

TACLAC: una plataforma colaborativa de intercambio de corpus y su aplicación en la traducción biosanitaria¹

M.ª del Mar Sánchez Ramos y Colleen McCarroll*

Resumen: El objetivo principal de este artículo es ofrecer y ejemplificar el uso de la plataforma colaborativa TACLAC en la fase de documentación previa al proceso de traducción dentro del ámbito biosanitario de la traducción e interpretación en los servicios públicos (TISP). Dicha plataforma está dirigida a profesionales, alumnos y docentes de TISP, y supone la creación e intercambio de distintos corpus con el fin de que puedan servir de herramientas documentales en las distintas fases de traducción especializada en forma de consulta terminológica, conceptual o fraseológica, o para la creación de otros recursos, como las memorias de traducción.

Palabras clave: documentación, herramientas de gestión de corpus, lingüística de corpus, traducción biosanitaria, traducción colaborativa, traducción e interpretación en los servicios públicos.

TACLAC: A collaborative platform for exchanging corpus in healthcare translation

Abstract: This paper presents a collaborative platform called TACLAC and exemplifies its use in the documentation stage prior to the translation process in the field of healthcare in Public Service Translation and Interpreting (PSTI). The platform is aimed at PSTI professionals, students and trainers. It is conceived as a platform to exchange different corpora (monolingual, parallel and comparable) to serve as documentation resources to solve doubts on terminology, concepts or phraseology during the translation process or to create other resources such as translation memories.

Keywords: collaborative translation, corpus linguistics, corpus management tools, documentation, medical translation, public service interpreting and translation.

Panacea 2019; XX (49): 73-81

Recibido: 24.II.2019. Aceptado: 06.V.2019.

1. Introducción

La disciplina conocida como TISP es un área de investigación que ha experimentado un gran avance y consolidación en las últimas décadas. La demanda de profesionales en situaciones que dificultan el proceso de comunicación entre los representantes de las administraciones públicas y los usuarios que no dominan una lengua es cada vez mayor, y, de este hecho, han tomado conciencia las distintas instituciones académicas. En este sentido, son cada vez más los programas universitarios cuya finalidad se centra en la formación de estos profesionales (Valero Garcés, 2013). Dentro de los diversos programas, independientemente del área de especialización, la preparación documental es de vital importancia (Palomares, 2000), y aún más en las tareas de traducción especializada o de la TISP, debido, en gran parte, a la variedad de situaciones concretas en donde el uso de una terminología precisa y adecuada es clave (p. ej., consultas médicas con inmigrantes). Es lógico pensar, por tanto, que el desarrollo de técnicas documentales debe estar presente en los diversos planes de estudios de las propuestas formativas en la TISP.

Lejos queda ya la exclusividad de los recursos en soporte papel a los que puede acceder todo traductor e intérprete, que han sido sustituidos por cantidades ingentes de información en formato digital y de fácil y rápido acceso. Dentro de los recursos documentales electrónicos (p. ej., bases de datos terminológicas, diccionarios electrónicos especializados), el uso de corpus ha tomado un papel central (Fernández Rodríguez, 2015; Gallego Hernández, 2015; Sánchez Ramos, 2017). Saber emplear los corpus en las distintas etapas del proceso traductor puede aportar varias ventajas (p. ej., como recurso para la elaboración de memorias de traducción, como acceso a textos reales donde consultar problemas de índole fraseológica o como material de referencia para la preparación de glosarios) para el profesional de la traducción y, en particular, para los profesionales de la traducción en los servicios públicos², ya que, con mucha frecuencia, las tareas dentro de este campo serán altamente especializadas. Sin embargo, el uso de una metodología de corpus requiere una preparación previa si queremos sacar el máximo partido a los corpus como recursos documentales, ya sea tanto para la consulta de los mismos como para la propia creación de corpus.

Es bajo este marco de necesidad documental específica dentro de la TISP de instruir en el uso de corpus donde se sitúa el presente trabajo, y se propone la creación de una plataforma de intercambio de corpus de diversa índole (TACLAC) como

* Universidad de Alcalá (Madrid, España). Dirección para correspondencia: mar.sanchezr@uah.es.

recurso para los procesos documentales en tareas relacionadas con la traducción biosanitaria, en concreto. Las siguientes secciones detallan cómo se ha integrado e implementado el uso de dicha plataforma como parte de la metodología de corpus realizada en el seno de la asignatura obligatoria Técnicas y Recursos para la TISP del Máster en Comunicación Intercultural, Interpretación y Traducción en los Servicios Públicos, que se ha impartido en la Universidad de Alcalá (Madrid) durante el curso académico 2018-2019.

2. Documentación y traducción biosanitaria en la TISP

La llegada de Internet y el uso generalizado de la tecnología en diversos ámbitos ha cambiado radicalmente la manera en la que se lleva a cabo la labor documental, y es que esta revolución tecnológica ha tenido una gran influencia en la forma que tiene un traductor de documentarse. De hecho, Corpas Pastor (2012: 81) comenta la estrecha relación entre los procesos documentales y la competencia tecnológica, por lo que se refiere a fuentes documentales electrónicas. Encuentra que la relación es tan cercana que se podría hablar de una «competencia tecnológica informacional para referirnos a la capacidad de seleccionar y manejar de forma adecuada las tecnologías de la traducción y la documentación electrónica necesarias para la preparación, producción y evaluación profesional de las producciones de traductores e intérpretes». No obstante, no podemos olvidar las palabras de Pinto Molina (2005: 19):

La información es un recurso en alza y en proceso de superabundancia incontrolada. Su crecimiento exponencial hace que la mitad de la información de la que hoy se dispone se haya acumulado en los últimos cincuenta años. Pero no solo es la cantidad y acopio elevado de datos lo que hace evolucionar constantemente a nuestra sociedad: el uso racional, ordenado, productivo e inteligente de la información se configura como el factor primordial de desarrollo. Las organizaciones, las empresas, los grupos sociales y, en definitiva, los individuos tienen la capacidad de transformar todo ese volumen de información en conocimiento inteligente y en saber productivo.

Queda bien claro que el éxito del proceso documental irá de la mano de una buena gestión de la información, es decir, el buen uso que de ella se haga, ya que, según Alcina, Soler y Estellés (2005: 222), «la gran cantidad y variedad de datos disponible a través de Internet puede convertirse [...] en una trampa para el traductor».

En relación con este vínculo entre la documentación y la competencia tecnológica y el saber utilizar la información que buscamos, Alcina, Soler y Estellés (2005: 222) subrayan dos cambios fundamentales que han surgido en el proceso de documentación tras el auge de Internet: la manera en la que se busca información y también el tipo de información que se puede encontrar:

La búsqueda de información ha cambiado porque *a)* el traductor puede hacer las búsquedas sin salir de su despacho [...] y *b)* la información a la que puede acceder a través de Internet es notablemente superior en cantidad y variedad. Por otra parte, también ha cambiado respecto a qué tipo de información puede buscar porque ahora, junto a la búsqueda de contenido temático y equivalencias terminológicas, el traductor puede obtener textos especializados de diferentes tipos textuales y niveles de especialización.

Existen trabajos que vienen a confirmar la consolidación de las fuentes electrónicas como recursos documentales en las labores de traducción. Tal es el caso de Durán Muñoz (2010), quien, en un estudio realizado entre profesionales de la traducción, afirma que los traductores prefieren solventar sus dudas, en este caso terminológicas, a través de la consulta de fuentes terminológicas en línea, debido en gran parte al fácil y rápido acceso de las mismas. Otro ejemplo lo encontramos en el trabajo realizado por Cid-Leal y Perpinya-Morera (2015) sobre el uso de herramientas documentales electrónicas de estudiantes de traducción, donde viene a confirmarse lo ya señalado por Durán Muñoz (2010) además de, y creemos que es lo realmente interesante desde nuestro punto de vista, 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. No podemos estar más de acuerdo con las conclusiones de estas autoras y nos atrevemos a afirmar que el alumnado de Traducción necesita desarrollar estrategias de búsqueda en Internet y conocer el funcionamiento de programas informáticos variados de descarga de información en la red para, con ello, ser capaces de crear su propio material documental, como pueden ser los llamados corpus.

La competencia documental (Pinto Molina y Sales Salvador, 2008), también denominada competencia heurístico-documental, competencia profesional e instrumental (Rabadán y Fernández, 2002) o infocompetencia (Sales Salvador, 2006), engloba cuatro competencias fundamentales: 1) cognitiva, relacionada con el conocimiento necesario para realizar la práctica traductora; 2) informativo-digital, relacionada con el dominio de plataformas y diversas herramientas informáticas; 3) comunicativa, relacionada con la interacción entre los distintos actores implicados en el proceso traductor; y 4) procedimental, relacionado con la aplicación de los conocimientos adquiridos. Castillo Rodríguez y Santamaría Urbieta (en prensa) inciden en la documentación como el estadio necesario anterior a la traducción «el cual debe ser igual de importante que la propia fase de trasvase y entrega del producto final». En relación con la TISP, esta competencia adquiere aún más importancia por la especificidad de la misma y la necesidad de encontrar los recursos adecuados. Valero Garcés (2005: 109) explica que el profesional de la TISP «se enfrenta a textos que aparentemente utilizan el lenguaje coloquial (escrito o hablado) pero que también contienen lenguajes de especialidad (médica, jurídica, farmacológica, administrativa, tema laboral, etc.)». Tratar estos lenguajes de especialidad y resolver los problemas que surgen exige del

profesional una documentación adecuada. Como bien señala Sánchez Ramos (2017: 179):

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 aún 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 re-expresión del 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.

El profesional de la TISP deberá recurrir a fuentes documentales para conseguir información sobre el texto original, para resolver problemas terminológicos que surjan por el uso de lenguajes especializados y para encontrar información fraseológica. De las distintas fuentes documentales, y a raíz de la especialización temática que la traducción biosanitaria presenta, creemos que los corpus juegan un papel fundamental a la hora de adquirir un conocimiento preciso terminológico y fraseológico en los contextos especializados (Ortego-Antón y Fernández Nistal, en prensa; Toledo Báez, en prensa) y que pueden «ayudar al estudiante de traducción a realizar un acercamiento al ámbito especializado y a aprender a gestionar y crear recursos documentales y terminológicos propios» (Rodríguez Martínez y Ortega Arjonilla, 2018: 120).

A día de hoy, los corpus ocupan un lugar destacado en el trabajo de documentación que debe realizar cualquier traductor antes de un encargo (Seghiri, 2017). Un corpus se define como «una colección de textos en soporte digital, que ha sido compilada conforme a unos criterios de diseño apropiados que le confieren representatividad» (EAGLES, 1996). Nosotros nos acercamos a los corpus denominados *corpus virtuales*, también llamados *ad hoc* (Aston, 1999), y que Seghiri (2011: 16) describe de la siguiente forma:

Aquel que se crea exclusivamente a partir de recursos electrónicos disponibles en la red Internet para la realización puntual de un determinado encargo de traducción o documentación de un bloque textual, y cuyo principal objetivo es satisfacer las principales necesidades documentales del traductor de forma económica (solo requiere de la conexión a la red Internet), rápida (los textos pueden descargarse en cuestión de segundos con un clic de ratón) y fiable (pues permite observar, entre otros, la macro- y microestructura textual, así como la terminología propia del campo de especialidad en vivo).

Comparados con otro tipo de corpus de amplias dimensiones (por ejemplo, el British National Corpus o el Brown Corpus), los corpus virtuales no destacan por su tamaño. En estos casos, no estamos hablando de millones palabras, sino de un tipo de corpus más reducido, con una especificidad temática y

extremadamente precisos, y en donde prima la calidad frente a la cantidad. El uso de este corpus responde a unas necesidades documentales concretas o bien puede emplearse como recursos pedagógicos del discente en la elaboración de material didáctico (Zanettin, 2012: 64). Los corpus virtuales pueden servir como fuentes documentales durante el proceso de traducción y ser de utilidad para «aprender términos especializados, identificar palabras que aparecen juntas, asimilar usos gramaticales, descifrar cuestiones estilísticas y genéricas, así como para profundizar en la temática conceptual de los textos a traducir» (Calzada Pérez, 2005: 170). No obstante, no puede obviarse la inversión de tiempo que ello supone. Es por ello que una instrucción precisa en la creación y compilación de corpus no puede faltar en la formación de traductores.

Por todo lo expuesto, el Máster en Comunicación Intercultural, Interpretación y Traducción en los Servicios Públicos de la Universidad de Alcalá (Madrid) oferta una asignatura obligatoria, llamada Técnicas y Recursos para la TISP, que incluye, junto con formación en traducción asistida por ordenador (TAO) y terminología, contenidos de metodología de corpus. Los principales objetivos de dicho módulo son: a) buscar, seleccionar y evaluar fuentes de documentación para el ámbito sanitario en TISP; b) saber dominar recursos de consulta lingüística y terminológica en Internet; c) saber dominar herramientas de consulta (corpus, programas de concordancias) como recursos documentales. Se ha optado por incluir este tipo de herramientas informáticas por varias razones. En primer lugar, suelen ser herramientas (casi) desconocidas por el alumnado, o bien, en caso de conocerlas, los alumnos afirman desconocer sus aplicaciones prácticas (Sánchez Ramos y Lázaro Gutiérrez, en prensa). A su vez, son herramientas que, además de solventar problemas de corte terminológico, fraseológico o conceptual, desarrollan también las destrezas más técnicas al trabajar con nuevos formatos y programas específicos. Y, por último, son herramientas que pueden ayudarles tanto en su plano profesional (por ejemplo, uso de corpus paralelos para la creación de memorias de traducción) como investigador.

3. La plataforma TACLAC

La propuesta para la plataforma TACLAC se sustenta sobre cuatro pilares que responden a las necesidades actuales a las que se enfrenta el alumnado de la TISP y cualquier profesional interesado en el uso de corpus en tareas de traducción. Se trata de una plataforma en línea donde se establecen los siguientes objetivos: (I) que los usuarios puedan formarse sobre qué son los corpus y qué utilidad tienen durante el proceso de documentación previa a los encargos de la TISP, (II) que los usuarios se familiaricen con distintas herramientas de gestión de corpus, (III) que los usuarios compartan con el resto de la comunidad los corpus que hayan creado y (IV) que los usuarios puedan acceder a los corpus ya creados.

La plataforma TACLAC se propone como una pequeña pieza del rompecabezas en la formación de los traductores en la TISP y, por ende, en la traducción especializada. Los primeros

FIGURA 1. Interfaz de la plataforma TACLAC



dos propósitos responden a la necesidad de formación que se ha repasado en los párrafos anteriores. Como se ha visto con anterioridad, la formación de los traductores es fundamental para el desempeño correcto de su trabajo. Entre las muchas competencias que debe desarrollar un profesional de la TISP, TACLAC pretende desarrollar la competencia documental. Se ofrece como un recurso adicional, un espacio colaborativo que incluya corpus especializados que puedan ayudar a docentes (p. ej., elaboración de material docente, creación de memorias de traducción), alumnos y profesionales (p. ej., uso de los corpus como recursos documentales de ayuda en futuras tareas de traducción). Como tal, la plataforma TACLAC puede servir tanto en la formación inicial de los profesionales de la TISP como en la formación continua o como recursos en tareas de traducción especializada. El tercer y cuarto objetivo de TACLAC responde a la dificultad que, en ocasiones, experimentan los profesionales de la TISP en la búsqueda de recursos ante trabajos especializados. Comentaba Seghiri (2011: 17) que «ante esta situación, al traductor no le queda otra alternativa que compilar él mismo su propio corpus virtual». Eso es cierto, pero la propuesta TACLAC va más allá y persigue ofrecer otro paso a seguir tras la elaboración de un corpus: compartirlo, para que otro traductor de la TISP pueda sacar el máximo provecho de ello. De igual forma, y como propuesta de innovación docente, la plataforma puede utilizarse para fines docentes entre aquellos que quieran incorporar una metodología de corpus en las clases de traducción especializada, o bien utilizar la plataforma como recursos para la preparación de material docente.

La plataforma TACLAC sigue la propuesta de Alcina, Soler y Estellés (2005: 235) sobre la creación de plataformas en Internet donde se puede presentar y obtener información especializada y fiable para los traductores. Las autoras observan que «la diversificación constante de disciplinas y especialidades y el aumento considerable de conocimiento que se produce en cada una de ellas hacen que la búsqueda, selección y almacenamiento de textos especializados sea una labor cada vez más ardua para el traductor especializado». Este tipo de plataforma que sugieren, ejemplificado por TACLAC, facilitaría esta difícil labor para el traductor. En ningún caso la plataforma actual pretende ser la versión final de la plataforma, ya que la creación de esta requie-

riría de apoyo de profesionales de la informática para su diseño y funcionamiento mucho más sofisticado, y, así, se ha diseñado un blog para servir como modelo de lo que podría ser una versión final de la plataforma.

TACLAC se ha creado con la ayuda de los conocidos gestores de contenido (*content management systems*, CMS). De entre ellos, se ha seleccionado Wordpress, uno de los gestores de contenidos más populares y de fácil manejo, y se ha proporcionado la siguiente dirección para su acceso: <http://www.taclac.wordpress.com>. Ya que el objetivo de la plataforma es crear una comunidad internacional, se ha redactado en inglés. El título del blog es *Take a Corpus, Leave a Corpus* y cuenta con los siguientes apartados: *About*, *What's a Corpus?*, *Virtual Corpora*, *Corpus Tools*, *How to Use TACLAC*, *Resources*, *Blog Posts* y *Contact* (figura 1).

3.1. Corpus virtuales disponibles y ejemplo de uso

TACLAC cuenta con dos corpus de muestra de temática bio-sanitaria³. Para cada uno de ellos, se ha seguido un protocolo de compilación, basado en Seghiri (2011) y Sánchez Ramos y Vigier Moreno (2016), que consta de tres fases: 1) fase de documentación, 2) fase de compilación y 3) fase de análisis, y que se ejemplifican en las siguientes secciones.

El primero de ellos, el corpus comparable monolingüe (EN) sobre cuidados paliativos (corpus 1), es el corpus de muestra que ofrece la plataforma, y se utilizó en la asignatura del máster anteriormente mencionado (curso 2018-2019) para ilustrar, de un lado, el uso y la finalidad de la plataforma, y, de otro, para la tercera fase de análisis. Los alumnos procedieron a la descarga del corpus y trabajaron con dicho corpus como herramienta de documentación previa a futuras tareas de traducción. De este modo, el principal objetivo de esta tarea con el corpus 1 era que dicho corpus sirviera para documentarse sobre un tema específico, tarea previa a cualquier proceso de traducción. El segundo de ellos, el corpus comparable bilingüe (EN-ES) sobre cardiopatía isquémica (corpus 2), es una muestra de la finalidad en sí de la plataforma, y responde al objetivo 3 anteriormente mencionado, puesto que es un corpus que los alumnos de posgrado crearon y que, tras la evaluación del mismo por parte del profesorado, se incluyó en TACLAC. En cada uno de ellos, y junto

con las fases anteriormente expuestas, se siguieron muy de cerca los parámetros de evaluación y compilación propuestos por Bowker y Pearson (2002) (tabla 1 y tabla 2).

La puesta en marcha de las actividades con TACLAC tuvieron lugar, como se ha mencionado al comienzo de este trabajo, en el seno de la asignatura Técnicas y Recursos para la TISP dentro del Máster en Comunicación Intercultural, Interpretación y Traducción en los Servicios Públicos (Universidad de Alcalá, Madrid). La actividad ocupó un total de 4 horas, repartida en dos sesiones de dos horas de duración, y contó con la participación de un total de 30 alumnos.

3.2. El corpus comparable monolingüe (EN) sobre cuidados paliativos (corpus 1)

El corpus 1 es un corpus integrado por textos escritos en inglés, de carácter científico y académico, extraídos de revistas especializadas y accesibles desde Internet (*Journal of Pain and Symptom Management*, *Palliative Medicine*, *The American Journal of Psychiatry* y *BMC Palliative Care*, entre otras), publicados en el periodo 2010-2015 (tabla 1). Como parte de la fase de documentación, se utilizó una búsqueda a través de sitios institucionales (Authermül, 2006), concretamente, revistas especializadas. Posteriormente, se pasó a la fase de compilación, que incluía la descarga y almacenamiento del material encontrado, junto con la conversión de los textos a texto plano (.txt). En su mayoría eran archivos .pdf, por lo que se utilizó el programa pdf2text para tal fin. Una vez que estaban los textos en forma de texto plano, se subieron a TACLAC.

TABLA 1. Criterios de creación del corpus 1 (Bowker y Pearson, 2002)

Criterios de compilación del corpus	Características del corpus
Número de palabras (<i>tokens</i> y <i>types</i> ²)	inglés → <i>tokens</i> ⁴ (122 079), <i>types</i> (7406)
Número de textos	26 textos
Medio	textos escritos
Temática	especializada: cuidados paliativos
Tipos de texto	artículos científicos extraídos de revistas especializadas
Idiomas	inglés británico, inglés americano
Periodo de tiempo	2010-2015

Se trata, por tanto, de un corpus virtual con un grado de especialidad alto, por lo que puede ser útil para consultas concretas o para la elaboración de glosarios previos a la traducción

de un texto especializado. Este primer corpus de muestra se ha utilizado en la clase de Técnicas y Recursos para la TISP para ejemplificar el uso de la plataforma a la vez que para instruir al alumnado en la fase de análisis y prepararlo para la creación de su propio corpus, que entrará a formar parte de la plataforma TACLAC.

Esta primera sesión de 2 horas se dedicó a describir los principales conceptos relacionados con los estudios de traducción basados en corpus, la tipología de corpus, estrategias de búsqueda en la web (uso de operadores filtro) y uso de distintos programas de gestión de corpus (Sketch Engine y AntConc). Posteriormente, se presentó la plataforma TACLAC y se hizo hincapié en el carácter colaborativo de la misma (*How to use TACLAC*), así como en las posibles utilidades que podría tener en sus futuras tareas de traducción. Sirvió también para ejemplificar la tipología de corpus anteriormente explicada, puesto que la plataforma cuenta con varios tipos de corpus (monolingües, comparables y paralelos) (*Learn about corpora*) y los distintos programas de gestión de corpus (p. ej., Sketch Engine o AntConc). Finalmente, los alumnos trabajaron con el corpus 1 y los dos programas de concordancias con el fin de comparar y ver las distintas utilidades. El profesorado les proporcionó instrucciones precisas como: 1) crear un listado de palabras ordenadas por frecuencia; 2) obtener un listado de unidades terminológicas monoléxicas y poliléxicas más frecuentes; 3) obtener el comportamiento sintáctico de una de las palabras más frecuentes del corpus. Como puede observarse, se trata de tareas sencillas, pero no debemos olvidar que el tiempo era muy

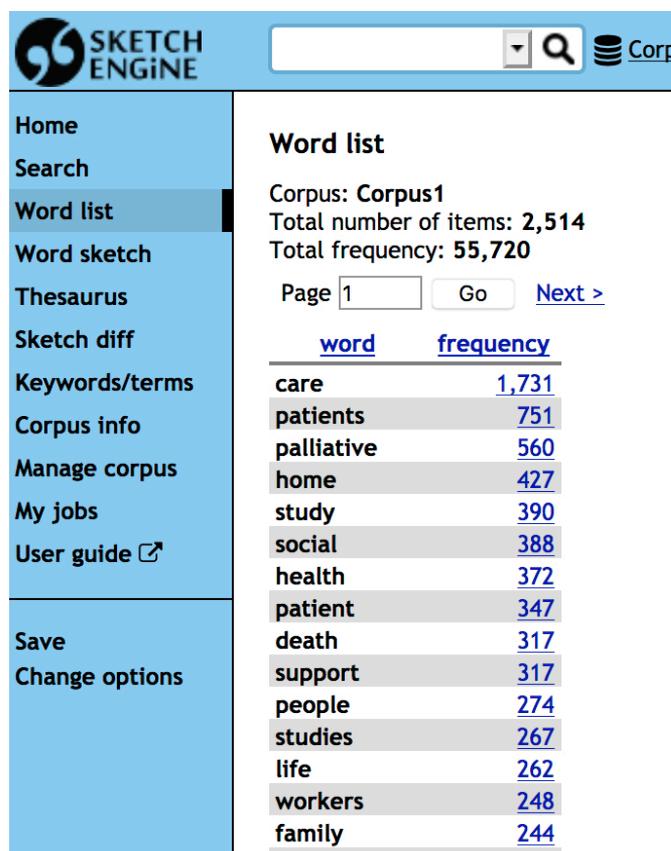


FIGURA 2. Ejemplo de word list creada con Sketch Engine

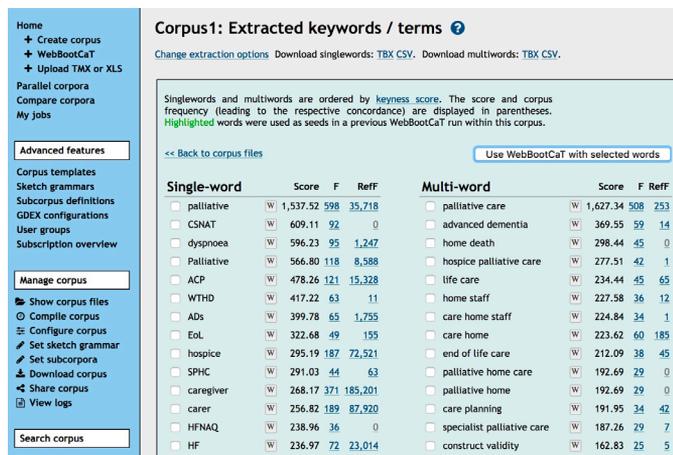


FIGURA 3. Ejemplo de la función keywords de Sketch Engine

reducido, por lo que las instrucciones debían ser muy precisas.

En este sentido, crearon listados de palabras con la función *word list* de ambos programas (figura 2) y pudieron comprobar que, para conseguir unos resultados óptimos, había que subir las llamadas *stoplist*⁵, a fin de no incluir el llamado «ruido» o palabras vacías de contenido léxico (*the, and, a, etc.*) y que, con seguridad, serían las más frecuentes en el corpus.

Los alumnos hicieron uso de la función *keywords* de Sketch Engine (figura 3), donde se ofrecían las unidades terminológicas más frecuentes monoléxicas y poliléxicas. De características similares, la opción de *Ngrams/Clusters* de AntConc les ayudó a conocer cómo este programa obtiene los patrones de dos palabras más frecuentes (p. ej., *palliative care, social workers, nursing home, health professional*).

Por otro lado, y a la vista de que la unidad terminológica más frecuente era *palliative*, se utilizó la función *word sketch* de Sketch Engine para comprobar el comportamiento sintáctico de un término concreto (figura 4).

3.3. El corpus comparable bilingüe (EN-ES) sobre cardiopatía isquémica (corpus 2)

Este segundo corpus fue compilado por alumnos de la asignatura de posgrado que previamente habían asistido a la sesión descrita en el apartado 3.2 y donde habían utilizado distintas

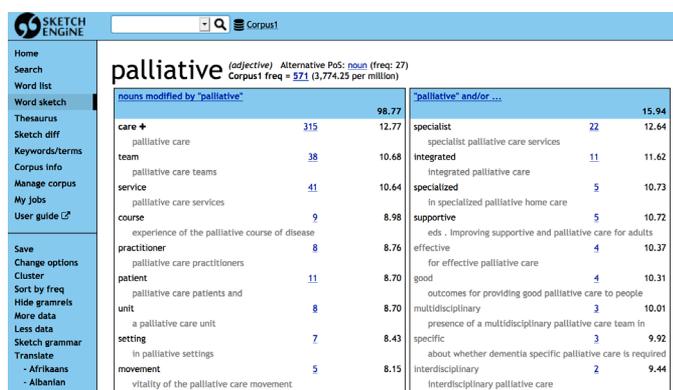


FIGURA 4. Ejemplo de la función word sketch de Sketch Engine

herramientas de gestión de corpus (SketchEngine y AntConc). La clase quedó dividida en grupos de seis y, como actividad, se pidió que compilaran un corpus especializado⁶, siempre teniendo en cuenta que el corpus compilado serviría como recurso documental para familiarizarse con alguno de los temas con los que trabajarían en las siguientes clases de traducción especializada. Junto con la mencionada tarea, los alumnos debían entregar un informe breve donde describirían cómo había sido su proceso de diseño del corpus de acuerdo con las fases de documentación, compilación y análisis. Se exponen a continuación las fases de diseño de uno de los grupos que compiló un corpus de temática biosanitaria (cardiopatía isquémica). En definitiva, la actividad suponía la creación de un corpus como recurso de documentación previo a la fase de traducción con el fin de que los estudiantes pudieran concienciarse de 1) la importancia de las necesidades documentales en el ámbito de la traducción biosanitaria y 2) la creación de material de calidad que permitiera una consulta inmediata y eficaz para la adquisición de conocimiento conceptual, terminológico y fraseológico en posteriores tareas de traducción, ya que el corpus se compartiría en la plataforma TAFLAC.

Puesto que se pedía el diseño de un corpus para una tarea concreta, el diseño de un corpus comparable (EN-ES) virtual especializado para tareas de documentación, no se necesitaba un corpus de grandes dimensiones, pero sí un corpus de calidad. Así, se propuso a los alumnos seguir los parámetros de Bowker y Pearson (2002) (tabla 2).

TABLA 2. Criterios de creación del corpus 2 (Bowker y Pearson, 2002)

Criterios de compilación del corpus	Características del corpus
Número de palabras (<i>tokens</i> y <i>types</i>)	inglés → <i>tokens</i> (26 264), <i>types</i> (3375) español → <i>tokens</i> (37 821), <i>types</i> (6904)
Número de textos	18 textos (español); 17 textos (inglés)
Medio	textos escritos
Temática	especializada: infarto de miocardio
Tipos de texto	portales especializados
Idiomas	inglés británico; español peninsular
Periodo de tiempo	2010-2016

En la fase de documentación, y con la ayuda de descriptores y búsquedas avanzadas, los alumnos procedieron a buscar y seleccionar textos a través de palabras clave como *myocardial infarction* y *coronary artery disease*, para el caso del subcorpus en inglés (EN), y *cardiopatía isquémica* o *infarto de miocardio*,

^ Back to top

Introduction

About

Symptoms

Diagnosis

Treatment

Aftercare

Causes

Complications

Prevention

FAQs

Diagnosis of a heart attack

Either in the ambulance or when you get to hospital, a doctor, nurse or paramedic will ask you about your symptoms. You'll have an **ECG (electrocardiogram)** as soon as possible to check the electrical activity of your heart. This can often show whether or not you're having a heart attack.

Your doctor will examine you - listening to your heart, taking your blood pressure and checking your heart rate. Other tests that your doctor may recommend, either immediately or over the next few days in hospital, are described here.

- Blood tests to check for any damage to your heart muscle. These will include a test for a protein called troponin, which is produced when heart muscle is damaged. Troponin can take between three and 12 hours to appear in your blood. Because of this delay your doctor will repeat the test at intervals to check for changes.
- Further ECGs - sometimes an ECG can be normal at first even if you've had a heart attack, so you may need to have the test again.
- A chest **X-ray**. This will help your doctor see whether there might be a cause for your chest pain other than a heart attack. It also checks for fluid in your lungs, which might happen if your heart isn't working well after a heart attack.
- A **coronary angiogram** - your doctor will give you an injection of a special dye into your coronary arteries to make them clearly visible on X-rays. This test can show where there are blockages or narrowings in your coronary arteries.
- An **echocardiogram** - this uses ultrasound (sound waves) to show the pumping action of your heart and valves. It can detect damage to your heart muscle, and how this might have affected how well it works.

FIGURA 5. Ejemplo de un portal especializado

para el subcorpus en español (ES), en sitios web institucionales o especializados⁷.

Además de utilizar palabras clave, los alumnos también emplearon los llamados *operadores filtro* para restringir los parámetros de dominio o sitio, puesto que se pedía un corpus comparable delimitado diatópicamente (inglés británico y español peninsular). De este modo, por ejemplo, utilizaron la ecuación «myocardial infarction site: bupa.co.uk» con el fin de que el buscador recuperara una selección de textos especializados dentro del sitio web bupa.co.uk donde apareciera el término *myocardial infarction*. Con la anterior ecuación, se restringía la búsqueda de manera doble: 1) a textos de procedencia bajo el dominio .co.uk (inglés británico) y 2) a textos dentro de un mismo portal especializado. En aras de buscar una mayor especialización, los alumnos emplearon la fórmula «intitle myocardial infarction», que delimitaba la búsqueda a aquellos sitios web que contenían las palabras especificadas en el elemento <title>.

Una vez ubicados los textos, los alumnos procedieron a la fase de compilación en sí. Se realizó la descarga, almacenamiento y conversión a texto plano (.txt) de todos los textos del subcorpus en inglés y del subcorpus en español. Los alumnos utilizaron la herramienta HTMLasText, que permite convertir en lote documentos html a txt. En cuanto al almacenamiento, los alumnos nombraron de forma manual los textos y asignaron el código del idioma más una cifra (es decir, EN1, EN2 para el subcorpus en inglés y ES1, ES2 para el subcorpus en español). Al ser los archivos html, el procedimiento de conversión fue crucial, ya que se exigía una limpieza de los mismos para evitar

etiquetas o expresiones como *Menu, Skip to main content, back to, Advertisement, Continue Reading Below*, etc.

En cuanto a la fase de análisis, los alumnos ya la habían practicado en la sesión anterior. No obstante, se les pidió que realizaran tareas de análisis similares a las llevadas a cabo en la sesión 1. Para el análisis del corpus creado, los alumnos utilizaron el programa de gestor de concordancias AntConc, que ya habían utilizado en la sesión anterior. El análisis se dividió en dos etapas relacionadas con las principales funciones del programa y las tareas requeridas: 1) análisis terminológico, para la elaboración de un glosario bilingüe, y 2) análisis de comportamiento colocacional de alguna de las palabras de mayor frecuencia. Así, en lo referido al análisis terminológico, los alumnos utilizaron la función *word list* para las unidades terminológicas monoléxicas, y *Ngrams/Collocations* para las unidades terminológicas poliléxicas, para generar un listado de las palabras más frecuentes en los dos subcorpus, monoléxicas y poliléxicas, que les permitiría posteriormente crear un glosario. Como puede verse en la figura 6, la comparación de ambos listados permite identificar equivalentes en ambos idiomas.

En el caso de los patrones colocacionales, mostramos como ejemplo las colocaciones más frecuentes que aparecen al utilizar la ecuación de búsqueda «coronary*», que buscará aquellas palabras más frecuentes que tengan delante el término *coronary* (figura 7).

Una vez identificado el patrón colocacional, el programa da la oportunidad de comprobar el contexto de cada uno de ellos (figura 8).

Rank	Freq	Word	Lemma Word Form(s)	Rank	Freq	Word
1	565	heart		1	341	infarto
2	318	attack		2	206	corazón
3	156	blood		3	179	miocardio
4	102	coronary		4	154	más
5	88	symptoms		5	118	síntomas
6	75	chest		6	116	dolor
7	69	pain		7	98	sangre
8	59	risk		8	89	riesgo
9	59	treatment		9	88	son
10	57	cardiac		10	79	agudo
11	50	arteries		11	73	cardíaco
12	49	people		12	71	tratamiento
13	46	artery		13	70	arterias
14	43	muscle		14	69	coronaria
15	39	attacks		15	67	ataque

FIGURA 6. Ejemplo de unidades terminológicas monoléxicas creadas con AntConc en ambos corpus

Total No. of Collocate Types: 43 Total No. of Collocate Tokens: 203

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	28	1	27	5.41631	artery
2	23	23	0	0	the
3	22	0	22	4.94809	arteries
4	18	1	17	1.16033	heart
5	17	17	0	0	a
6	15	2	13	5.71747	angioplasty
7	10	10	0	0	your
8	10	1	9	5.86948	thrombosis
9	5	5	0	6.13251	percutaneous
10	5	0	5	6.13251	intervention
11	4	4	0	3.96259	blocked
12	4	1	3	6.54755	angiography
13	3	3	0	0	with
14	3	3	0	0	have
15	2	0	2	4.81058	syndrome
16	2	2	0	0	having
17	2	2	0	2.12251	disease

Search Term: Words Case Regex Window Span: Same
 From... TL To... 1R

 Sort by: Invert Order Min. Collocate Frequency:

FIGURA 7. Ejemplo de colocaciones creadas con AntConc

4. Conclusiones

La necesidad de servicios de traducción e interpretación en las administraciones públicas ha supuesto una demanda de una formación acorde a dichas necesidades. Si bien al comienzo los avances fueron lentos, son muchos los programas formativos que existen en el ámbito universitario. Esta formación debe sustentarse en el desarrollo de una serie de habilidades adaptadas a la especificidad de la TISP. Nuestro trabajo se ha centrado en la competencia documental y en los recursos llamados corpus que, a nuestro parecer, pueden ser de gran ayuda en las fases previas de traducción y que suelen ser recursos desconocidos, en su mayor parte, por los alumnos de traducción.

La propuesta de plataforma TAFLAC surge precisamente de esta necesidad concreta de solventar la falta de familiarización de los alumnos de la TISP con estas herramientas electrónicas, además de servir como lugar de intercambio de estos recursos documentales. TAFLAC, que se encuentra en su primera fase de desarrollo, cuenta con dos corpus de temática biosanitaria que ya han sido utilizados en el aula de TISP (ámbito biosanitario). Es, por tanto, una propuesta que puede ayudar tanto al docente como al alumnado de TISP. De un lado, TAFLAC es una fuente de recursos para el docente, quien puede utilizarla para crear sus propios materiales de trabajo o bien para la enseñanza en el uso de corpus y gestores de corpus. Por otro, los discentes o cualquier profesional interesado en corpus puede emplear la plataforma para documentarse sobre un tema concreto o bien para crear materiales de ayuda en sus futuras tareas de traducción (p. ej., glosarios o memorias de traducción).

Notas

1. El presente trabajo se enmarca en el seno del grupo de innovación docente «Herramientas informáticas para la traducción e interpretación y su docencia» (Universidad de Alcalá, Madrid) y la red temática TRAJUTEC (Universidad de Málaga).

Concordance Hits 13

Hit	KWIC	File
1	atheter. This procedure is known as coronary angioplasty, usually called	EN3.txt
2	main ways to do this. A coronary angioplasty (also known as	EN8.txt
3	s a planned procedure afterwards. A coronary angioplasty aims to widen y	EN8.txt
4	in all circumstances. Occasionally a coronary angioplasty doesn't work, c	EN8.txt
5	: one week if you had a coronary angioplasty, it was success	EN8.txt
6	r your heart is pumping properly. Coronary angioplasty A coronary an	EN8.txt
7	properly. Coronary angioplasty A coronary angioplasty is a procedure	EN8.txt
8	cedure to widen the coronary artery (coronary angioplasty). Coronary ang	EN15.txt
9	continued for some time after PCI. Coronary angioplasty Coronary angic	EN15.txt
10	time after PCI. Coronary angioplasty Coronary angioplasty is a potentially	EN15.txt
11	of the UK's regions. During coronary angioplasty, a tiny tube kno	EN15.txt
12	orse. Coronary artery bypass graft A coronary angioplasty may not be tec	EN15.txt
13	s some cases, further treatment with coronary angioplasty or coronary art	EN15.txt

Search Term: Words Case Regex Search Window Size: 50
 From... TL To... 1R
 1
 Kwic Sort

FIGURA 8. Ejemplo de KWIC (Key word in wontext) en AntConc

2. A partir de ahora, nos referiremos a la traducción debido a que nos centraremos en los distintos trabajos sobre corpus en esta área y a que la ejemplificación del uso de TAFLAC se realizó dentro de la asignatura Técnicas y Recursos para la TISP para el ámbito de la traducción.
3. Junto con los dos corpus de temática biosanitaria que se muestran, TAFLAC ofrece dos ejemplos de corpus, uno comparable y otro paralelo, dentro del ámbito jurídico-administrativo hasta la fecha, si bien el objetivo final es que este número se vaya incrementando.
4. El concepto de *token* está relacionado con el de palabra, y *type*, con el de la forma de la palabra. De este modo, *type* se corresponde con todas las palabras distintas de un texto, mientras que *token* tiene que ver con todas las palabras de un texto, sin importar si se repiten o no.
5. Algunas direcciones útiles para la descarga de las llamadas *stoplists* son <http://www.lextek.com/manuals/onix/stopwords1.html> y <http://www.ranks.nl/stopwords>.
6. Los alumnos podían seleccionar entre los ámbitos de traducción biosanitaria o jurídico-administrativa, puesto que eran los módulos de especialidad del máster y las áreas para las que les podría servir el corpus creado.
7. Algunos de los portales especializados fueron: <https://familydoctor.org/condition/heart-attack>, <https://labtestsonline.es/conditions/infarto-agudo-de-miocardio-y-sindrome-coronario-agudo> y https://www.onmeda.es/enfermedades/infarto_miocardio.html.

Referencias bibliográficas

Alcina, Amparo; Victoria Soler y Anna Estellés (2005): «Internet como instrumento para la documentación en terminología y traducción. Hacia las plataformas de recursos electrónicos para el traductor especializado», en Dora Sales Salvador (ed.): *La biblioteca de Babel. Documentarse para traducir*. Granada: Comares, pp. 221-241.

- Aston, Guy (1999): «Corpus use and learning to translate», *Textus*, 12: 289-314.
- Bowker, Lynne y Jennifer Pearson (2002): *Working with specialized language: A practical guide to using corpora*. Londres: Routledge.
- Calzada Pérez, María (2005): «Corpus electrónicos como herramientas de documentación y formación de traductores», en Dora Sales Salvador (ed.): *La biblioteca de Babel. Documentarse para traducir*. Granada: Comares, pp. 163-199.
- Castillo Rodríguez, Cristina y Alexandra Santamaría Urbieta (en prensa): «Extracción de patrones lingüísticos en inglés y en italiano: un caso práctico de corpus etiquetado en el ámbito del turismo de salud y belleza», en Miriam Seghiri (ed.): *Integración de nuevas herramientas TIC basadas en corpus en el aula de traducción especializada*. Berna: Peter Lang.
- Cid-Leal, P. y R. Perpinyá-Morera (2015): «Competencia informacional en Traducción: análisis de los hábitos de los estudiantes», *BID, Textos universitarios de biblioteconomía y documentación*, 34.
- Corpas Pastor, Gloria (2012): «Corpus, Tecnología y Traducción», en Miguel Casas y María García Antuña (eds.): *XII Jornadas de Lingüística*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz, pp. 75-98.
- Durán Muñoz, Isabel (2010): «Specialized lexicographical resources: a survey of translators' needs», en Sylviane Granger y Magali Paquot (eds.): *eLexicography in the 21st century: New challenges, new applications. Proceedings of ELEX2009*. Lovaina-La-Nueva: Presses Universitaires de Louvain, pp. 55-66.
- Expert Advisory Group on Language Engineering Standards (EAGLES) (1996): *Text corpora Working Group reading Guide. EAGLES Document EAG-TCWG-FR-2*. <<http://www.ilc.cnr.it/EAGLES/corpintr/corpintr.html>> [consulta: 20.II.2019].
- Fernández Rodríguez, Áurea (2015): «La competencia documental en el aprendizaje de la traducción económica francés-español-francés», *InTralinea*, Special issue: New insights into specialized translation. <http://www.intralinea.org/specials/article/la_competencia_documental_en_el_aprendizaje_de_la_traducción_economica> [consulta: 19.II.2019].
- Gallego Hernández, Daniel (2015): «The use of corpora as translation resources. A study based on a survey of Spanish professional translators», *Perspectives: Studies in Translatology*, 23 (3): 375-371.
- Ortego Antón, María Teresa y Purificación Fernández Nistal (en prensa): «Estudio contrastivo de la terminología de embutidos en inglés y en español con ParaConc y tlCorpus a partir del corpus paralelo P-GEFEM y del comparable C-GEFEM», en Miriam Seghiri (ed.): *Integración de nuevas herramientas TIC basadas en corpus en el aula de traducción especializada*. Berna: Peter Lang.
- Palomares Perrault, Paloma (2000): *Recursos documentales para el estudio de traducción*. Málaga: Universidad de Málaga.
- Pinto Molina, María y Dora Sales Salvador (2008): «INFO-LITRANS: A model for the development of information competence for translators», *Journal of documentation*, 64 (3): 413-437.
- Rabadán, Rosa y Purificación Fernández Nistal (2002): *La traducción inglés-español: fundamentos, herramientas, aplicaciones*. León: Servicio de publicaciones de la Universidad de León.
- Rodríguez Martínez, Manuel C. y Emilio Ortega Arjonilla (2018): «El corpus de prospectos farmacéuticos como recurso didáctico en el aula de traducción especializada francés-español», *MonTi, Monografía de Traducción e Interpretación*, 10: 117-140.
- Sales Salvador, Dora (2006): *Documentación aplicada a la traducción: presente y futuro de una disciplina*. Gijón: Ediciones Trea.
- Sánchez Ramos, M.ª del Mar y Raquel Lázaro Gutiérrez (en prensa): «Traducción de consentimientos informados y lingüística de corpus: una propuesta metodológica para el aprendizaje de la traducción de textos médico-jurídicos», en Miriam Seghiri (ed.): *Integración de nuevas herramientas TIC basadas en corpus en el aula de traducción especializada*. Berna: Peter Lang.
- Sánchez Ramos, M.ª del Mar (2017): «Compilación y análisis de un corpus ad hoc como herramienta de documentación electrónica en Traducción e Interpretación en los Servicios Públicos (TISP)», *Estudios de Traducción*, 7: 177-190.
- Sánchez Ramos, M.ª del Mar y Francisco J. Vigier Moreno (2016): «Using monolingual virtual corpora in public service legal translator training», en Elena Martín-Monje, Iduerne Elorza y Blanca García Riaza (eds.): *Technological Advances in Specialized Linguistic Domains: Practical Applications and Mobility*. Londres: Routledge, pp. 228-239.
- Seghiri, Miriam (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): 13-30.
- Seghiri, Miriam (2017): «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): 43-64.
- Toledo Báez, Cristina (en prensa): «SketchEngine en traducción científico técnica (francés-español): creación y explotación del corpus ad hoc comparable GeneCorp», en Miriam Seghiri (ed.): *Integración de nuevas herramientas TIC basadas en corpus en el aula de traducción especializada*. Berna: Peter Lang.
- Valero Garcés, Carmen (2005): «La difícil tarea de documentarse en traducción e interpretación en los servicios públicos», en Dora Sales Salvador (ed.): *La biblioteca de Babel. Documentarse para traducir*. Granada: Comares, pp. 99-121.
- Valero Garcés, Carmen (2013): «Formación de traductores e intérpretes en una sociedad multicultural. El programa de la Universidad de Alcalá, Madrid», *Cuadernos de Aldeu*, 25: 215-238.
- Zanettin, Federico (2012): *Translation-driven corpora*. Manchester: St. Jerome.