



Compilación y análisis de un corpus *ad hoc* como herramienta de documentación electrónica en Traducción e Interpretación en los Servicios Públicos (TISP)

M^a del Mar Sánchez Ramos

Recibido: 2 de junio de 2016 / Aceptado: 14 de marzo de 2017

Resumen. La formación de profesionales cualificados que sirvan como comunicadores con la población extranjera se presenta como imprescindible en la disciplina conocida como Traducción e Interpretación en los Servicios Públicos (TISP), principalmente en aquellos casos en los que los profesionales de TISP provienen de entornos académicos distintos. Esta área de investigación engloba distintos contextos como el sanitario, la educación, el plano legal o el administrativo. Las fuentes de documentación tradicionales (textos paralelos, glosarios y bases de datos léxicas o diccionarios especializados) no son suficientes en el caso de las tareas específicas que pueden darse en TISP. Este artículo tiene como objetivo ilustrar el uso de corpus *ad hoc* como herramientas de documentación electrónica en un curso de posgrado en TISP. Se describen las distintas fases que forman parte en el diseño y el análisis del mismo.

Palabras clave: Lingüística de corpus, corpus especializados, TISP, programas informáticos de gestión de corpus

Compiling and analysing *ad hoc* corpora as electronic documentation tools in Public Translation and Interpreting (PSTI)

Abstract. Training qualified professionals to serve as communicators with foreigners is seemed as a must in the discipline known as Public Service Interpreting and Translation (PSIT), especially when PSIT practitioners come from different academic backgrounds. This area of research covers different areas, such as healthcare, educational, legal or administrative settings. The use of different sources of information (parallel texts, different types of glossaries and lexical databases or specialized dictionaries) sometimes is not enough to fill the gap that specific tasks in PSIT may present. This paper aims at illustrating the use of *ad hoc* corpora as electronic documentation tools in a postgraduate PSIT training context. We describe the different stages involved in the design and analysis of an *ad hoc* corpus.

Keywords: Corpus linguistics, specialized corpus, PSIT, concordance software

Sumario: 1. Introducción. 2. Documentación electrónica y traducción. 3. Estudios de traducción basados en corpus. 4. Diseño y análisis de un corpus *ad hoc*. 5. Conclusión

Cómo citar: Sánchez Ramos, M.^a del M. (2017) Compilación y análisis de un corpus *ad hoc* como herramienta de documentación electrónica en Traducción e Interpretación en los Servicios Públicos (TISP), en *Estudios de Traducción* 7, 177-190.

¹ Universidad de Alcalá de Henares
mar.sanchezr@uah.es

1. Introducción

El continuo cambio que está experimentando la sociedad actual ha hecho que se establezcan nuevas modalidades de comunicación (Mikkelson, 1996). Estos cambios inciden directamente en el modo de concebir la formación de traductores e intérpretes que intervienen en la comunicación entre los proveedores de servicios públicos y sus usuarios (Valero Garcés, 2013), que, en su mayoría, desconocen o no dominan con fluidez la lengua de uso. Valero Garcés (2005: 101) resalta como principal característica de la disciplina denominada TISP la de ser

un medio para establecer la comunicación con un público específico que responde a una minoría cultural y lingüística, que posee un nivel educativo y adquisitivo generalmente inferior al de la mayoría y que, con frecuencia, desconoce o no domina la nueva realidad social del país en el que se encuentra, viéndose agravada tal situación por la distancia cultural entre ambos.

Distintos programas se han centrado en consolidar esta disciplina y son variadas las propuestas formativas existentes que claman una mayor atención por la vertiente pedagógica dentro de esta área de investigación (Angelelli, 2006; D'Hayer, 2012; Mikkelson, 2015; Vargas Urpí, 2016). La formación, por tanto, parece cuanto menos indispensable. Siguiendo esta línea pedagógica se ubica el Máster en Comunicación Intercultural, Traducción e Interpretación en los Servicios Públicos² (Universidad de Alcalá), un programa de posgrado que cuenta con más de una década de trayectoria, y se asienta sobre tres pilares fundamentales: investigación, formación y práctica (Valero Garcés, 2013: 218). Su plan de estudios se divide en cinco módulos: 1) Comunicación Intercultural e Interlingüística (módulo en línea); 2) Módulo sobre Traducción e Interpretación Sanitaria; 3) Módulo sobre Traducción e Interpretación Jurídico-Administrativa; 4) Prácticas en empresas e instituciones; y 5) Trabajo fin de Máster. Dentro de los módulos dedicados a la traducción sanitaria y la traducción jurídico-administrativa se imparte la asignatura Técnicas y Recursos aplicados a TISP, marco docente donde se enmarca nuestra propuesta, y que abarca, principalmente, contenidos que pueden ayudar al futuro traductor e intérprete en TISP en sus labores de documentación. Entre los objetivos que guían dicha asignatura, y por ende nuestro trabajo, se encuentran: a) proporcionar las herramientas documentales, preferentemente electrónicas, que puedan facilitar el proceso de trasvase en ámbitos especializados y cubran las necesidades documentales de los futuros traductores e intérpretes en TISP; y b) proporcionar la formación necesaria en uso de tecnologías para la extracción de información y elaboración de material documental (EMT, 2009), siguiendo, de este modo, las pautas proporcionadas por la European Master's in Translation (EMT) sobre las distintas competencias que deben trabajarse. De las distintas herramientas informáticas documentales, nosotros abogamos por utilizar los gestores de gestión de corpus por dos razones: 1) suelen ser herramientas (casi) desconocidas por el alumnado; 2) son herramientas que, además de solventar

¹ Este programa de formación de posgrado pertenece a la prestigiosa red *European Commission's European Master's in Translation* (EMT). Posibilita la formación en distintos pares de lenguas (español e inglés, francés, árabe, chino, ruso o rumano). Puede encontrarse más información en el siguiente enlace: <http://www3.uah.es/master-tisp-uah/presentacion/>

problemas de corte terminológico, fraseológico o conceptual, desarrollan también las destrezas más técnicas al trabajar con nuevos formatos y programas específicos.

Tras un breve repaso de las necesidades documentales en traducción y el uso de corpus en los Estudios de Traducción, detallaremos cómo se ha integrado la metodología de corpus (diseño y compilación de un corpus y uso de corpus de referencia) en la formación del traductor e intérprete en TISP y cómo los discentes compilaron un corpus *ad hoc* para poder elaborar recursos documentales que les fueran de ayuda en futuros trabajos o encargos de traducción e interpretación.

2. Documentación electrónica y traducción

La automatización y digitalización de la información es constante en nuestros días (Austermül, 2001: 2). Es innegable que las herramientas informáticas han cambiado (y siguen cambiando) la labor de los traductores e intérpretes. Las palabras de Valero Garcés y De la Cruz Cabanillas (2001:11) siguen aun vigentes hoy en día:

Frente a la situación anterior caracterizada por ser una profesión con escasos medios auxiliares (unos cuantos diccionarios y consultas a hablantes nativos [...]), estamos asistiendo a una verdadera revolución [...]. El ordenador constituye el elemento esencial e imprescindible en este proceso.

La documentación, ya sea en formato papel o electrónico, se supone como uno de los pilares básicos de la formación del traductor e intérprete, y aun lo es más en el caso de TISP, ya que los contextos comunicativos suelen ser muy específicos y el acceso a la información suele ser más sesgada. La comprensión del texto original y la reexpresión de mismo se deben al éxito y buen uso que el traductor haga de sus conocimientos lingüísticos y documentales (terminológicos, fraseológicos y textuales), entre otros, formando parte estos últimos de la denominada competencia documental (Pinto Molina y Sales Salvador, 2008).

Palomares Perrault (2000) establece los tres puntos clave por los que un traductor debe acudir a las distintas fuentes documentales: a) para obtener información sobre el texto origen; b) para obtener información terminológica que resuelva problemas de vocabulario especializado; y c) para obtener información fraseológica. El uso de fuentes de documentación viene a solventar carencias que puedan presentarse tanto al inicio como al final del proceso de traducción. Lo expuesto anteriormente adquiere una mayor dimensión al hablar de la traducción especializada, caracterizada por un contenido altamente especializado, una terminología y fraseología específica y una tipología textual concreta (Montalt Resurrecció y González Davies, 2007).

En lo referido al plan de documentación en traducción encontramos la propuesta de Orozco Jutorán (2012: 220), que destaca los distintos tipos de documentación que el traductor especializado debe dominar: la documentación temática o conceptual, la terminológica y la contextual y co-textual. En primer lugar, y relacionado con la documentación temática, es labor del traductor e intérprete el adquirir los conocimientos necesarios de un determinado campo de especialidad para una adecuada comprensión del texto (Orozco Jutorán, 2012: 222). La autora menciona los tesauros y obras que varían en su grado de especialización (textos divulgativos, tex-

tos escritos por especialistas) entre los recursos que pueden ayudar en la documentación temática. La identificación y comprensión de los términos especializados del texto origen es crucial en la llamada documentación terminológica. Las obras que la autora propone son, en este caso, de corte lexicográfico (diccionarios generales, especializados, glosarios, bases de datos terminológicas o diccionarios visuales especializados). Por último, la documentación contextual y co-textual, en donde, y en palabras de la estudiosa: “el traductor debe comprobar que la macroestructura y la microestructura del TM sean adecuadas a la función del TM y que éste se lea de forma fluida y natural por parte del lector final [...]” (Orozco Jutorán, 2012: 225), y aboga por el uso de los textos paralelos como fuente documental.

De la clasificación de Orozco Jutorán (2012) se observa que el proceso de documentación va más allá de la mera consulta a distintas herramientas lexicográficas. Junto con los recursos en línea que puedan encontrarse, textos paralelos, glosarios (monolingües, bilingües y multilingües), bases de datos, portales especializados, el discente debe saber crear sus propios materiales documentales, ya que pueden darse casos de encargos sumamente especializados y en los que los recursos documentales sean escasos. Distintos estudios ya arrojan datos interesantes sobre los hábitos y usos de herramientas documentales electrónicas de los estudiantes de traducción (Cid-Leal y Perpinya-Morera, 2015) y de los profesionales de la traducción (Désilets et al., 2009). Dichos resultados concluyen que tanto estudiantes como profesionales emplean recursos, principalmente electrónicos, y que la formación en este tipo de herramientas así como la búsqueda eficiente de las mismas debe incorporarse en la formación de traductores e intérpretes. En nuestra opinión, y a la vista de dichos distintos estudios y nuestra experiencia como docente, los discentes necesitan desarrollar estrategias de búsqueda en Internet, conocer distintos formatos digitales, programas informáticos de descarga de información en la red y herramientas de tratamiento de textos, por ejemplo, para así poder elaborar su propio material documental que no es otro que material real, o lo que es lo mismo, utilizar la web como fuente para la creación de material auténtico y que se conoce como *Web for Corpus* (Sinclair, 2005).

3. Estudios de traducción basados en corpus

Desde que la académica Mona Baker asentara las bases de los llamados Estudios de Traducción basados en Corpus (*Corpus-Based Translation Studies*) han sido muchas las propuestas que han apostado por el uso de la metodología de corpus en distintos ámbitos de la investigación en traducción. Un corpus puede definirse como:

a computerized collection of authentic texts amenable to automatic or semiautomatic processing or analysis. [...]. The texts are selected according to explicit criteria in order to capture the regularities of a language, a language variety or a sub-language (Tognini-Bonelly (2001: 55)

El uso y estudio de corpus en los Estudios de Traducción cuenta con una línea de investigación ya consolidada. Distintos trabajos centran su atención en estudiar las características específicas de diversos géneros (Corpas, 2008; Sánchez Ramos y Vigier Moreno, 2016) así como las aplicaciones pedagógicas de los mismos (Monzó

Nebot, 2008; Zanettin, 2003); el uso de los corpus como recursos en entornos profesionales (Gallego-Hernández, 2015) o el uso de una metodología de corpus para la compilación de herramientas documentales (Seghiri, 2017). Según Calzada Pérez (2007) los Estudios Traductológicos de Corpus destacan:

por su flexibilidad y capacidad de adaptación, los CTS aúnan metodologías descriptivas y lingüísticas; análisis del proceso y producto; exégesis de detalles o amplios patrones de comportamiento de interés tanto por cuestiones formales como por facetas culturales, ideológicas e incluso literarias. (Calzada Pérez, 2007:216)

En lo que respecta a una clasificación de corpus, diversos estudiosos dibujan una tipología (Bowker, 2003; Granger, 2003), si bien nos parece apropiada la distinción de Laviosa (2002: 36), que resume los principales corpus que se utilizan en los estudios de traducción. Esta académica divide los corpus monolingües en simples (recopilación de textos en una única lengua) y comparables (recopilación de textos originales en una lengua A y textos traducidos en esa misma lengua A); los bilingües quedan clasificados en: paralelos (textos originales en lengua A y sus traducciones en lengua B) y comparables (textos originales en lengua A y textos originales en lengua B); y los multilingües se agrupan en: paralelos (textos originales en lenguas diversas con sus respectivas traducciones), comparables (“bi/multilingual corpus made up to two or more sets from the same subject domain(s)”, Laviosa (2002: 36).

Otro tipo de corpus que cada vez está siendo utilizado en la enseñanza de la traducción, es el llamado corpus *ad hoc* (Aston, 1999). El empleo de este tipo de corpus responde a unas necesidades concretas, como tareas de documentación, o bien como recurso pedagógico para la elaboración de material didáctico, y cuya fuente no es otra que textos electrónicos extraídos de Internet y que han seguido una evaluación de dichos textos y un criterio o protocolo de compilación concreto. Zanettin (2012: 64) señala las ventajas de este tipo de corpus: se trata de corpus creados como respuesta a un problema traductológico concreto, y de ahí su especificidad. El objetivo último es reunir la mayor información posible. La calidad no es la principal preocupación en estos casos, sino la selección de textos que documentan el corpus así como su evaluación tanto temática como textual en relación al texto origen desde el que se parte y que ha generado el diseño del corpus virtual. Este tipo de corpus es considerablemente útil como recursos de documentación en temas especializados, donde los recursos de documentación pueden llegar a ser escasos. Es por ello que puede recurrirse al a compilación de un corpus, si bien hay que ser conscientes de la inversión de tiempo que ello supone.

Independientemente del tipo de corpus, los beneficios pedagógicos del uso de la metodología de corpus en traducción también han sido destacados (Bernardini y Castagnoli, 2008; Bowker y Pearson, 2002: 10). Una de las principales ventajas es el desarrollo de la llamada subcompetencia instrumental (PACTE 2003: 53) o lo que el grupo de expertos de EMT (2009) denomina *information mining competence*. Estos últimos inciden en la necesidad que, debido a los avances tecnológicos en los que vivimos inmersos, los traductores e intérpretes tienen de conocer y manejar herramientas de concordancia. Rodríguez Inés (2010) identifica una subcompetencia dentro de la ya subcompetencia instrumental identificada por PACTE (2003) y habla de “the ability to meet a number of learning outcomes: identifying the princi-

ples that lie at the basis of the use of corpora; creating corpora; using corpus-related software; and solving translation problems by using corpora” (Rodríguez Inés 2010: 253). El uso de los programas de concordancias y el análisis de corpus electrónicos permite a los traductores e intérpretes la realización de una serie de tareas, como puede ser la identificación y verificación de patrones léxicos, fraseológicos y textuales (Neunzig 2003).

4. Diseño y análisis de un corpus *ad hoc*

En este apartado describiremos cómo la metodología de corpus sirvió como recurso de documentación en la clase de Técnicas y Recursos para TISP (Módulo Sanitario). La finalidad era que los alumnos tomaran conciencia de uso de herramientas informáticas documentales adicionales a las ya tradicionales (diccionarios electrónicos, bases de datos en línea, textos paralelos, etc.), desarrollaran aspectos técnicos (tratamiento de textos y manejo de programas informáticos) y supieran crear sus propios recursos documentales basados en textos reales. Para cumplir nuestros objetivos, se diseñó una actividad en la que el alumnado tenía que compilar un corpus *ad hoc* especializado monolingüe (inglés británico) como fuente de documentación conceptual, terminológica y fraseológica como primera aproximación a un texto especializado que versaba sobre la llamada anemia drepanocítica (“sickle-cell anemia”). La compilación de dicho corpus le permitiría la creación de un glosario, la comprensión de las distintas etapas de compilación, el uso de operadores booleanos para restringir las búsquedas en la Web y la manipulación de distintos formatos y programas informáticos concretos (Antconc, HTMLasText).

Como señala Seghiri (2011), es imprescindible establecer unos parámetros de diseño de un corpus o protocolo de compilación. Siguiendo muy de cerca la metodología establecida por esta autora, nuestra metodología de compilación de corpus quedó dividida en tres fases: a) fase documental, b) fase de compilación y c) fase de análisis.

En la primera etapa (documentación) los alumnos pusieron en prácticas conocimientos adquiridos en clases previas teóricas (estrategias de búsqueda en la web (operadores básicos, expresiones booleanas y operadores filtro), además de conocimientos teóricos sobre lingüística de corpus y traducción, tipología de corpus, uso de corpus de referencia (por ejemplo, CREA, Corpus de Referencia del Español Actual de la Real Academia de la Lengua Española). El uso de estrategias de búsqueda fue de especial relevancia, ya que una búsqueda precisa posibilitaría a nuestros estudiantes encontrar fuentes de información especializada para la creación de su corpus. Los alumnos también se formaron en el uso de programas de concordancias como AntConc y se familiarizaron con sus principales funciones (generador de listados de palabras frecuentes, uso de concordancias, recuperación de colocaciones y paquetes léxicos). Una vez que los discentes contaban con un andamiaje sólido, según las necesidades de nuestra actividad, se les encargó compilar un corpus *ad hoc* monolingüe (inglés británico) que les sirviera como herramienta documental de un texto especializado sobre la anemia drepanocítica. También se les proporcionó un corpus de referencia que utilizarían en la última fase (análisis).

Tras la lectura de textos paralelos y la familiarización con el tema, los estudiantes procedieron con la búsqueda de textos electrónicos a través de palabras clave

como “sickle-cell anemia”, “hemoglobin”, “erythrocytes”, “Hb SS disease” y la combinación de las mismas en los motores de búsqueda (Google). Se usaron búsquedas avanzadas (operadores filtro) para restringir los parámetros de dominio o sitio tales como **sickle cell site: www.nhlbi.nih.gov**, donde el buscador recupera una selección de textos especializados dentro del sitio web señalado que contenga la palabra “sickle-cell”. Por el contrario, **sickle cell site: co.uk** ofrece páginas que incorporan la palabra “sickle cell”, pero que se encuentran bajo el dominio co.uk, lo que delimita diatópicamente nuestras búsquedas. Si restringimos la búsqueda por etiqueta la etiqueta <title> (**intitle:sickle cell anemia**) el buscador devuelve sitios web que incluyen el termino especificado en su etiqueta <title>, lo cual es un indicio de especificidad temática:

The image shows a screenshot of the Mayo Clinic website. At the top, there is a navigation bar with the Mayo Clinic logo, a menu icon, a search bar containing 'Search Mayo Clinic', and social media icons for Twitter, Facebook, YouTube, and Pinterest. Below the navigation bar, the main content area is titled 'Sickle cell anemia' under the heading 'Patient Care & Health Information > Diseases & Conditions'. There are tabs for 'Overview', 'Symptoms & causes', 'Diagnosis & treatment', and 'Self-management'. The 'Overview' tab is selected. The text under 'Overview' states: 'Sickle cell anemia is an inherited form of anemia — a condition in which there aren't enough healthy red blood cells to carry adequate oxygen throughout your body.' It further explains that normally red blood cells are flexible and round, but in sickle cell anemia, they become rigid and sticky, shaped like sickles or crescent moons, and can get stuck in small blood vessels. To the right of the text is an illustration of red blood cells, with some labeled 'Normal' (round) and others 'Sickle cell' (crescent-shaped). Below the illustration is the caption 'Sickle cell anemia'.

Figura 1. Ejemplo de búsqueda por etiqueta <title>

El segundo estadio en el protocolo lo compone la compilación en sí, que incluye la descarga, almacenamiento y conversión a texto plano (.txt) de los documentos, al ser este el formato que podría considerarse estándar para utilizar en los programas de gestión de corpus. La descarga puede realizarse de forma manual, con la descarga de cada página web, o bien se pueden utilizar gestores de descarga de sitios web (WinHTTrack Website Copier). En cuanto al almacenamiento y conversión, existen diversos programas para nombrar archivos en lote (Personal Renamer, FastFile Rename y Lupass Rename) y programas para la conversión de formatos (HTMLas-Text y PdfToText). Las extensiones que normalmente se van a encontrar en Internet

son.html y.pdf, por lo que se hacen necesarias las herramientas mencionados, o de lo contrario la labor de descarga podría convertirse en un proceso difícil que requeriría de una importante inversión de tiempo.

Tras las fases de documentación y compilación, donde el objetivo era que el alumnado desarrollara sus destrezas técnicas, como tratamiento de formatos, así como familiarizarse con diversos programas informáticos, se entra en la fase de análisis. Antconc 3.4.3, programa informático desarrollado por Laurence Anthony³, de coste cero, fácil de usar y con una interfaz sencilla, fue la herramienta que los alumnos utilizaron. Antconc incluye distintas funciones: generar concordancias, listados de frecuencias de las palabras del corpus, patrones colocacionales y paquetes léxicos, además de poder incluir un corpus de referencia para así extraer vocabulario específico basado en frecuencias y poder elaborar, por ejemplo, un glosario. Los alumnos comenzaron a utilizar su corpus virtual como herramienta terminológica para la creación de un glosario en inglés. Los discentes generaron una lista de palabras de términos monoléxicos ordenados por orden alfabético y frecuencia de aparición. A su vez, crearon lo que se llama “stoplist” para filtrar palabras como artículos, adverbios y preposiciones. De este modo, se evitaba que palabras irrelevantes o el llamado “ruido” aparecieran como palabras más frecuentes en el corpus. Este tipo de listas, que pueden subirse al programa AntConc, ofrece la posibilidad de crearlas manualmente o bien subir alguna de las listas que pueden encontrarse por Internet. Antconc también facilita el uso de corpus de referencia (función *Keyword list*), lo que permitiría al alumno la tarea de una extracción terminológica mucho más sencilla y precisa. Para ello se proporcionó un listado extraído del corpus de referencia Bank of English (BoE). Tras el uso del corpus de referencia, los alumnos pudieron comprobar la especificidad de su corpus y pudieron extraer vocabulario específico basado en la frecuencia para crear su glosario monolingüe.

Los distintos términos también pueden visualizarse en su contexto con la función de concordancias (*concordance*). Así, a la hora de utilizar el corpus como fuente de información conceptual y sinonímica, el uso de comodines se presentó realmente útil para manejar la dicha función. Por ejemplo, si en la casilla de búsqueda se utiliza *disease**, el programa arrojará resultados de la palabra en singular y plural (Figura 3). Al pinchar en cada uno de los resultados, el programa presenta el texto completo en el que se encuentra el término (información contextual). A la hora de buscar definiciones de términos desconocidos, los discentes utilizaron patrones combinatorios para conocer el significado de algunos términos: *sickle* is*, *sickle-cell anemia*or*, *hemoglobin * is*; *hemoglobin * or*, *erythrocytes * are known as* y así elaborar fichas documentales (Figura 4).

³ El programa informático puede descargarse en <http://www.laurenceanthony.net/software/antconc/>

Frecuencia de aparición	Término LO	Término LM
618	Sickle	Falciforme
580	Cell	Célula
425	Blood	Sangre
366	Hemoglobin	Hemoglobina
240	Disease	Enfermedad
219	Red [blood cell]	Glóbulo rojo
209	Hb (hemoglobin)	Hb (hemoglobina)
208	Haemolysis	Hemólisis
191	Iron	Hierro
181	Patients	Pacientes
179	Haemoglobin	Hemoglobina
163	Anemia	Anemia
108	Sickling	Falciforme
102	Oxygen	Oxígeno
101	Hb S (hemoglobina S)	HbS (hemoglobina S)
100	Labile	Lábil
99	RBCS (Red Blood Cells)	Glóbulos rojos (RBC)
96	Gene	Gen

Figura 2. Ejemplo de glosario elaborado por los alumnos

Los distintos términos también pueden visualizarse en su contexto con la función de concordancias (*concordance*). Así, a la hora de utilizar el corpus como fuente de información conceptual y sinonímica, el uso de comodines se presentó realmente útil para manejar la dicha función. Por ejemplo, si en la casilla de búsqueda se utiliza *disease**, el programa arrojará resultados de la palabra en singular y plural (Figura 3). Al pinchar en cada uno de los resultados, el programa presenta el texto completo en el que se encuentra el término (información contextual). A la hora de buscar definiciones de términos desconocidos, los discentes utilizaron patrones combinatorios para conocer el significado de algunos términos: *sickle* is*, *sickle-cell anemia*or*, *hemoglobin * is*; *hemoglobin * or*, *erythrocytes * are known as* y así elaborar fichas documentales (Figura 4).

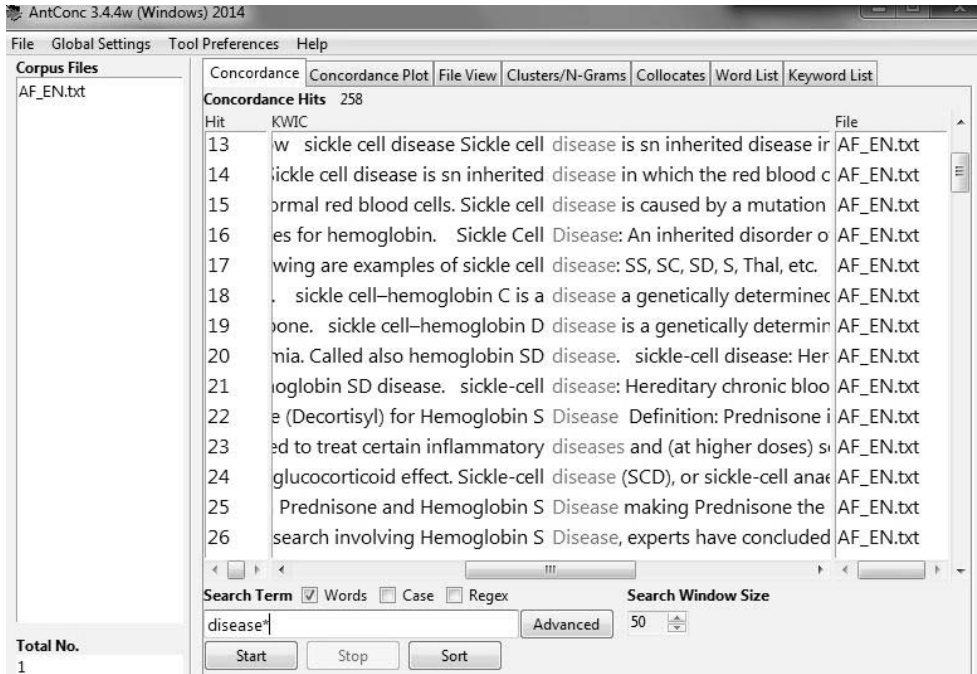


Figura 3. Ejemplo función concordance

<i>Sickle-cell anemia (SCA)</i>	Definición	Sickle cell anemia is an inherited blood disorder, identified by the sickle shape of red blood cells which carry less oxygen and break easily, causing anemia. The sickling trait, the less serious form, occurs from the inheritance of only one parent; however, both parents must exhibit the disease in order for full symptoms to take place. It is caused by an error in the gene that tells the body how to make hemoglobin. The disorder is usually found in people of African descent because the gene mutation that occurred thousands of years ago in parts of Africa.
	Sinónimos	Sickle-cell disease (SCD)
		Hemoglobine SS-disease
		Drepanocytosis

Figura 4. Ejemplo de ficha creada por los alumnos

Otra función con la que los alumnos trabajaron fue la llamada *clusters/N-grams* para buscar conjuntos léxicos formados por más de una palabra y ver su comportamiento con otros términos. “Hemoglobin” fue la segunda palabra más frecuente tras comparar el corpus con un corpus de referencia (366 resultados). Al buscarla con la función *clusters/N-grams*, se mostraron los siguientes resultados de candidatos a términos (el tamaño del conjunto léxico se limitó a dos gramas):

47	hemoglobin ss
41	hemoglobin aa
32	hemoglobin s
18	hemoglobin a
17	hemoglobin ss erythrocytes
16	hemoglobin aa erythrocytes
10	hemoglobin and
10	hemoglobin electrophoresis
10	hemoglobin is
9	hemoglobin as
9	hemoglobin sc
8	hemoglobin c
8	hemoglobin molecule
8	hemoglobin s and
7	hemoglobin iron

Tabla 1

Gracias a esta función, el glosario incluyó tanto unidades monoléxicas como poliléxicas, tomando la frecuencia como criterio para la selección de candidatos a término y comprobar su utilidad a la hora de reconocer distintos tipos de hemoglobina o comprobar sinónimos para dichos tipos.

5. Conclusión

Nuestro trabajo ha descrito y ejemplificado el uso de la metodología de corpus en el contexto de la formación de traductores e intérpretes en TISP. El tiempo de formación fue reducido, por lo que el diseño y secuenciación de las actividades formativas fue esencial para que el proceso y el resultado fuese el adecuado. El enfoque adoptado pretendía, de un lado, proporcionar el andamiaje necesario para que los discentes contaran con el uso de corpus como herramientas documentales adicionales a las tradicionales en tareas de índole terminológica y fraseológica y, de otro, pudieran trasladar lo aprendido (compilación de corpus, tratamiento de formatos, análisis del corpus) a otros contextos dentro del aula de traducción y, por qué no, en su futura tarea profesional.

Como hemos señalado a lo largo del trabajo, el uso de una metodología de corpus requiere un proceso formativo claro. Nuestra propuesta se articuló en dos fases: a) contextualización, donde se familiariza al alumnado con los aspectos teóricos fundamentas en Lingüística de Corpus y Estudios de Traducción basados en Corpus;

y b) proceso de compilación, donde el alumno es parte central y entra de lleno en el diseño del corpus (en nuestro caso, un corpus *ad hoc*).

En suma, en nuestra opinión, la formación de traductores e intérpretes en TISP debe incorporar cada vez más recursos documentales que integren el desarrollo de distintas competencias y la formación tecnológica. La metodología de corpus en el aula de TISP se presenta con un importante valor didáctico que necesita (y debe) ser explorado.

Referencias bibliográficas

- Angelelli, C., «Designing curriculum for healthcare interpreter education: A principles approach», en: Roy, C., (ed.), *New approaches to interpreter education*. Washington D.C.: Gallaudet UP 2006, 23-46.
- aston, g., «Corpus use and learning to translate». *Textus* 12 (1999), 289-314.
- Austhermül, F., *Electronic tools for translators*. Manchester St Jerome 2001.
- Bernardini, S. y Castagnoli, S., «Corpora for Translation Education and Translation Practice», en Yuste Trigo, E., (ed.), *Topics in Language Resources for Translation and Localization*, Amsterdam-Filadelfia: John Benjamins 2008, 39-55.
- Bowker, J., y Pearson, J., *Working with specialized language: A practical guide to using corpora*. London: Routledge 2002.
- Bowker, L., «Corpus-based applications for translator training: Exploring possibilities», en: Granger, S., Lerot, S. y Petch-Tyson, J. (eds.), *Corpus-based approaches to Contrastive Linguistics and Translation Studies*. Amsterdam/Nueva York: Rodopi 2003, 169-183.
- calzada perez, c. *El espejo traductológico. Teorías didácticas para la formación del traductor*. Barcelona: Ediciones Octaedro 2007.
- Cid-Leal, P. y Perpinyá-Morera, R., «Competencia informacional en Traducción : análisis de los hábitos de los estudiantes», *BID, Textos universitarios de bibliotecnología y documentación* 34 (2015).
- Corpas Pastor, G., *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt: Peter Lang 2008.
- D'Hayer, D., «Public service Interpreting an translation: Moving towards a (virtual) community of practice», *Meta* 57 (2012), 235-247.
- Désilets, A., Melacon, Ch. ; Patenaude, G. ; Brunette, L., «How translators use tools and resources to resolve translation problems: an ethnography study». Actas del Congreso Beyond Translation Memories. NRC Publications Archive (NPARC) 2009.
- EMT Expert Group., «Competences for professional translators, experts in multilingual and multimedia communication», Recuperado el 28 de marzo, 2016 [http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf]
- Fantinuoli, C. y Zanettin, F., (eds). *New directions in corpus-based translation studies*. Berlin: Language Science Press 2015.
- Gallego Hernández, d., «The use of corpora as translation resources. A study based on a survey of Spanish professional translators». *Perspectives: Studies in Translationology* 23 (3) (2015), 375-371.

- Granger, s., Lerot, J., Petch-Tyson, S., *Corpus-based approaches to contrastive linguistics and translation studies*. Ámsterdan y Nueva York: Rodopi 2003.
- Laviosa, S., *Corpus-based Translation Studies. Theory, Findings, Applications*. Ámsterdan / Nueva York: Rodopi 2002.
- Mikkelson, H., «Evolution of public service interpreting training in the U.S. », *Fitispas International Journal* 1 (2014), 9-22.
- Mikkelson, H., «Community interpreting: An emerging profession», *Intpreting* 1, 1 (1996), 125-129.
- Montalt Resurrecció, V. y González Davies, M., *Medical translation step by step. Learning by drafting*. Manchester: St. Jerome 2007.
- Monzó Nebot, E., «Corpus-based Activities in Legal Translator Training». *The Interpreter and Translator Trainer* 2 (2) (2008), 221-252.
- Neunzig, W., «Tecnologías de la información y traducción especializada inversa», en: Kelly, D. et al (eds.), *La direccionalidad en Traducción e Interpretación. Perspectivas teóricas, profesionales y didácticas*. Granada: Atrio 2003, 189-205.
- Orozco Jutorán, M., *Metodología de la traducción directa del inglés al español. Materiales didácticos para traducción general y especializada*. Granada: Comares 2012.
- Palomares Perrault, P., *Recursos documentales para el estudio de traducción*. Málaga: Universidad de Málaga 2000.
- Pinto Molina, M. y Sales Salvador, D., «INFOLITRANS: A model for the development of information competence for translators», *Journal of documentation* 64, 3 (2008), 413-437.
- Rodríguez Inés, P., «Electronic corpora and other information and communication technology tools. An integrated approach to translation teaching», *The Interpreter and Translator Trainer*, 4, 2 (2010), 251-282.
- Sánchez Ramos, M.M y Vigier Moreno, F., “Using Monolingual Virtual Corpora in Public Service Legal Translator Training”. En E. Martín-Monje, I. Elorza y B. García Rianza (Eds.), *Technological Advances in Specialized Linguistic Domains: Practical Applications and mobility*. Londres/Nueva York: Routledg 2016, 228-239.
- Seghiri, M., «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 (2011), 13-30.
- Seghiri, M., «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores», *Babel* 63 (1).
- Tognini-Bonelli, E., *Corpus linguistics at work*. Ámsterdan/Filadelfia: John Benjamins Publishing Company 2001.
- Valero Gracés, C. y De la Cruz Cabanillas, I. (Eds.). *Traducción y nuevas tecnologías*. Alcalá de Henares: Servicio de Pubzlicaciones de la Universidad de Alcalá 2001.
- Valero Garcés, C., «La difícil tarea de documentarse en traducción e interpretación en los servicios públicos», en Sales Salvador, D. (ed.), *La biblioteca de Babel. Documentarse para traducir*. Granada: Comares 2005, 99-121.
- Valero Garcés, C., «Formación de traductores e intérpretes en una sociedad multicultural. El programa de la Universidad de Alcalá, Madrid», *Cuadernos de Aldeu* 25 (2013), 215-238.

- Vargas Urpí, M. «La difícil tarea de dar respuesta a las necesidades de formación en Interpretación en los Servicios Públicos (ISP) en Cataluña: 10 años de avances y retrocesos», *Fitispos International Journal* 3 (2016), 9-22.
- Zanettin, F., Bernardini, S. y Dominic, S., (Eds.), *Corpora in Translator Education*. Manchester: St. Jerome 2003.