

# CÓRPORA *AD HOC* EN LA PRÁCTICA TRADUCTORA ESPECIALIZADA: APLICACIÓN AL ÁMBITO DE LAS ENFERMEDADES NEUROMUSCULARES

Tamara Varela Vila  
[tamaravarela@gmail.com](mailto:tamaravarela@gmail.com)

## ABSTRACT

En este trabajo se analizan las ventajas de la utilización de corpórea monolingües especializados y creados *ad hoc* en la práctica traductora. Consideramos que la utilización de un corpus monolingüe como base de datos especializada es una opción muy recomendable, especialmente si el traductor lo crea según sus necesidades precisas de documentación. Dicho corpus constituirá una fuente de información muy fiable y fácil de explotar gracias a las numerosas herramientas existentes para ello.

Propondremos, por otro lado, una metodología simple y efectiva para crear corpórea *ad hoc*, siempre orientados a resolver los problemas puntuales derivados de la práctica traductora especializada. Asimismo, compilaremos, a modo de ejemplo, un corpus en español sobre un subdominio de la medicina en el que estos momentos estamos trabajando: el de las enfermedades neuromusculares. Se trata de un conjunto muy diverso de enfermedades raras que tienen como efecto común la pérdida de la capacidad muscular.

**PALABRAS CLAVE:** corpus *ad hoc*, enfermedades neuromusculares, traducción especializada, traducción médica.

## I. INTRODUCCIÓN

Es innegable que la actividad traductora ha cambiado sustancialmente con la democratización de las nuevas tecnologías de la información y la comunicación (TICs) y sobre todo con la llegada de Internet. Los traductores han tenido que adaptarse a esta nueva situación, de modo que ya no sólo tienen que saber traducir, sino que se espera que tengan ciertos conocimientos informáticos, de maquetación, de comunicación electrónica, etc. Por otro lado, estas tecnologías han supuesto un avance indiscutible en las labores documentales y terminológicas, ya que con sólo un ordenador y una conexión a Internet, el traductor puede acceder a una cantidad inimaginable hasta hace pocos años de información.

De manera específica, la lingüística de corpus pone al alcance del traductor de hoy en día un buen número de herramientas que permiten el acceso a una gran cantidad de información, que éste tendrá que saber aprovechar si quiere sacar el máximo partido a su tiempo. Si bien el trabajo con corpórea es ya habitual, pues los traductores han comenzado a ser conscientes de las ventajas que supone su explotación, exige un esfuerzo considerable y una inversión de tiempo que éste, en muchos casos, no puede permitirse. Por este motivo, los traductores se apoyan a veces en corpórea ya creados por otros especialistas u organismos, que intentan adaptar a sus objetivos. Sin embargo, en muchos casos los corpórea existentes no proporcionan la información deseada, por lo que es importante que el traductor sea capaz de crear un corpus *ad hoc* que responda a sus necesidades de documentación.

En este trabajo proponemos una metodología simple y efectiva para que los traductores puedan crear sus propios *córpora* monolingües *ad hoc* especializados sin tener que recurrir a un experto en dicha disciplina<sup>1</sup>.

## II. LA DOCUMENTACIÓN EN TRADUCCIÓN ESPECIALIZADA

Como sabemos, la documentación es un pilar fundamental del proceso de traducción. Cuanto más especializado sea un texto, mayor será la necesidad de adquirir conocimientos sobre ese ámbito para comprenderlo a la perfección, conocer la terminología utilizada, dominar las convenciones textuales, etc.

Por lo tanto, existe una estrecha relación entre documentación y traducción, pues cada una de ellas actúa como ciencia auxiliar para la otra (Recoder & Cid, 2004). Resulta indispensable que el traductor se documente a fondo cuando está realizando una traducción, por lo que deberá conocer las técnicas y procedimientos del ámbito documental.

Mientras que tradicionalmente la documentación se basaba en la consulta de fuentes como enciclopedias, diccionarios en formato papel o manuales especializados, hoy en día Internet es la fuente básica de la que sacar información. En efecto, el proceso documental ha ido modificándose poco a poco gracias a las nuevas tecnologías y especialmente a Internet. En primer lugar, el traductor ya no tiene que salir de su lugar de trabajo para acceder a la información que precisa. En segundo lugar, «la dificultad no está en hacer un gran acopio de datos, ya que eso ya nos lo ofrece Internet, sino cómo filtrar la información relevante, almacenarla masivamente y extraer de forma inequívoca los datos buscados» (Alcina, Soler & Estellés, 2005). Es importante, por lo tanto, analizar las fuentes encontradas para excluir aquellas que no ofrecen información totalmente fiable o relevante.

Efectivamente, la web es un recurso de gran utilidad para realizar búsquedas de información, aunque encontrar los datos que nos interesan en ella exige un buen dominio de los motores de búsqueda. De hecho, la utilización de estas herramientas no está exenta de dificultades. Las desigualdades en la cantidad de información encontrada en los distintos idiomas, la dudosa calidad de la información, la poca precisión de los resultados y los problemas del motor de búsqueda para localizar determinada información (Alcina, Soler & Estellés, 2005) son algunas de ellas. El traductor tendrá que recurrir a distintas estrategias para sacar el máximo partido a la información contenida en la web.

Además de la utilización de motores de búsqueda, también podemos utilizar las recopilaciones de información que han sido seleccionadas por otros especialistas y que, en muchos casos, se encuentran disponibles en ciertas páginas web, revistas especializadas, manuales sobre traducción especializada, etc. En ellas encontraremos vínculos a páginas web de especial interés para nuestro ámbito de estudio.

Por otra parte, los expertos en un determinado ámbito son una gran fuente de información a la que no debemos dudar en recurrir. Podemos contactar con expertos directamente a través del correo electrónico, o a través de medios más indirectos como listas de distribución o centros de expertos. Ellos podrán tanto resolvernos dudas conceptuales o terminológicas como ofrecernos referencias relevantes sobre el ámbito que nos interesa.

Toda esta información que el traductor obtenga debe ser almacenada metódicamente para poder localizarla posteriormente con la mayor facilidad, ya sea guardando los enlaces en ‘marcadores’ o ‘favoritos’, o descargando esa información en nuestro disco duro. Asimismo,

todos esos documentos pueden servirnos para compilar corpórea especializados, y especialmente corpórea *ad hoc*, en cuya descripción y utilidades nos centraremos más adelante.

### III. LA DOCUMENTACIÓN EN TRADUCCIÓN MÉDICA

Si bien la necesidad de documentarse es aplicable a cualquier traducción sobre un área especializada, nuestro trabajo se centra en el ámbito de la biomedicina, y más particularmente en el subdominio de las enfermedades neuromusculares<sup>2</sup>, por lo que nuestras necesidades documentales serán ligeramente diferentes.

La documentación cobra una especial relevancia en la traducción médica, debido precisamente a las peculiaridades de este ámbito y a lo que está en juego: la vida de los pacientes. La medicina está subdividida en múltiples especialidades y subespecialidades, lo cual hace imposible (incluso para los médicos) tener conocimientos sobre todas ellas. Por lo tanto, es imprescindible que el traductor se especialice en un determinado ámbito y que se documente bien antes de realizar una traducción de este tipo. No obstante, el trabajo del traductor no está exento de cuestionamientos. Como sabemos, se ha discutido mucho sobre su capacidad para realizar traducciones médicas de buena calidad sin ser un especialista médico.

Por otra parte, los conceptos médicos están sometidos a una constante evolución debido al desarrollo de las diferentes áreas que componen el ámbito médico y a los avances tecnológicos que cada vez son más rápidos, por lo que la necesidad de documentación es constante. El traductor tendrá que utilizar todo los medios a su disposición para documentarse (diccionarios especializados, corpórea, tesauros, enciclopedias, la web...), sin olvidar la consulta con expertos del ámbito médico sobre el que traduce, pues éstos son la mayor fuente de conocimiento especializado.

Los cuatro tipos básicos de información que un traductor médico necesita son los siguientes (Montalt & González Davies, 2007):

- Conocimiento del ámbito

Hace referencia a aquella información que sirve como introducción a un área de conocimiento y que nos proporciona una base de conocimiento sobre el tema del texto. Algunas de las fuentes que podemos consultar son enciclopedias, libros de texto, artículos especializados, etc.

- Terminología

Se refiere tanto a los términos, como a las abreviaciones, nomenclaturas, símbolos, etc. Aunque los recursos más habituales son diccionarios o glosarios, también deben utilizarse enciclopedias, libros de texto, artículos de investigación, expertos en el ámbito, etc.

- Información sobre géneros textuales

Necesitamos especialmente este tipo de información cuando el texto original no pertenece al mismo género que el texto meta. Por lo tanto, debemos familiarizarnos con las convenciones textuales y conocer las diferencias estructurales, estilísticas, discursivas, terminológicas, etc.

- Información lingüística

Se utiliza tanto para resolver dudas gramaticales como para conseguir una fluidez en la traducción. Dicha información puede encontrarse en diccionarios, tesauros, gramáticas y manuales de estilo.

En efecto, Montalt y González Davies (2007) consideran importante tanto la consulta de fuentes de introducción a temas médicos, como de diccionarios y enciclopedias especializados, portales, directorios o buscadores médicos, información sobre epónimos, abreviaciones, unidades de medida, nomenclaturas internacionales, patentes, etc. Cada traductor deberá saber qué información necesita y buscar aquellas fuentes que más se adapten a sus necesidades de información. El uso de corpórea como fuente de documentación médica no excluye el uso de estos otros recursos, sino que ambos recursos deben complementarse para obtener los mejores resultados.

#### **IV. CÓRPORA: CONCEPTOS BÁSICOS**

Como bien sabemos, un corpus no es sólo un conjunto de textos. Es un conjunto de textos escogidos utilizando criterios muy precisos, que sirven para representar una lengua o una parte de la lengua. Una de las definiciones más acertadas nos la ofrece Sinclair (2004), considerado el padre de la lingüística de corpus: «A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research».

La construcción de un corpus debe basarse en un pilar básico: la representatividad, es decir, la capacidad del corpus de representar una lengua o una variedad lingüística. Como podemos observar en la definición, Sinclair es bastante escéptico sobre dicha capacidad de representación. Ésta es una idea que recogen otros especialistas, pues consideran que ésta dependerá de las elecciones que el creador del corpus haga y más concretamente de su intuición (McEnery et al., 2006). Hay otros dos aspectos que pueden considerarse importantes en la construcción de un corpus: el muestreo y el equilibrio. El muestreo hace referencia al número de muestras incluidas en el corpus para representar cada variedad textual. El equilibrio tiene que ver con la inclusión en el corpus de textos sobre todas las subáreas que componen dicha parcela del conocimiento. Son aspectos muy relacionados entre ellos y con la representatividad, y también altamente subjetivos, por lo que el creador del corpus tomará decisiones que dependerán de su propia concepción de un corpus y de la lengua.

Una cuestión que ha creado muchas controversias es el tamaño de los corpórea. Evidentemente, el desarrollo de herramientas informáticas estos últimos años ha permitido la construcción de grandes corpórea con relativa facilidad. A pesar de todo, los especialistas no se ponen de acuerdo sobre lo que ellos consideran el tamaño perfecto de un corpus. En realidad, el tamaño depende del criterio del especialista y del objetivo del corpus (McEnery et al. 2006). Leech (1991), por ejemplo, considera que el tamaño del corpus no es tan importante, ya que un pequeño corpus puede contener ejemplos suficientes de los trazos lingüísticos que nos interesan. De acuerdo con este punto de vista, el tamaño necesario para crear un corpus de calidad es muy subjetivo y la calidad de los resultados va a depender también de la capacidad del especialista para sacar conclusiones a partir de las ocurrencias encontradas. Lo que sí es cierto es que los corpórea especializados son más reducidos que aquellos que intentan representar la lengua común.

Otra cuestión que continúa creando polémica entre los especialistas es la pertinencia de utilizar fragmentos de textos o textos completos. Para Biber (1993), es suficiente con

utilizar fragmentos de unas 2000 palabras, dado que los trazos lingüísticos se distribuyen en el texto de una forma homogénea. Sin embargo, aconseja utilizar extractos del principio, del medio y del final para asegurar el equilibrio. Otros especialistas, como Sinclair (2004), afirman que no es buena idea utilizar fragmentos, ya que una parte del texto no siempre es representativa del texto completo.

El corpus con el que trabajemos debe adaptarse siempre al objetivo que deseemos conseguir con él. Por ello existen diferentes tipos de córpora, de entre los que deberemos seleccionar aquel que más se adapte a nuestras necesidades. Existen múltiples maneras de clasificar los córpora atendiendo a sus características: si están formados por textos orales o escritos, si incluyen textos completos o fragmentos, según el número de lenguas representadas, el tipo de etiquetado, etc. (Laviosa, 2002: 33-42). Corpas (2004) incluye una nueva distinción entre córpora estables y córpora virtuales. Los estables son aquellos que presentan un tamaño y un diseño más o menos fijo, y que se encuentran listos para su explotación. Los córpora virtuales, por su parte, son aquellos en los que sólo se incorporan documentos cuyo formato original sea electrónico y su diseño se ve condicionado por la finalidad concreta que ha motivado su compilación. Los córpora *ad hoc*, en los que nos centraremos, son un subtipo de córpora virtuales.

## V. CÓRPORA *AD HOC*

Corpas (2004) define un corpus *ad hoc* como «un corpus virtual que se compila puntualmente para la realización de un determinado encargo de traducción en cualquier dirección (directa, inversa o indirecta). En su diseño no prima tanto la cantidad como la calidad: por regla general, el corpus *ad hoc* no incluye un número demasiado elevado de textos, pero sí textos muy adecuados, equiparables al texto original en cuanto a la temática, el género y la variedad textual». Dichos córpora han sido denominados por otros autores ‘disponible’ (Zanettin, 2002), ‘do-it-yourself/DIY’ (Zanettin, 2002) o ‘virtual’ (Corpas, 2004).

Por su parte, Zanettin (2002) considera que un corpus *ad hoc* es una colección de documentos de Internet, creada como respuesta específica a una necesidad de traducción. Además, afirma que a este corpus se puede añadir más material si es necesario y que no está destinado a formar parte de un corpus permanente.

Los córpora *ad hoc* se adaptan a la definición básica de corpus (Corpas, 2004), pero tienen la ventaja de haber sido creados para responder a cuestiones específicas. Pueden ser tanto comparables como monolingües (Zanettin, 2002), si bien los más habituales son los monolingües, debido a la mayor facilidad que supone su compilación. Sin embargo, es más difícil extraer la información que nos interesa de ellos, por lo que es necesario ser creativo al explotarlos, inventando estrategias para obtener resultados (Bowker & Pearson, 2002). En efecto, Bowker (1998) demostró que un corpus monolingüe especializado es un recurso muy eficiente para traducir, pues supone una gran ayuda para la comprensión del ámbito, la elección términos correctos y el uso de expresiones idiomáticas adecuadas.

Existen numerosos motivos por los que crear nuestro propio corpus *ad hoc*, aunque también es cierto que su compilación supone una inversión de tiempo importante. Aston (2002) nos da algunas razones por las que compilar un corpus propio. En primer lugar,

cuando compilamos un corpus tenemos un control sobre él y, además, podemos utilizar nuestros propios criterios para crear corpórea que no estén disponibles. Nos ofrece asimismo seguridad, ya que podemos conocer mejor sus límites y realizar los ajustes necesarios para mejorarlos, además de ser muy gratificante participar en la creación de una herramienta útil. Aston también afirma que es importante la capacidad crítica que desarrollaremos, pues poco a poco comprenderemos cómo afectan los criterios de diseño al resultado. Por último, considera que es importante crear corpórea *ad hoc* porque esto fomenta la comunicación y las discusiones orientadas a mejorar los resultados. Veremos ahora qué tipo de búsquedas podemos hacer en los corpórea *ad hoc* monolingües para mejorar el proceso traductológico.

En primer lugar, un corpus puede ayudarnos a comprender un texto. Siempre es aconsejable que los traductores lean textos sobre el ámbito de trabajo para familiarizarse con él antes de comenzar a traducir. Un corpus ofrece muchas ventajas en ese sentido, ya que proporciona numerosos textos fiables sobre ese ámbito. Por otra parte, si no comprendemos un fragmento concreto, posiblemente se deba a que no conocemos un término o que no entendemos el concepto. Si el problema es este último, basta con buscar este término en el corpus para encontrar otros contextos en los que su significado esté más claro. Se trata de una búsqueda similar a la que todos los traductores realizan en los motores de búsqueda de la web, con la ventaja de que el corpus, si se ha compilado bien, sólo va a ofrecernos material fiable. Esto nos ayudará a cometer menos errores derivados de la mala comprensión del texto origen.

Por otro lado, un corpus monolingüe es un recurso básico para reducir el número de errores (lingüísticos, estilísticos, terminológicos, etc.) cometidos al redactar el texto meta. Evidentemente, para un traductor no basta con entender los conceptos y conocer los términos equivalentes, sino que tiene que saber cómo utilizarlos. Una de las formas más simples y útiles es estudiar las colocaciones existentes en el corpus. Como bien nos dice Sinclair (1991), las colocaciones son «combinations of words that, according to the conventions of a given language, are habitually associated».

En general, los corpórea proporcionan más información estilística y sobre colocaciones que los diccionarios (Bowker & Pearson, 2002). La literatura especializada es siempre la mejor fuente de información para conocer las convenciones de un ámbito especializado.

En este sentido, el análisis de un corpórea *ad hoc* puede ayudarnos también a verificar o rechazar decisiones tomadas basándonos en otras herramientas, como diccionarios. A veces éstos nos proponen varias soluciones terminológicas, algunas poco adecuadas, otras con un contexto insuficiente para determinar cuál es el término que necesitamos. En caso de duda, el corpus puede ser una base de datos útil para ayudarnos a decidir cuál es la solución más adecuada. Además, también podemos valernos de este tipo de corpórea para buscar el equivalente a un término en otra lengua, basándonos en nuestras intuiciones, o simplemente buscando todas las palabras que contengan su misma raíz, un prefijo o un sufijo, etc.

Por otra parte, se trata de una herramienta muy valiosa para estudiar las convenciones textuales, que son tan importantes para conseguir que el texto meta sea aceptable en la lengua y en la cultura receptoras.

No debemos olvidar que los usos que se le puede dar a un corpus son infinitos y que, como ya hemos dicho, la información que se extraiga de ellos dependerá de si son representativos del ámbito de especialidad que se desea estudiar y, además, de la creatividad del traductor y su capacidad para extraer conclusiones.

## VI. CRITERIOS SEGUIDOS PARA LA CREACIÓN DE NUESTRO CORPUS *AD HOC* SOBRE ENFERMEDADES NEUROMUSCULARES

Como hemos visto, un corpus puede proporcionarnos, si no toda, una gran parte de la información que un traductor especializado necesita para realizar un trabajo de calidad. Para ilustrar la utilidad de los *córpore* monolingües especializados creados *ad hoc* nos centraremos en la creación de un corpus sobre el subdominio médico de las enfermedades neuromusculares, ámbito en el que actualmente estamos trabajando. Se trata de un conjunto de patologías normalmente de origen genético que se caracterizan por una afección de los nervios que controlan los músculos voluntarios, llevando a una debilitación muscular. Las enfermedades neuromusculares son múltiples (hay alrededor de 150 diagnósticos diferentes) pero afectan a un número de personas muy reducido (menos del 0,05% de la población), por lo que se consideran enfermedades 'raras'. Hasta ahora no se han hecho los suficientes esfuerzos económicos para que se investiguen en profundidad. Esta falta de investigación sobre ellas, especialmente en el ámbito del estado español, se ve reflejada también en la existencia de poca documentación al respecto en español (la mayor parte de ella traducida de otras lenguas) y de una pobre normalización de la terminología utilizada para denominarlas. Esto causa serias dificultades en el proceso traductológico: el proceso de documentación resulta complicado, ya que a la reducida existencia de documentación podemos añadir la confusión terminológica causada por la proliferación de sinónimos y términos polisémicos<sup>3</sup>.

Consideramos que sería interesante que los traductores aprendiesen a crear sus propios *córpore* *ad hoc* para resolver sus dudas durante el proceso de traducción. Aunque en un principio su compilación puede resultar demasiado laboriosa, estamos seguros de que la información que el traductor obtendrá servirá para compensar todo el tiempo invertido en la creación.

A pesar de que nos vamos a centrar en proponer un método sencillo y eficaz para crear un corpus *ad hoc* sobre las enfermedades neuromusculares, dichos pasos pueden ser extrapolados y utilizados para crear otros *córpore* especializados en otros ámbitos o en otras lenguas.

Antes de comenzar, debemos tener en cuenta que los *córpore* deben ser una muestra de la variedad de lengua o ámbito con el que se desea trabajar. Por lo tanto, un corpus no es una cantidad de texto recogida de manera aleatoria, sino que es un conjunto de documentos seleccionados siguiendo criterios muy precisos para que puedan representar el ámbito que deseamos. Debemos, entonces, decidir qué criterios utilizaremos para delimitar nuestro ámbito de estudios, de manera que el corpus se adapte a nuestras necesidades y sea representativo.

Para comenzar, utilizaremos dos criterios muy generales pero que nos parecen de gran utilidad, y que fueron propuestos por EAGLES (Sinclair, 1996):

- Cantidad

Los *córpore* especializados, como hemos afirmado, suelen ser más reducidos que los generales, pero la información que contienen está más condensada. No obstante, deben construirse *córpore* lo más amplios posible. Además, éstos deben estar formados por textos completos, dado que un fragmento puede no ser representativo un texto en su totalidad.

- Calidad

Los textos deben cumplir estrictamente el criterio de calidad, tanto en lo referente al autor del texto, como al contenido del mismo. Los autores deben ser especialistas en el ámbito que se estudia y el contenido de dichos textos correcto y pertinente.

Además de éstos, hay otros criterios más específicos que debemos seguir al crear córpora de este tipo, como los siguientes:

- Pertenencia al ámbito de especialidad

Debemos delimitar nuestro ámbito de especialidad para poder determinar qué textos podrán formar parte del corpus y qué textos, por ser demasiado generales o no adaptarse a dicho ámbito, no podrán hacerlo. En realidad, ésta es una tarea complicada y subjetiva, pues las fronteras entre las diferentes lenguas de especialidad son muy confusas (Cabré, 1993).

- Fecha de producción de los documentos

Normalmente, los creadores prefieren compilar córpora con documentos de reciente producción, puesto que éstos describen de una manera más precisa el estado en el que se encuentra ese lenguaje de especialidad.

- Procedencia y lengua de los autores

Es preciso también valorar si nos interesa trabajar con documentos originales o si, por el contrario, el hecho de que sean traducciones no afectará a la explotación del corpus. Aún así, debemos tener en cuenta que en el ámbito científico existe una gran internacionalización que hace difícil en muchos casos saber si un texto es original o traducido. Además, es necesario tener en cuenta si los autores de los textos son nativos (aunque hoy en día es también difícil de determinar) y si nos centraremos en autores de un determinado ámbito geográfico o si es indistinto.

- Género textual

Debemos diferenciar los géneros textuales existentes dentro del conjunto de los textos especializados. Se analizará el ámbito de especialidad en busca de los géneros más representativos, que deberán formar parte del corpus. En el caso del corpus sobre enfermedades neuromusculares, por ejemplo, los géneros textuales que tendremos en cuenta serán los artículos o manuales especializados, los artículos o manuales semi-especializados y los artículos de divulgación científica.

De entre los múltiples criterios que se pueden utilizar para crear un corpus especializado, éstos nos parecen los más relevantes.

En lo referente a nuestro ámbito de trabajo, hemos utilizado textos en español que traten el subdominio médico de las enfermedades neuromusculares y que hayan sido publicados a partir del año 2000, de manera que tanto la terminología como las convenciones textuales sean actuales. Por otro lado, no hemos tenido en cuenta la procedencia de los autores, puesto que muchos de estos textos son traducciones. Por último, hemos creído conveniente utilizar textos completos, como Sinclair (2004) recomienda, y siguiendo el criterio de cantidad que antes mencionábamos.

En total, el corpus está formado por 93 textos y cerca de 320.000 palabras. De los textos seleccionados, un 35% son artículos especializados, publicados en revistas científicas. Un 33% de los textos son semi-especializados, pues provienen de universidades, congresos u hospitales, pero sus receptores no son específicamente especialistas médicos, sino también los afectados y su entorno. De éstos, un 17% han sido publicados por asociaciones, especialmente por ASEM<sup>4</sup>. El restante 32% de los textos son de divulgación científica.



En nuestro corpus se incluyen los siguientes géneros textuales: artículos de revisión, notas clínicas, artículos de investigación, casos clínicos, editoriales, cartas al director, una tesis doctoral, comunicaciones, conferencias, informes, guías para el paciente, fichas técnicas, manuales, artículos de revistas, artículos de periódico y entrevistas. El porcentaje de textos perteneciente a cada género va a depender del ámbito científico en el que trabajemos y de la cantidad de textos de un u otro género existentes. Lo que sí es cierto es que, en nuestro trabajo de compilación de un corpus sobre enfermedades neuromusculares, optamos por no descartar ningún texto siempre que cumpliera el criterio de calidad y perteneciese a uno de los referidos géneros textuales más habituales, dada la escasez de textos sobre dicho ámbito.

Una vez reunidos todos los textos que cumplieran los criterios establecidos, procedimos a la creación del corpus propiamente dicho. Primero, pasamos todos los textos a formato TXT utilizando el programa Notepad++, que nos parece de gran utilidad, al permitir el tratamiento de varios documentos a la vez. Realizamos un tratamiento sistemático de todos los textos para que el programa de concordancias nos permita un análisis del corpus lo más claro posible. Eliminamos todo aquello que no formaba parte del texto original, como los números de página o los títulos del artículo o de un capítulo que se repiten en todas las páginas. Cada texto se guardó en un documento TXT junto con el texto original y se nombraron con la sigla 'CAHENM' (corpus *ad hoc* sobre enfermedades neuromusculares) y números correlativos.

## VII. EXPLOTACIÓN DEL CORPUS

Una vez realizados todos estos pasos, debemos estudiar nuestro corpus con un programa de concordancias. Para ello, hemos elegido el programa MonoConc Pro, pues nos parece de sencilla utilización para aquellas personas que no están acostumbradas a trabajar este tipo de programas. MonoConc Pro nos permite obtener listas de frecuencias, listas de palabras, líneas de concordancia KWIC (*keyword in context*), datos estadísticos, etc.

Como ya hemos dicho, debemos ser creativos a la hora de explotar el corpus. Aquí damos algunas ideas de cómo extraer información relevante de él, que no deben considerarse pasos a seguir sino simples propuestas de explotación:

### VII.1. Extracción palabras más frecuentes

Para comenzar, siguiendo los pasos de Corpas (2001), podemos extraer una lista de frecuencias de todo el corpus, con el objetivo de conocer las palabras que más se repiten en él. De ésta forma podremos saber si éstas reflejan los conceptos más importantes que, en principio, el corpus debería tratar. Éstas son las 20 palabras más repetidas en nuestro corpus:

Tabla 1 - Lista de frecuencias del CAHENM

1185: 0,3739% muscular
986: 0,3111% enfermedad

696: 0,2196% pacientes
617: 0,1947% gen
611: 0,1928% distrofia
566: 0,1786% enfermedades
501: 0,1581% forma
432: 0,1363% tipo
423: 0,1335% musculares
419: 0,1322% expresión
414: 0,1306% células
410: 0,1294% años
409: 0,1290% músculo
405: 0,1278% músculos
402: 0,1268% casos
387: 0,1221% paciente
378: 0,1193% cromosoma
369: 0,1164% proteína
369: 0,1164% vida
361: 0,1139% diagnóstico

MonoConc Pro nos ofrece el número de ocurrencias de cada palabra en el corpus junto con el porcentaje de aparición en él. Como podemos ver, las palabras que más aparecen son bastante reveladoras, pues son aquellas que describen estas enfermedades. Aunque el término 'distrofia', que se utiliza como sinónimo de 'enfermedad neuromuscular', sí es uno de los más utilizados en el corpus, la palabra 'neuromuscular' aparece en él 500 veces, 242 de ellas en singular.

## VI.2. Buscar definiciones

Podemos utilizar la cadena 'es una enfermedad' para buscar definiciones de algún tipo de patología. Por ejemplo, haciendo esa búsqueda en nuestro corpus hemos obtenido 110 definiciones diferentes de enfermedades neuromusculares.

También podemos encontrar información sobre una determinada enfermedad, haciendo una búsqueda como 'distrofia muscular de Duchenne es'. De esta forma, hemos encontrado esta información:

Tabla 2 - Contextos de 'distrofia muscular de Duchenne'

<ol style="list-style-type: none"><li>1. ... nsmisoras . ¿ cómo se manifiesta ? La [[distrofia muscular de Duchenne es]] una enfermedad que afecta al conjunto d ...</li><li>2. ... sentado en tu vida ? Mi enfermedad , la [[Distrofia Muscular de Duchenne es]] de las más graves dentro de las distrof ...</li><li>3. ... cker , distrofia de cintura , etc . la [[distrofia muscular de duchenne es]] la enfermedad neuromuscular más común e ...</li><li>4. ... nsmisoras . ¿ cómo se manifiesta ? La [[distrofia muscular de Duchenne es]] una enfermedad que afecta al conjunto d ...</li><li>5. ... n DMD , según los investigadores . la [[distrofia muscular de Duchenne es]] una enfermedad muscular degenerativa de ...</li></ol>
---

Como podemos ver, estos contextos ya nos dan algunas ideas de la información que podremos obtener. Por otro lado, es posible utilizar patrones como 'es un', 'es el' o 'un tipo de' para encontrar contextos definitorios (Bowker & Pearson, 2002).

Tabla 3 - Contextos definitorios

<ol style="list-style-type: none"><li>1. ... uso de inhibidores del mecanismo de inestabilidad somática de las repeticiones. [[un tipo de]] estrategia que tenga como diana la mutación a nivel de DNA per se ha sido motiv ...</li><li>2. ... sas o de necrosis nos permitirá comprobar si las células están sensibilizadas a [[un tipo de]] muerte celular programada o a otro . implicación de factores asociados ...</li><li>3. ... desconoce , podría tratarse de una enzima de tipo E3 ubiquitina ligasa , que es [[un tipo de]] enzimas que identifican las proteínas alteradas y las dirigen hacia el proteaso ...</li><li>4. ... el caso del ratón mdx . esto se puede hacer mediante inyección con plásmidos ( [[un tipo de]] vectores de DNA ) que contienen la parte codificante del gen o cDNA directament ...</li><li>5. ... conducción eléctrica de las aurículas hacia los ventrículos , bloqueo AV, , es [[un tipo de]] desorden de la conducción visto muchas veces en la DM tipo Emery Dreifuss . en ...</li></ol>
---

Podemos observar que, entre los cinco primeros contextos encontrados, los tres últimos nos ofrecen definiciones o explicaciones sobre determinados conceptos. Esta información puede ser muy útil a la hora de comprender un texto.

### VI.3. Encontrar equivalentes de términos en otras lenguas

Muchas veces, cuando buscamos el equivalente a un término extranjero, suponemos cuál es el término en nuestro idioma. El corpus es una herramienta muy útil para confirmar esas suposiciones. Si, por el contrario, no intuimos de qué término se puede tratar, podemos comenzar a buscarlo utilizando la raíz por la que podría estar formado, un prefijo o un sufijo.

Para realizar este tipo de búsquedas podemos utilizar los comodines, que son los siguientes:

- ?

Representa una sola letra y puede servirnos, por ejemplo, para determinar cómo se escribe cierto término, si tenemos dudas con una sola letra ('en?ima' para saber si se trata de 'encima' o 'enzima').

- % - letra opcional

Representa una letra opcional y es muy útil si deseamos buscar las ocurrencias de una misma palabra en singular y en plural, en caso de que éste se forme añadiendo una 's' ('distrofia%').

- \*

Comodín que representa una o varias letras, o una palabra, y que puede servir tanto para buscar qué palabra acompaña habitualmente a otras ('enfermedad de' serviría para buscar las enfermedades denominadas a través de epónimos) o para buscar, por ejemplo, palabras que contengan determinado prefijo o sufijo (la búsqueda 'mio\*' proporciona 'miopatía', 'miotonía', 'miotónica', 'miositis', 'mioblastos'...).

- @

La arroba representa un número determinado de palabras. En las opciones de búsqueda (*search options*) podemos determinar a cuántas palabras puede equivaler. Serviría para buscar determinadas secuencias, como 'enfermedad @ neuromuscular'. De esta forma podemos saber que en el corpus se utiliza 5 veces 'enfermedad genética de la unión neuromuscular'.

### VI.4. Conocer las colocaciones

Podemos utilizar la opción de colocaciones (*advance collocations*) para saber qué palabras acompañan normalmente a otras. Además, esta opción nos permite determinar qué posición ocupan las palabras que nos interesan. Por ejemplo, pueden interesarnos las dos palabras que aparecen a la derecha de aquella que buscamos o quizás la palabra anterior.

Por ejemplo, podemos buscar el término 'neuromuscular' e interesarnos por la primera palabra situada a su izquierda (1L). Así sabremos que en nuestro corpus, 'neuromuscular' va acompañado habitualmente de 'enfermedad' (82 veces), pero también de otros sustantivos, que pueden ser relevantes:

Tabla 4 - Colocaciones de 'enfermedad'

82	33,8843%	enfermedad neuromuscular
49	20,2479%	unión neuromuscular
21	8,6777%	patología neuromuscular
14	5,7851%	bloqueo neuromuscular
8	3,3058%	función neuromuscular
8	3,3058%	transmisión neuromuscular
7	2,8926%	enfermo neuromuscular
5	2,0661%	trastorno neuromuscular
4	1,6529%	biopsia neuromuscular
4	1,6529%	monitorización neuromuscular
3	1,2397%	clínica neuromuscular
2	0,8264%	patología neuromuscular
2	0,8264%	laboratorio neuromuscular
2	0,8264%	huso neuromuscular
2	0,8264%	causa neuromuscular
2	0,8264%	desorden neuromuscular

Si, por el contrario, queremos saber cuáles son las dos palabras que siguen al término 'miopatía', obtendremos la siguiente lista (de la que mostramos sólo las primeras ocurrencias):

Tabla 5 - Colocaciones de 'miopatía'

24	8,1356%	miopatía distal de
16	5,4237%	miopatía de cinturas
13	4,4068%	miopatía congénita de
13	4,4068%	miopatía del paciente
10	3,3898%	miopatía de bethlem
9	3,0508%	miopatía de duchenne
8	2,7119%	miopatía de miyoshi
7	2,3729%	miopatía y establecer
7	2,3729%	miopatía congénita miotubular
6	2,0339%	miopatía congénita centronuclear
5	1,6949%	miopatía de ver

5	1,6949%	miopatía distal con
---	---------	---------------------

Evidentemente, algunas ocurrencias no son relevantes, puesto que no muestran el patrón completo. Sin embargo, encontramos colocaciones habituales como 'miopatía de cinturas', 'miopatía de Bethlem', 'miopatía de Duchenne', 'miopatía de Miyoshi', etc. Sumergiéndonos en dichas listas de colocaciones y realizando diferentes búsquedas sucesivas podremos encontrar otros patrones habituales.

### VI.5. Investigar la sinonimia

También podremos utilizar el corpus para determinar cuál de dos sinónimos es el más utilizado en este tipo de textos. Por ejemplo, en nuestro ámbito se habla tanto de la 'marcha' como de la 'deambulación' para hacer referencia a la capacidad de los afectados de caminar. Si dudamos sobre cuál de las dos utilizar, podremos consultar el corpus. En este caso, 'marcha' se utiliza mucho más (170 ocurrencias) que 'deambulación' (22 ocurrencias).

### VI.6. Conocer el significado de una palabra

Además de esto, podemos buscar contextos de una determinada palabra para conocer su significado y cómo se utiliza, en el caso de que el corpus no contenga ninguna definición de ella. Nosotros hemos realizado la búsqueda del término 'expresión', que evidentemente no tiene el mismo significado en la lengua común que en este ámbito especializado. De ese modo, podemos obtener los siguientes contextos, que nos darán una idea de su significado:

Tabla 6 - Contextos de 'expresión'

<p>1. ... en pacientes con DM1 y en ratones mutantes homocigotos para Dmpk los niveles de [[expresión]] de DMPK decrecen y alteran el estado de la fosforilación de CUG, BP1 facilitando ...</p> <p>2. ... debidas al efecto de la metilación del locus DM1 y por tanto el aumento de su [[expresión]] en DM congénita por la pérdida de la actividad del elemento aislante ( Cho y Ta ...</p> <p>3. ... al . , 2000 ; Reddy et al . , 1996 ) . también se testó la hipótesis de que la [[expresión]] de los genes adyacentes a DMPK pudiera estar afectada en presencia de expansión ...</p> <p>4. ... 1 . , 2003a ) ( véase Apartado 5 . 2 de Introducción ) . además , el patrón de [[expresión]] de los genes MBNL solapa ampliamente con el de DMPK , lo cual es un requerimien ...</p> <p>5. ... IR e incluyendo el exón 7a para el CIC,1 . el patrón fetal del IR se debe a la [[expresión]] de una isoforma con más baja capacidad de señalización que la isoforma adulta . ...</p>
---

## VI.7. Confirmar datos de otras fuentes

Este tipo de *cópora* también es válido para confirmar si nuestra intuición o los datos proporcionados por otra fuente (por ejemplo, un diccionario), son correctos. Para ello podemos buscar la solución propuesta y estudiar las ocurrencias con el objetivo de determinar si se ajusta a aquello que estamos buscando. Por ejemplo, si conocemos la denominación francesa 'myopathie à bâtonnets', posiblemente un diccionario especializado nos ofrezca sólo la solución 'miopatía nemalínica'. Podemos comprobar la utilización de éste término en español, que nos servirá para descubrir que también existe la denominación 'miopatía con bastoncillos'.

## VI.8. Estudio de las convenciones textuales

Asimismo, a través del corpus podremos estudiar las convenciones textuales. Por ejemplo, si lo que nos interesan son los artículos científicos, podemos analizar qué tipo de contenido se suele incluir en la introducción o en la conclusión, o si es habitual que los autores introduzcan una parte dedicada a los agradecimientos.

Es posible también analizar los textos que componen el corpus para saber si en ellos se utiliza en gran medida la pasiva, la pasiva refleja, el gerundio, el condicional... En ese sentido, el estudio de un corpus resulta un recurso inagotable.

Como podemos ver, un corpus *ad hoc* puede servirnos como fuente de documentación para realizar traducciones especializadas de gran calidad. En este caso, el corpus está en español, puesto que (idealmente) ésta será la lengua hacia la que traduzcamos. Si la lengua meta de nuestra traducción fuese otra diferente a nuestra lengua nativa, un corpus *ad hoc* en dicha lengua sería de especial utilidad, puesto que nos proporcionaría un modelo fiable y real. Cada especialista debe analizar su situación y determinar cuáles son sus necesidades lingüísticas, y a partir de ahí crear un corpus adaptado a dichas necesidades.

## CONCLUSIONES

La traducción científica implica una gran necesidad de documentación. Comprender el texto que se va a traducir es imprescindible para realizar una traducción de calidad, transmitiendo los contenidos de una lengua a otra con total precisión. Además, el conocimiento de la terminología y de las convenciones de un determinado ámbito son aspectos totalmente básicos para redactar un texto que se adapte a la lengua y a la cultura receptoras.

El uso de *cópora* es un recurso muy valioso del que los traductores deberían servirse para aumentar de forma notable la calidad de los resultados de su trabajo. En muchos casos, los *cópora* existentes no se adaptan a las necesidades específicas de la traducción especializada, sobre todo si se trata de un ámbito muy reducido. Por ello es importante que el traductor sepa crear sus propios *cópora ad hoc* y así pueda resolver los obstáculos que se le planteen al realizar una traducción, ya sean problemas puntuales de terminología o

estructuras, o problemas más generales de comprensión o relacionados con las convenciones textuales.

En este trabajo se procedió a crear, a modo de ejemplo, un corpus en español sobre enfermedades neuromusculares. Primero se determinaron los criterios estructurales que dicho corpus para representar el ámbito especializado con el que se desea trabajar. Posteriormente se creó el corpus propiamente dicho, buscando documentos para formarlo, teniendo en cuenta sobre todo los criterios de calidad y cantidad. Se trató este corpus con el programa de concordancias MonoConc Pro, para obtener la información que consideremos necesaria. Lo que se pretendía era esencialmente dar ciertas ideas básicas de qué información se puede obtener con un programa de concordancias, para que los traductores, utilizando la imaginación, puedan trabajar con ellos y sacar sus propias conclusiones.

Estamos seguros de que este tipo de córpora podrá mejorar la actividad traductora, sobre todo en ámbitos muy especializados, en los que conseguir información no es fácil. Lo que no debemos olvidar es que el hecho de que el usuario de un determinado corpus sea también su creador es una enorme ventaja. De este modo, él mismo puede determinar sus necesidades, delimitar el ámbito de especialidad y utilizar los criterios estructurales que considere necesarios, siempre que sean coherentes, para crear un corpus fiable y útil. De todos modos, el traductor no debe olvidar que la consulta con especialistas sobre dicha parcela de la ciencia es indispensable.

Sin duda sería interesante que se hiciesen comparaciones entre los resultados que puede obtener un traductor profesional especializado al utilizar un corpus *ad hoc* creado por él mismo como recurso lingüístico y al utilizar otros recursos más habituales, como diccionarios u otras fuentes especializadas.

---

<sup>1</sup> Este trabajo forma parte del proyecto «Construcción eficiente de recursos lingüísticos multilingües» (INCITE08PXIB302179PR), financiado por la Consellería de Innovación e Industria de la Xunta de Galicia.

<sup>2</sup> En lo referente a las enfermedades neuromusculares, recomendamos la consulta de Sánchez Trigo (2006 y 2008) y Miquel Vergès y Sánchez Trigo (2008).

<sup>3</sup> El corpus MYOCOR constituye una fuente importante de documentación en francés y español sobre enfermedades neuromusculares. Está disponible a través del buscador MYOCOR en la página web de ASEM-Galicia (<http://www.asemgalicia.com/myocor/myocor.html>). Fue desarrollado en el marco del proyecto de investigación «Creación y explotación de recursos documentales sobre enfermedades neuromusculares» (PGIDIT 04SIN065E), liderado por ASEM-Galicia, y contó con la colaboración externa de la Universidade de Vigo y el Hospital del Meixoeiro (Vigo).

<sup>4</sup> La Federación Española de Enfermedades Neuromusculares traduce, gracias a un acuerdo con la AFM (Asociación Francesa contra las Miopatías), textos sobre enfermedades neuromusculares de francés a español, para favorecer la existencia de documentación en esta última lengua. Este trabajo de traducción, que está siendo realizado desde el año 1997 por la profesora Elena Sánchez Trigo y su equipo (Departamento de Traducción e Interpretación de la Universidade de Vigo) junto con médicos del Hospital del Meixoeiro (Vigo), ha permitido la publicación en castellano de varias monografías y un buen número de textos de diversos géneros sobre las enfermedades neuromusculares. Parte de estos textos pueden consultarse en las páginas web de ASEM (<http://www.asem-esp.org/>) y ASEM-Galicia (<http://www.asemgalicia.com/>).

## REFERENCIAS BIBLIOGRÁFICAS

Alcina, A., Soler, V. & Estellés, A. (2005). Internet como instrumento para la documentación en terminología y traducción. Hacia las plataformas de recursos



- electrónicos para el traductor especializado. In Soles Salvador, D. (Ed.), *La biblioteca de Babel. Documentarse para traducir*. Granada: Editorial Comares.
- Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (Eds.) *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi. 9-25.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8, 243-257.
- Bowker, L. & Pearson, J. (2002). *Working with specialized language. A practical guide to using corpora*. Londres y Nueva York: Routledge.
- Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta*, 43, 4, 631-651.
- Cabré, M. T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Ed. Antártida/Empúries.
- Corpas Pastor, G. (2001). Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada. *TRANS. Revista de traductología*. 5. 155-184.
- Corpas Pastor, G. (2004). La traducción de textos médicos especializados a través de recursos electrónicos y corpus virtuales. In L. González & P. Hernández (Eds.), *Las palabras del traductor. Actas del II Congreso Internacional «El español, lengua de traducción»*, 20 y 21 de mayo, Toledo 2004. Bruselas: Comisión Europea/ESLETRA. 137-164.
- Laviosa, S. (2002). *Corpus-based Translation Studies. Theory, Findings, Applications*. Amsterdam y Nueva York: Rodopi.
- Leech, G. (1991). The State of the Art in Corpus Linguistics. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. Londres: Longman, pp. 8-29.
- McEnery, T. et al. (2006). *Corpus-based language studies: an advanced resource book*. Londres: Routledge.
- Miquel Vergès, J. & Sánchez Trigo, E. (2008). El modelo social de la traducción y su aplicación a los buscadores especializados en salud en Internet: el buscador de enfermedades neuromusculares de ASEM (Marco teórico e implementación computacional). *Cadernos de Tradução*, vol. 2.
- Montalt, V. & González Davies, M. 2007. *Medical translation Step by Step*. Manchester (Reino Unido) & Kinderhook (EEUU): St Jerome Publishing.
- Recoder, M.J. & Cid, P. 2004. La documentación en la traducción especializada. In C. Gonzalo García & V. García Yebra (Eds.), *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco Libros.
- Sánchez Trigo, E. (2006). MYOCOR: creación y explotación de un corpus bilingüe (FR-ES) sobre enfermedades neuromusculares. *Confluências. Revista de Tradução Científica e Técnica*, vol. 4, 67-83.
- Sánchez Trigo, E. (2008). MYOCOR: création d'un corpus bilingue et d'un moteur de recherche sur les maladies neuromusculaires. In *Proceedings of the XVIII FIT World Congress*. Shanghai (China): FIT, 350-360.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Sinclair, J. (1996). *Preliminary recommendations on Corpus Typology*. EAG-TCWG-CTYP/P. Pisa: EAGLES.
- Sinclair, J. (2004). Corpus and Text - Basic Principles. En Wynne, M. (Ed.). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. Disponible en: <<http://ahds.ac.uk/linguistic-corpora>> [Consulta: 13/02/09].
- Zanettin, F. (2002). DIY Corpora: The WWW and the Translator. In B. Maia, J. Haller & M. Urlrych (Eds.), *Training the Language Services Provider for the New Millennium*. Oporto: Faculdade de Letras, Universidade do Porto. Disponible en: <<http://www.federicozanettin.net/DIYcorpora.htm>>. [Consulta: 03/03/09].