñoz Ramos (2012: 637), los alumnos no han recibido una formación en tecnologías de la traducción y, por consiguiente, el uso que estos hacen de

dichas tecnologías es escaso (*ibid*: 637). Si nos adentramos en el campo de la interpretación, las tecnologías contribuyen también de forma bastante notoria en el proceso interpretativo pero, en este caso, el desconocimiento es todavía mayor y, a día de hoy, muy pocos son los estudios que hablan del vínculo interpretación y corpus. Asimismo, hemos de recordar que las tecnologías siempre irán un paso por delante. Por tanto, para garantizar la calidad y la precisión de los trabajos, contar con la ayuda de las tecnologías (y, como veremos más adelante, de los corpus lingüísticos) no se presenta ya como una opción, sino más bien como una necesidad y obligación. Por ello, el objetivo principal de este trabajo gira en torno a hacer partícipe a la interpretación de las tecnologías de la traducción e interpretación mencionadas con anterioridad a través de la metodología del corpus centrándose, sobre todo, en la evaluación de la calidad de las interpretaciones en el ámbito judicial (y más concretamente, de los niveles morfosintáctico y léxico) a través de un corpus compuesto por textos jurídicos que nos permita estudiar las interpretaciones, así como el propio lenguaje jurídico y judicial más de cerca.

Hoy en día, las tecnologías lingüísticas suponen una gran ayuda en este contexto. Por ello, el objetivo es hacer que nuestro producto, en este caso la interpretación, gane en calidad y fiabilidad gracias a las tecnologías mencionadas en materia de Lingüística de corpus. A partir de ahora es muy necesario desglosar los puntos principales que funcionarán a modo de sostén durante todo el desarrollo del trabajo:

- a) Contextualización de la interpretación judicial en España
- b) Evaluación de las interpretaciones en el ámbito judicial
- c) Tratamiento del lenguaje jurídico en el proceso interpretativo a través del corpus escogido
- d) Creación semiautomática de un glosario para el intérprete judicial

2. Corpus e interpretación

Una definición bastante completa y aceptada en general es la que ofrecen Bowker y Pearson (2002: 9), en la que un *corpus* es «a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria». Existen otras definiciones anteriores a la mencionada con anterioridad pero que, sin duda alguna, conciben los corpus igualmente de forma empírica. Destacamos, por ello, la forma en la que McEnery y Wilson (1996: 87) conciben el concepto de *corpus*: «a body of text which is carefully sampled to be maximally representative of a language or language variety». De estas dos definiciones se extraen las cuatro piezas que determinan el concepto de corpus en su totalidad; hablaríamos por tanto de una **colección de textos** que han de ser **representativos en una lengua o parte de la misma**, reunidos mediante **formato electrónico** y atendiendo a una serie de **criterios específicos** (Corpas Pastor, 2001 y 2008).

Partimos de la base de un corpus lo forma una gran cantidad de textos. Hoy día, con la revolución tecnológica, llevar a cabo esa recopilación de textos de forma manual se concibe como una tarea casi imposible. Por ello, la necesidad de organizar y estructurar los textos queda cubierta gracias a la tecnología que hace que este trabajo se reduzca de forma considerable.

Antes de concluir este epígrafe, es preciso dar a conocer la clasificación que propone Corpas Pastor (2001: 157-159). Esta responde, *grosso modo*, a:

- a) Según el porcentaje y distribución de los textos: corpus grande o extenso, equilibrado, piramidal, monitor, comparable y paralelo.
- b) Según la especialidad de los textos: corpus general, especializado, genérico, canónico, periódico o cronológico y diacrónico.
- c) Según la cantidad de texto contenida: corpus textual y de referencia.

2.1 El corpus en interpretación

Partimos de la base de que en todo proceso interpretativo (y traslativo) la documentación es el cimiento sobre el que se sustenta el trabajo; es algo común a la traducción e interpretación. Por ello, la documentación se concibe como un cimiento fundamental en toda actividad traslativa o interpretativa y una de las competencias básicas a la vez que primarias de todo traductor e intérprete. Satisface por ende las necesidades de estos profesionales, así como de los trabajos que estos llevan a cabo, cubriendo total o parcialmente un campo de especialización. Permite, además, al traductor o intérprete ser cuasi docto en la materia a tratar, ampliando sus conocimientos y, por consiguiente, proporcionando a su trabajo calidad y fiabilidad. Por ello, todo traductor e intérprete ha de documentarse para que su trabajo sea, al fin y al cabo, un «producto de calidad». Claro que, a diferencia de la traducción, el intérprete ha de documentarse con antelación dada la dificultad de hacerlo durante su trabajo. Por ello, y esclarecida la importancia de la documentación, vemos que es aquí donde entran en juego los corpus. Los usos que se podrían distinguir, a saber, del corpus en interpretación son tres:

- a) Estudio de la terminología *in vivo* (Cabré, 1999) para la traducción especializada
- b) Creación de glosarios
- c) Para la figura del docente, evaluación de la calidad de las interpretaciones

Sin embargo, se podría añadir otro factor que ayudase al profesional a alcanzar una calidad óptima en su interpretación, así como ayudar al estudiante a tener una mejor formación: **el corpus oral**¹. Esta figura permitiría, sin alejarse de las definiciones anteriormente expuestas, tener acceso a archivos de sonido reales llevados a cabo en distintas modalidades y campos de especialidad que permitan al intérprete hacer su práctica diaria más amena. Sin embargo, el principal escollo de esta cuestión es la poca cantidad de corpus orales que existen para propulsar la metodología, en este caso, de la interpretación (y más concretamente de la interpretación simultánea). Existiendo factores críticos en el proceso interpretativo, surge la necesidad de crear corpus orales que abarquen diversas temáticas y que, por supuesto, estén disponibles en el mayor número de lenguas posibles. Con ello, se pretende una mejoría y la eliminación (o reducción) de los esos factores críticos en el trabajo del intérprete.

¹ En interpretación, el tipo de corpus que se usa es mayoritariamente el escrito. Para una mayor aproximación al corpus como formación de intérpretes, consúltese Pérez Pérez (2013) y Araújo y Correia (2015).

El presente trabajo pretende, a través de un corpus escrito, evaluar y crear glosarios que permitan al intérprete judicial tener una mayor aproximación de carácter jurídico y judicial.

3. Introducción a la interpretación en el ámbito judicial

In its narrower sense, court or judicial interpreting is the oral interpretation of speech from one language to another in a legal setting.

(Edwards, 1995: 1)

Partiendo de la definición que propone Edwards (1995), la figura del intérprete judicial cobra especial importancia en escenarios judiciales como, por ejemplo, los juzgados. En esta misma línea, es preciso comentar que en el campo de interpretación para los servicios públicos, la mejor traducción para el término inglés *legal* sería *judicial* ya que, a veces, dicho término funciona como un falso amigo en ciertos contextos no conviene traducirlo como *legal* por la confusión que este término puede generar, ya que, según la RAE (2014), este hace referencia a lo «prescrito por la ley o conforme a ella». Hablamos, por tanto, de la interpretación en casos judiciales y no de la interpretación de la ley como tal, tarea que más bien le es conferida a los jueces. Esta disciplina pertenece, en su conjunto, a la interpretación en los servicios públicos (ISP), donde encontramos otras especialidades afines tales como la interpretación en los servicios sanitarios, interpretación policial, interpretación en los servicios educativos, interpretación en los servicios sociales, etc. El entorno en el que se desenvuelve esta práctica requiere con mucha frecuencia la técnica de la interpretación de enlace (Valero Garcés, 2006). Sin embargo, el caso de los juicios requiere, a veces, la traducción a la vista cuando, por ejemplo, un juez pide que se lea un escrito de acusación al encausado².

Contextualizada la interpretación judicial, pasaremos a ver los parámetros interpretativos para garantizar la calidad en la interpretación judicial y cuáles son las necesidades del intérprete para garantizar dichos parámetros.

3.1 La calidad en interpretación judicial

Cuando hablamos de traducción e interpretación, hablamos de lenguas y culturas diferentes entre sí. En ello encontramos la verdad universal de saber transmitir correctamente la información de una lengua o cultura a otra y hacer que la comunicación, ya sea oral o escrita, llegue a buen puerto. Para garantizar, en este caso, la calidad en los servicios de traducción, encontramos la norma UNE-EN ISO 17100 (2015). Sin embargo, dicha norma no es aplicable en ningún momento a los servicios de interpretación ni a las pruebas

²Anteriormente, esta figura se denominaba *acusado*. Sin embargo, con la Ley Orgánica 13/2015, de 5 de octubre, por la que se modifica la Ley de Enjuiciamiento Criminal, dicho término sufre una modificación por la cual pasa a denominarse *encausado*.

realizadas por el Ministerio de Asuntos Exteriores y Cooperación para la expedición del título de Traductor-Intérprete Jurado. A partir de esto, podemos decir que no hay ley escrita ni pautas a seguir para garantizar la calidad en servicios de interpretación. Sin embargo, esto no ha conllevado a que la calidad en interpretación quedara apartada y no se tuviera en cuenta. Muchos son los autores los que se pronuncian para tratar de esclarecer y aportar una serie de parámetros de incidencia (cf. Collados Aís, Pradas Macías *et al.*, 2007) para unificar criterios e intentar ganar en calidad al realizar un servicio de interpretación.

Muchos son los parámetros existentes para la evaluación de las interpretaciones. Uno de los primeros estudios fue analizar el campo de especialidad de los usuarios (Kurz, 1993 citado en Collados Aís, 2010) para medir la calidad. Tras este estudio se pudo comprobar que la calidad variaba considerablemente al cambiar de un usuario a otro. En el campo de la interpretación judicial, la especialización es muy alta, por ello necesitamos intérpretes altamente cualificados y especializados en materia jurídica y judicial e incluso que no sólo tengan estudios en Traducción e Interpretación, sino que estén también en posesión de algún posgrado relacionado con el Derecho o Relaciones Internacionales. No obstante, en lo que nos concierne, hablaremos de las necesidades lingüísticas del intérprete para garantizar la calidad en términos de terminología, sintaxis y pragmática funcional. Para ello, se seguirán los parámetros propuestos por Collados Aís, Pradas Macías *et al.* (2007):

- a) Parámetro *transmisión correcta* del discurso original
- b) Parámetro *transmisión completa* del discurso original
- c) Parámetro *terminología*

Todos estos parámetros se comentarán en el apartado correspondiente una vez hayamos concluido con la metodología de corpus y la correspondiente extracción terminológica.

3.2 Necesidades del intérprete

En estas últimas décadas se ha hablado y escrito mucho acerca de las habilidades que un intérprete ha de desarrollar para poder realizar de manera correcta su labor. Claro ejemplo de ello es la lista de destrezas que elabora la AIIC (Association Internationale des Interprètes de Conférence) (2017):

- a) Dominio perfecto de su lengua materna en todos los registros y dominios
- b) Dominio completo las lenguas no nativas
- c) Conocimiento de las culturas de los países en los que se hablan las lenguas de trabajo
- d) Compromiso de ayudar a otros a comunicarse
- e) Interés y comprensión de la actualidad, unidos a una curiosidad insaciable
- f) Experiencia internacional fuera del hogar y de la escuela, y una educación general
- g) Buena formación y, al menos, una titulación universitaria
- h) Capacidad para concentrarse a medida que se desarrolle el discurso
- i) Voz agradable
- j) Actitud afable
- k) Capacidad para apaciguar los nervios, el tacto, el juicio y el sentido del humor

- l) Voluntad para respetar las normas de conducta, como por ejemplo la confidencialidad

Nos centraremos en los puntos a) y c) dado que, para asegurar esas destrezas, las necesidades que un traductor necesita son las puramente lingüísticas y que harán referencia explícitamente a la terminología y a la pragmática de un campo de especialización en concreto. A continuación, pasaremos a desglosar cada una de esas necesidades lingüísticas para delimitar claramente su función en cuanto a la labor del intérprete se refiere:

3.2.1 La necesidad terminológica

Constituye el eje central de toda necesidad del intérprete. Las distintas ramas de especialización a las que se puede enfrentar el intérprete son muy variadas. Ello hace que el dominio terminológico de un campo de especialización sea total y que el mensaje se transmita de forma pura sin perder ningún matiz. En este caso, los ámbitos jurídico y judicial son campos cuyas necesidades terminológicas están dotadas de un altísimo grado de especialización y que no todo el mundo conoce a no ser que sea un profesional en el campo del Derecho o, simplemente, que sea traductor o intérprete y tenga una amplia formación tanto en el campo de la traducción jurídica o en el campo del Derecho. Estamos, por tanto, ante lo que Alcaraz Varó, Hughes y Gómez (2014: 15) denominan «lenguas profesionales y académicas», siendo el lenguaje jurídico una de las variantes más importantes del Español Profesional y Académico (*ibid*: 16). Por ello, un intérprete con una buena capacidad terminológica jurídica y judicial tendrá como resultado una interpretación completa y eficaz en el ámbito que trabaje. En el punto 4, y siguiendo la metodología de corpus, intentaremos cubrir esa necesidad para que la precisión lingüística sea mayor.

3.2.2 La necesidad pragmática

Como se viene diciendo a lo largo de todo el trabajo, la función principal de todo traductor e intérprete es saber transmitir el mensaje de forma correcta, sin dejarse nada relevante por el camino. Esta necesidad pragmática está íntimamente relacionada con la terminológica, dado de que la lengua del Derecho es especialmente característica por el oscurantismo en su léxico y que puede ocasionar importantes pérdidas semánticas si no se sabe manejar con la precisión que requiere. Como el intérprete trabajará con dos lenguas y dos culturas, en este caso, la española y la británica, hemos de tener en cuenta las dos grandes familias del Derecho que entrarán en juego: el Derecho civil, emanado de los romanos, y el *Common Law*. Por ello, la necesidad pragmática se hace más patente en este contexto si queremos acercar a nuestros oyentes a un enfoque jurídico común o similar al de las lenguas de partida. Esto se conseguirá si hacemos que nuestras interpretaciones estén dotadas de lo que se denominaría **pragmática funcional**. La **pragmática funcional** consistiría en traducir o interpretar de la lengua origen a la lengua de destino, teniendo en cuenta las diferencias culturales, en este caso las diferencias jurídicas de ambas culturas, de tal forma que, bien se haga una explicación un término jurídico, se opte por buscar un equivalente del mismo en la lengua de destino o se traduzca de forma literal, no se pierdan en ningún momento ni el concepto ni el sentido de lo que ese término significaba en la lengua de origen. Alcanzar el concepto de pragmática funcional en el contexto jurídico no resulta una tarea fácil, pues trabajamos con

dos sistemas jurídicos totalmente opuestos. Sin embargo, siempre podremos conseguir una aproximación cualitativa en términos de equivalencia semántica para que la comunicación llegue a buen puerto y el lector o la persona a la que estemos interpretando no pierda el hilo del discurso en ningún momento.

Todas las necesidades mencionadas con anterioridad intentarán verse resueltas en el siguiente punto a través de la metodología de corpus. La creación de nuestro glosario basado en una compilación de textos jurídicos de la Unión Europea y unas pautas posteriores para la evaluación de la calidad de la interpretación judicial a través de los parámetros **transmisión correcta, transmisión completa y terminología** del discurso original podrán cubrir casi de forma total las necesidades principales del intérprete y, por supuesto, la interpretación resultará ser una comunicación satisfactoria.

4. Creación de glosarios para el intérprete judicial y evaluación de la calidad: estudio de caso

Dejando a un margen el contacto con la agencia o con el cliente, en lo que respecta a la labor de documentación, todo trabajo de traducción o interpretación comienza con una labor de análisis previo de la temática con la que trabajaremos. Sin embargo, la documentación ligada a la traducción incurriría de forma notoria al proceso de descodificación del texto origen (fraseología, terminología...) y al posterior proceso de codificación del texto meta. Por el contrario, la documentación ligada a la interpretación tendría de igual forma un papel notorio pero, en este caso, no trabajamos con texto escrito, por lo que nos es muy difícil saber qué elementos se van a emplear. En la mayoría de las ocasiones, cuando hablamos de las técnicas de interpretación consecutiva o simultánea, lo más seguro es que la interpretación sea de conferencias, por lo que se insta a los organizadores a entregar al intérprete una breve monografía de las ponencias o, en algunas ocasiones, el discurso del orador. En este caso, la labor de documentación estaría relegada, en su mayoría, a cubrir la necesidad terminológica del intérprete así como conocer de primera mano el tema a tratar. Hemos de recordar que el intérprete apenas dispondrá de tiempo para documentarse durante la interpretación, por ello la documentación deberá de igual forma seguir teniendo un peso mayor, pero quedará supeditada en este caso a la necesidad terminológica del intérprete.

En nuestro caso haremos referencia al trabajo de un intérprete judicial. Como se ha comentado con anterioridad, la figura de este intérprete desarrolla su labor principalmente en los juzgados. Ello conlleva a tener que manejar el lenguaje jurídico y judicial casi de la misma forma que lo haría un jurista. Por ello, en los siguientes apartados veremos como podremos solventar el tema principal de la terminología a través del corpus propuesto para, posteriormente, poder evaluar de forma cualitativa la calidad terminológica y de transmisión del discurso hacia la lengua de destino.

4.1 Descripción del corpus

Este punto constituye, junto con el siguiente, una de las piedras angulares de nuestro trabajo. Una vez tenemos la temática con la que trabajaremos concertada, procedemos al diseño de nuestro corpus. Para el diseño de nuestro corpus nos podemos ayudar del buscador

de CURIA, el cual nos permite, mediante diferentes criterios, afinar nuestra búsqueda y prescindir de aquello que nos resulte irrelevante.

Dicho diseño seguirá los criterios propuestos por Torruella y Llisterri (1999):

- a) Género textual
- b) Lengua
- c) Corpus comparable o paralelo
- d) Textos completos o parciales

Género textual

A día de hoy nos encontramos con una gran variedad de géneros textuales jurídicos. El propicio para nuestro glosario es el de las sentencias. Según la Real Academia y el Consejo General del Poder Judicial (2016), una *sentencia* es una «resolución que decide definitivamente el pleito o causa en cualquier instancia o recurso, o que, según las leyes procesales, debe revestir esta forma». Este documento, junto con las providencias y los autos, forman las órdenes judiciales. Sin embargo, los autos y las sentencias, a diferencia de las providencias, están motivadas. Por ello, y al ser el documento que pone fin al proceso judicial, la sentencia puede considerarse como el más amplio y completo. En definitiva, esto ha sido lo que ha llevado a la sentencia a ser el género escogido para la compilación del corpus. Nuestras sentencias estarán dictadas entre los años 2014 y 2018 y se extraerán de CURIA, el portal web del Tribunal de Justicia de la Unión Europea donde podemos encontrar una recopilación en versión electrónica de la jurisprudencia del Tribunal.

The image shows a screenshot of the CURIA search interface. The interface is organized into a grid of search criteria on the left and their corresponding input fields on the right. At the top right, there are buttons for 'Vaciar los campos' (Clear fields) and 'Buscar' (Search). The search criteria include:

- Situación de los asuntos:** Radio buttons for 'Todos los asuntos', 'Asuntos concluidos' (selected), and 'Asuntos pendientes'.
- Tribunal:** Radio buttons for 'Todos', 'Tribunal de Justicia' (selected), 'Tribunal General', and 'Tribunal de la Función Pública'.
- Número de asunto:** A text input field with a placeholder '(p. ej.: C-17/05; 17/05)'.
- Nombre de las partes:** A text input field.
- Documentos:** A dropdown menu showing 'Documentos publicados en la Recopilación : Sentencias'.
- ECLI:** A text input field.
- Palabras del texto:** A text input field.
- Período o fecha:** A date range selector with 'Fecha de pronunciamiento' selected. The range is set from '01/01/2014' to '13/05/2018'. Other options include 'Desde hace 8 días', 'Desde hace 1 mes', 'Desde hace 1 año', and 'Desde hace 5 años'.
- Materia:** A text input field.
- Procedimiento y resultado:** A text input field.
- Citas de jurisprudencia o de legislación:** Radio buttons for 'Buscar en' with 'La motivación' selected, and 'El fallo', 'Las conclusiones'. A 'Categoría' dropdown menu is also present.
- Plan de clasificación sistemática:** A text input field.
- Formación del Tribunal:** A radio button for 'Ampliar al antiguo/nuevo plan' is selected.
- Juez Ponente:** A text input field.
- Abogado General:** A text input field.
- Origen de una cuestión prejudicial:** A text input field.
- Versión lingüística auténtica:** Radio buttons for 'Lengua de procedimiento' (selected) and 'Lengua de las conclusiones'. A dropdown menu shows 'español', 'inglés'.

At the bottom left, there is a link for 'Comentarios y sugerencias'. At the bottom right, there are buttons for 'Vaciar los campos' and 'Buscar'.

Figura 1: Buscador de recopilaciones electrónicas en CURIA

Lengua

El par de lenguas con el que trabajaremos será inglés y español. Por ello, buscaremos sentencias en inglés con su traducción en español. Al ser un portal de la Unión Europea, CURIA nos permite filtrar por todas las lenguas oficiales de la Unión, lo cual hace que la variedad lingüística sea mucho mayor.

Corpus comparable o paralelo

Como la misión principal de nuestro corpus es la creación de un glosario, es de obligada necesidad que nuestro corpus sea paralelo para una mayor facilidad a la hora de extraer el glosario. Además, para la posterior evaluación de las interpretaciones en ámbito judicial, nos servirá de gran ayuda el hecho de que podamos comparar el texto en su versión bilingüe para analizar el discurso y la terminología en contexto mediante la herramienta KWIC (*Key Word in Context*).

Textos completos o parciales

Para que nuestro corpus gane en calidad terminológica, los textos estarán completos. Nos prescindiremos, en ningún momento, de partes de las sentencias extraídas dado que de por sí, este género textual no admite quebrantamiento alguno entre sus partes.

Analizados todos nuestros criterios de diseño, procederemos ya a la compilación del corpus. Para ello, seguiremos los 4 pasos propuestos por Seghiri (2011):

- a) Búsqueda
- b) Descarga
- c) Normalización
- d) Almacenamiento

La búsqueda de nuestros textos estará siempre vinculada al género textual que hayamos propuesto. En este caso ya tenemos identificado el portal desde el cual descargaremos las sentencias. En caso contrario, nos tendríamos que valer de palabras clave relacionadas con nuestra búsqueda: *sentencias*, *Tribunal Superior de Justicia*, *Unión Europea*, etc. Una vez tengamos identificados un portal o diferentes portales de los que extraer nuestros textos, hemos de estar muy atentos al proceso de descarga, ya que trabajaremos con *software* que no reconocen ciertos formatos.

A la búsqueda le sigue la fase de descarga. A día de hoy, la gran mayoría de textos descargables de Internet se encuentran en formato *Portable Document Format* (.pdf) y la mayoría de *software* encargados del análisis de corpus no trabajan con este tipo de documentos. Este problema se verá solucionado en la siguiente fase.

La fase de normalización consiste en convertir nuestros archivos no reconocibles por el programa de análisis de corpus por un formato que este sí reconozca. La mayoría de estos trabajan con archivos de texto plano (.txt). Esta conversión se puede realizar desde el propio lector de archivos en .pdf, el cual permite exportar dicho archivo a texto plano. En caso

contrario, en Internet encontramos herramientas online que permiten la conversión, como por ejemplo *PDFtoText*.

Por último, nos encontramos con la fase de almacenamiento, centrada en organizar nuestros textos por lengua y con una denominación que nos permita distinguir entre el texto origen y su traducción. Como ejemplo podríamos proponer SENTEN 1 para la primera sentencia en lengua inglesa y SENTES 1 para su traducción en español.

Seguidos tanto el proceso de diseño como los pasos para la formación, el resultado ha sido la compilación de un corpus de 40 sentencias de recursos de casación del Tribunal de Justicia de la Unión Europea en inglés y sus traducciones al español. Una vez lo tenemos todo reunido y organizado, procederemos a su análisis y a la extracción del glosario para el intérprete.

4.2 Metodología y sistema de gestión

Nuestro glosario contendrá las palabras con mayor índice de frecuencia del corpus compilado. Para ello, nos valdremos de ParaConc³ (Barlow, 2001), un *software* que nos permite extraer los términos ordenados por orden de mayor frecuencia en ambas lenguas del corpus complicado (en nuestro caso, español e inglés), para así poder ver qué palabras se usan con mayor frecuencia en un determinado género textual y poder construir así nuestro glosario. Una vez introduzcamos nuestra lista de textos en la herramienta, así como un listado de *stopword lists*, es decir, listas de palabras vacías de significado para que la herramienta nos muestre únicamente los términos relevantes del campo en el que estamos trabajando, la herramienta nos mostrará un listado de palabras ordenadas por frecuencia de aparición tal y como muestra la siguiente figura:

³ ParaConc se puede descargar desde <http://www.paraconc.com/demo.html>. La versión con la que hemos trabajado es una demostración, pero permite la extracción de un gran número de términos, así como obtener hasta un máximo de 150 resultados en una búsqueda de equivalencias.

The screenshot shows the ParaConc software window with a 'Corpus Frequency List - All' dialog box. The dialog displays two columns: English (United Kingdom) and Spanish (Modern Sort). Each column has three sub-columns: Count, Pct, and Word. The English list starts with 'court' (1122, 0.9494%) and 'appeal' (991, 0.8385%). The Spanish list starts with 'tribunal' (1214, 0.9050%) and 'general' (901, 0.6717%). The status bar at the bottom indicates '15 parallel files loaded' and '118.182 / 134.145'.

English (United Kingdom)			Spanish (Modern Sort)		
Count	Pct	Word	Count	Pct	Word
1122	0.9494%	court	2661	1.9837%	
991	0.8385%	appeal	1214	0.9050%	tribunal
886	0.7497%	general	901	0.6717%	general
770	0.6515%	commission	892	0.6650%	comisión
738	0.6245%	judgment	787	0.5867%	sentencia
568	0.4806%	issue	784	0.5844%	apartado
561	0.4747%	regulation	730	0.5442%	reglamento
441	0.3732%	decision	677	0.5047%	artículo
349	0.2953%	case	547	0.4078%	casación
328	0.2775%	law	462	0.3444%	recurso
324	0.2742%	ground	416	0.3101%	motivo
315	0.2665%	eurocean	380	0.2833%	decisión
			350	0.2609%	derecho

Figura 2: Listado de términos ordenados por frecuencia

Para la extracción de este listado de frecuencias, debemos dirigirnos al menú *Frequency*, seleccionar la opción *Frequency Order* y después, *All*.

Una vez finalizados estos pasos, estaremos ante lo que ha sido una extracción automática de términos ordenados por frecuencia. Ya solo falta buscar sus equivalentes en las listas y, si tenemos cualquier duda, consultar en un diccionario especializado. Lo que hemos conseguido con esto es conocer qué palabras se usan más en un campo de especialización en concreto para que así tanto el intérprete como el traductor lo tengan como referencia. En el siguiente apartado describiremos nuestro glosario completo que servirá al intérprete durante su trabajo.

4.3 Descripción del glosario

Para lograr una transmisión correcta y completa a la hora de interpretar, consideramos que nuestro glosario deberá contener los siguientes campos:

- a) Término en inglés (EN)
- b) Definición del término en inglés. Dicha definición deberá extraerse de un diccionario especializado.
- c) Equivalente o traducción en español (ES)
- d) Contexto en ambas lenguas. El contexto nos lo dará el programa Antconc.

4.4 Evaluación de la calidad

Como se ha venido comentando a lo largo del presente trabajo y extraído el glosario a través del corpus compilado, lo que nos ocupa para finalizar es la evaluación de la calidad del servicio prestado. Contar ya con herramientas como el corpus y el glosario supone un avance sobresaliente. Por ello, y tal como se comentó en el apartado 3.1 y siguiendo a Collados Aís, Pradas Macías *et al.* (2007), estos son los dos parámetros de incidencia que procederemos a evaluar:

- a) Parámetro **transmisión correcta** del discurso original
- b) Parámetro **terminología**

Estos parámetros podrían considerarse análogos dado que por lo que se vela es por que se transmita el mensaje en su totalidad y que no haya confusión pragmática durante el discurso, es decir, que se transmita el mensaje de forma cohesiva, coherente y, por supuesto, fidedigna, quedando solventado el parámetro **transmisión completa** del discurso original mediante la fidelidad. Sin embargo, el parámetro terminología es un complemento imprescindible para que el parámetro transmisión correcta del discurso original quede patente en todo momento durante la actividad interpretativa. Ambos parámetros quedan solventados gracias al corpus compilado. La función *Concordance* que nos brinda AntConc (Anthony, 2014) nos permite trabajar con palabras clave en contexto (Key Word in Context, KWIC) así como estudiar el número de ocurrencias en nuestro corpus.

Caso: El uso de colocaciones en el lenguaje jurídico y judicial

En el lenguaje jurídico existen muchas cadenas sintácticas que el público general suele desconocer: *conocer de (un caso)*, *interponer (una demanda)*, *contraer (matrimonio)* y un largo etcétera. Todos los ejemplos anteriores, atendiendo a la clasificación propuesta por Koike (2001)⁴, se corresponden con las estructuras **verbo+sustantivo (CD)** y **verbo+preposición+sustantivo**. De nada sirve que nuestro intérprete o estudiante sepa que *appeal* se traduce por *recurso* si no conoce la colocación, en este caso el verbo que funciona con el término anterior. Por ello, la solución a ello se encuentra en nuestro corpus. A través de la opción *Clusters⁵/N-Grams* de nuestro software AntConc (Anthony, 2014) podremos ver los co-términos (verbos) que suelen ir junto a *recurso*.

Supongamos que el intérprete tiene que trasladar el contenido de la frase inglesa *dismiss an appeal*. Tras consultar en varios diccionarios, nos encontramos con que *dismiss* puede significar *desestimar*, *denegar*, *anular* o *rechazar*.

Supongamos que, por otro lado, tiene que trasladar al español *bring an appeal*. Tras consultar en varios diccionarios, nos encontramos con que *bring* puede significar *poner*, *interponer*, *presentar* o *formular*. Ante estos casos, ¿qué verbo es el más correcto y preciso para ambas oraciones?

La imagen siguiente muestra los *clusters* extraídos de nuestro corpus de textos en lengua española a través de la herramienta AntConc para poder ver así qué verbo es el que se emplea con este término:

⁴ Para la clasificación completa de colocaciones, consúltese Koike (2001). Esta completa la propuesta por Corpas Pastor (1996).

⁵ En palabras de Baker, Hardie y McEnery (2006), *cluster* se usa para describir un grupo de palabras que aparece de forma secuencial en un texto.

Total No. of Cluster Types		121		Concordance	Concordance Plot
Total No. of Cluster Tokens		479			
Rank	Freq	Range	Cluster		
1	36	14	de su recurso		
2	30	1	sala de recurso		
3	20	11	mediante su recurso		
4	18	14	desestime el recurso		
5	18	14	sobre el recurso		
6	17	7	interponer un recurso		
7	16	8	de un recurso		
8	16	6	motivo del recurso		
9	15	13	cuando el recurso		
10	15	11	desestimar el recurso		
11	14	14	objeto un recurso		
12	14	7	que el recurso		
13	12	8	interpuso un recurso		
14	11	5	en su recurso		
15	11	3	objeto del recurso		
16	10	6	del presente recurso		
17	9	1	interposición del recurso		
18	8	4	inadmisibilidad del recurso		
19	8	4	vías de recurso		
20	7	3	de ese recurso		
21	7	2	pretensiones del recurso		
22	6	5	desestimó el recurso		
23	6	5	marco del recurso		

Figura 3: Listado de clusters en torno al término base recurso

Con esta imagen vemos que lo más correcto sería hablar de *desestimar un recurso* y de *interponer un recurso*. El uso de ciertos verbos como *anular* sería incorrecto, dado que lo que en contexto jurídico se anula o se revoca serían las sentencias. No estamos, por ello, transmitiendo el mensaje de forma correcta y clara ya que perderíamos ese matiz semántico que posee el verbo.

Por ello, y tras comprobar el ejemplo anterior, vemos que ambos parámetros son complementarios entre sí y uno no puede verse resuelto sin la ausencia del otro. En nuestro contexto, dejando a un lado parámetros como el **acento**, la **agradabilidad de la voz**, la **fluidez** y la **cohesión lógica** (cf. Collados Aís, Pradas Macías *et al.*, 2007), la **transmisión correcta del discurso** y la **terminología** son los puntos de partida para poder realizar una evaluación de calidad adecuada en contextos altamente especializados como es el jurídico y el judicial. El resto de los parámetros podrán verse afectados si no se logra un uso correcto de la terminología del campo de especialidad en el que estamos trabajando ni se transmite el discurso correctamente.

Para concluir, a través de la jurisprudencia de EurLex (Unión Europea, s.f.)⁶ podremos disponer de nuestros textos de forma paralela para realizar consultas rápidas. Es una opción muy visual que ayuda bastante durante el proceso de búsqueda de concordancias para así

⁶ Acceso directo al buscador de jurisprudencia de EurLex desde <https://eur-lex.europa.eu/collection/eu-law/eu-case-law.html?locale=es>

poder estudiar igualmente los parámetros **terminología y transmisión correcta y completa** del discurso original.

<p>Provisional text JUDGMENT OF THE COURT (Fifth Chamber) 17 May 2018 (*) (Reference for a preliminary ruling — Intellectual property — Trade-mark law — Regulation (EC) No 207/2009 — Article 13 — Exhaustion of the rights conferred by a trade mark — Parallel imports — Repackaging of the product bearing the mark — New labelling — Conditions applicable to medical devices) In Case C-642/16, REQUEST for a preliminary ruling under Article 267 TFEU from the Bundesgerichtshof (Federal Court of Justice, Germany), made by decision of 6 October 2015, and received at the Court on 14 December 2016, in the proceedings Junek Europ-Vertrieb GmbH v Lohmann & Rauscher International GmbH & Co. KG THE COURT (Fifth Chamber), composed of J.L. da Cruz Vilaça, President of the Chamber, E. Levits, A. Borg Barthet (Rapporteur), M. Berger and F. Biltgen, Judges, Advocate General: M. Bobek, Registrar: K. Malacek, Administrator, having regard to the written procedure and further to the hearing on 24 January 2018, after considering the observations submitted on behalf of: - Junek Europ-Vertrieb GmbH, by J. Sachs and C. Sachs, Rechtsanwälte, - Lohmann & Rauscher International GmbH & Co. KG, by C. Rohnke and M. Stütz, Rechtsanwälte, - the German Government, by T. Henze and M. Hellmann, acting as Agents, - The Italian Government, by G. Palmieri, acting as Agent, assisted by M. Russo, avvocato dello Stato, - the European Commission, by G. Braun, É. Gippini Fournier and T. Scharf, acting as Agents, having decided, after hearing the Advocate General, to proceed to judgment without an Opinion, gives the following judgment</p>	<p>Edición provisional SENTENCIA DEL TRIBUNAL DE JUSTICIA (Sala Quinta) de 17 de mayo de 2018 (*) «Procedimiento prejudicial — Propiedad intelectual — Derecho de marcas — Reglamento (CE) n.º 207/2009 — Artículo 13 — Agotamiento del derecho conferido por la marca — Importación paralela — Reenvasado del producto que lleva la marca — Nuevo etiquetado — Requisitos aplicables a los dispositivos médicos» En el asunto C-642/16, que tiene por objeto una petición de decisión prejudicial planteada, con arreglo al artículo 267 TFUE, por el Bundesgerichtshof (Tribunal Supremo Civil y Penal, Alemania), mediante resolución de 6 de octubre de 2015, recibida en el Tribunal de Justicia el 14 de diciembre de 2016, en el procedimiento entre Junek Europ-Vertrieb GmbH Y Lohmann & Rauscher International GmbH & Co. KG , EL TRIBUNAL DE JUSTICIA (Sala Quinta), integrado por el Sr. J.L. da Cruz Vilaça, Presidente de Sala, y los Sres. E. Levits y A. Borg Barthet (Ponente), la Sra. M. Berger y el Sr. F. Biltgen, Jueces; Abogado General: Sr. M. Bobek; Secretario: Sr. K. Malacek, administrador; habiendo considerado los escritos obrantes en autos y celebrada la vista el 24 de enero de 2018, consideradas las observaciones presentadas: - en nombre de Junek Europ-Vertrieb GmbH, por el Sr. J. Sachs y la Sra. C. Sachs, Rechtsanwälte; - en nombre de Lohmann & Rauscher International GmbH & Co. KG, por los Sres. C. Rohnke y M. Stütz, Rechtsanwälte; - en nombre del Gobierno alemán, por los Sres. T. Henze y M. Hellmann, en calidad de agentes; - en nombre del Gobierno italiano, por la Sra. G. Palmieri, en calidad de agente, asistida por la Sra. M. Russo, avvocato dello Stato; - en nombre de la Comisión Europea, por los Sres. G. Braun, É. Gippini Fournier y T. Scharf, en calidad de agentes; vista la decisión adoptada, oído el Abogado General, de que el asunto sea juzgado sin conclusiones; dicta la siguiente</p>
---	--

Figura 4: Ejemplo de texto paralelo en EurLex

5. Conclusiones

El tratamiento de las LSP (*Languages for Specific Purposes*) nunca ha sido una tarea fácil, sobre todo porque la mayoría de traductores o intérpretes no suelen ser doctos en la materia que están trabajando. La calidad en los trabajos de traducción e interpretación puede verse muy reducida si no se lleva a cabo una labor de documentación precisa. Como hemos podido comprobar, esa labor se ve fuertemente enriquecida gracias a las TIC en conjunto con la metodología de corpus. El estudio de ejemplos extraídos de lenguaje real ayuda a afrontar las lenguas de especialidad de forma más metodológica, precisa e intuitiva, y estudiando la materia como si fuésemos profesionales de ese campo de especialidad. Por consiguiente, los trabajos ganarán en calidad y en fidelidad.

Este trabajo ha pretendido demostrar dos aspectos. En primer lugar, que el uso de corpus especializados puede permitir que el intérprete realice sus propios glosarios basados en contextos lingüísticos reales. Debemos recordar que el intérprete, al igual que el traductor, basa su trabajo diario en el trabajo con lenguas, y más específicamente, que la mayor parte de su trabajo se centra en el trabajo de la terminología de áreas temáticas altamente especializadas. Por ello, el glosario supone el cimiento principal de su quehacer diario. Si este cimiento falla, muy difícilmente se podrá asegurar un buen trabajo *a posteriori*. Un glosario bien fundamentado de principio a fin es sinónimo de seguridad, fiabilidad y calidad en el trabajo. Este primer punto nos lleva al segundo, que se centra en evaluar la calidad de

la interpretación, en este caso en el ámbito judicial, a través de la metodología de corpus. Tal y como se ha comentado en el apartado 4.4, la terminología es el punto de partida, a la vez que un requisito imprescindible para poder decidir entre si se ha conseguido la transmisión completa y correcta del discurso o no. Por ello, este parámetro ha de ser un muro infranqueable. Además, el uso de la herramienta KWIC nos permitirá ver si ese término se usa en el contexto adecuado y no se aleja del discurso y de la intención del orador ya que estaremos nuevamente ante un error terminológico y pragmático. Por supuesto, todo el marco práctico del presente trabajo es fácilmente extrapolable a la labor de traducción, que puede complementar perfectamente la labor de interpretación.

Para finalizar y teniendo en cuenta los objetivos propuestos al principio de este trabajo y tras observar todo el conjunto práctico del mismo, solo nos queda afirmar que las tecnologías lingüísticas, y más concretamente, el uso de corpus en traducción e interpretación, suponen un avance de gran categoría que dará lugar a muchas más facilidades para que la labor traslativa e interpretativa se vea casi automatizada, aunque siempre tenemos que tener en cuenta que es el traductor y el intérprete esa figura humana que ha de tener en todo momento la última palabra.

Referencias bibliográficas

- AENOR (Asociación Española de Normalización y Certificación) (2015): *Norma española UNE-EN ISO 17100 para los servicios de traducción*. Madrid: AENOR.
- AIIC (Association Internationale des Interprètes de Conférence) (2017). Advice to students wishing to become conference interpreters. Disponible en: <<https://aiic.net/page/56/advice-to-students-wishing-to-become-conferenceinterpreters/lang/1>>.
- Alcaraz Varó, E., Hughes, B. y Gómez, A. (2014). *El español jurídico*. Barcelona: Ariel.
- Anthony, L. (2014). *AntConc* (Versión 3.4.3). [Programa informático]. Tokyo: Waseda University. Disponible en: <<http://www.laurenceanthony.net/>>.
- Araújo, S. Y Correia, A. (2015). Corpus y formación de intérpretes: un estudio exploratorio. En M.T. Sánchez Nieto, S. Álvarez Álvarez *et al.* (Eds.), *Metodología y aplicaciones en la investigación en traducción e interpretación con corpus* (pp. 165-181). Valladolid: Ediciones Universidad de Valladolid. Disponible en: <<https://dialnet.unirioja.es/servlet/libro?codigo=580083>>.

- Baker, P., Hardie, A. y McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edimburgo: Edinburgh University Press.
- Barlow, M. (2001). *ParaConc* (Versión 1.0). [Programa informático]. Houston: Atheslan.
Disponible en: <<http://www.paraconc.com/demo.html>>.
- Bowker, L. y Pearson, J. (2002). *Working with Specialized Language: A practical guide to using corpora*. Londres/Nueva York: Routledge.
- Cabré, M.T. (1999). *La terminología, representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Collados Aís, A., Pradas Macías, E.M. et al. (eds.) (2007). *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia*. Granada: Comares.
- Collados Aís, A. (2010). La evaluación de la calidad en interpretación simultánea: pautas evaluadoras según usuarios. En R. Rabadán, M. Fernández López y T. Guzmán González (Coords.), *Lengua, traducción, recepción: en honor de Julio César Santoyo* (pp. 67-90). León: Universidad de León.
- Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid: Gredos

- Corpas Pastor, G. (2001). Compilación de un corpus ad hoc para la enseñanza de traducción inversa especializada. *TRANS. Revista de Traductología*, 5, 155-184.
- Corpas Pastor, G. (2008). *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt am Main: Peter Lang.
- Corpas Pastor, G. (2012). Corpus, Tecnología y Traducción. En M. Casas (Dir.) y M. García Antuña (Ed.), *XII Jornadas de Lingüística* (pp. 75-98). Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Edwards, A.B. (1995). *The Practice of Court Interpreting*. Ámsterdam/Filadelfia: John Benjamins.
- Koike, K. (2001). *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá de Henares.
- Kurz, I. (1993). Conference interpretation: expectations of different user groups. *The Interpreters' Newsletter*, 5. Disponible en: <<https://www.openstarts.units.it/bitstream/10077/4908/1/KurzIN5.pdf>>.
- McEnery, T y Wilson, A. (1996). *Corpus Linguistics*. Edimburgo: Edinburgh University Press.
- Muñoz Ramos, M. (2012). *Hacia una redefinición de las materias de tecnologías de la traducción en el marco del EEES: conceptos, competencias, procesos de enseñanza-*

aprendizaje y evaluación (Tesis doctoral). Universidad de Málaga, Málaga. Disponible en: <<https://riuma.uma.es/xmlui/handle/10630/5048?show=full>>.

Pérez Pérez, P.S. (2013). *La enseñanza de la interpretación de conferencias: un estudio empírico con metodología de corpus* (Tesis doctoral inédita). Universidad de Málaga, Málaga.

Real Academia Española y Consejo General del Poder Judicial. (2016). *Diccionario del español jurídico*. Madrid: Espasa-Calpe.

Real Academia Española. (2014). *Diccionario de la lengua española*. Madrid: Espasa-Calpe.

Seghiri, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad. *RLA. Revista de Lingüística Teórica y Aplicada*, 49(2). Doi: <http://dx.doi.org/10.4067/S0718-48832011000200002>.

Torruella, J. y Llisterri, J. (1999). Diseño de corpus orales y textuales. En J.M. Bleca, G. García Clavería *et al.* (Eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos* (pp. 45-81). Barcelona: Universitat Autònoma de Barcelona.

Valero Garcés, C. (2006). *Formas de mediación intercultural. Traducción e interpretación en los servicios públicos. Conceptos, datos, situaciones y práctica*. Granada: Comares.

Unión Europea (s.f.). *Eur-Lex*. Disponible en: <<https://eurlex.europa.eu/homepage.html>>.