

MYOCOR: CREACIÓN Y EXPLOTACIÓN DE UN CORPUS BILINGÜE (FR-ES) SOBRE ENFERMEDADES NEUROMUSCULARES

ELENA SÁNCHEZ TRIGO

Departamento de Traducción e Lingüística, Universidade de Vigo, España

Resumen:

La evolución de las tecnologías de tratamiento de la información ha dado lugar a un conjunto creciente de trabajos relacionados con los corpus textuales. En el ámbito científico, el omnipresente dominio del inglés ha recibido mayor atención que otras lenguas en el planteamiento de diversos proyectos que, de modo general, adoptan una perspectiva terminológica.

En este artículo presento los criterios que estamos siguiendo para la elaboración y explotación de un corpus electrónico bilingüe (francés-español) relacionado con un subdominio de la traducción médica (enfermedades neuromusculares). La disponibilidad de textos en francés de máxima actualidad permite definir un acercamiento más próximo a las necesidades de la traducción profesional del francés al español.

La adopción de una perspectiva comunicativa, la multidisciplinariedad del subdominio seleccionado, la creación de herramientas específicas y su vinculación con un proyecto real de traducción permiten definir líneas de avance en un espacio que, como señala Habert (1999), reclama un esfuerzo específico a los investigadores.

Palabras clave:

Corpus electrónicos; Traducción textos médicos; Enfermedades neuromusculares; Francés; Español.

Resumo:

A evolução das tecnologias de tratamento da informação deu lugar a um conjunto crescente de trabalhos relacionados com os *corpora* textuais. No âmbito científico, o omnipresente domínio do inglês recebeu maior atenção do que outras línguas na

formulação de diversos projectos que, de maneira geral, adoptam uma perspectiva terminológica.

Neste artigo apresento os critérios que estamos a aplicar para a elaboração e exploração de um *corpus* electrónico bilingue (francês–espanhol) relacionado com um subdomínio da tradução médica (doenças neuromusculares). A disponibilidade de textos em francês da máxima actualidade permite definir uma abordagem mais próxima das necessidades da tradução profissional do francês para o espanhol.

A adopção de uma perspectiva comunicativa, a multidisciplinaridade do subdomínio seleccionado, a criação de ferramentas específicas e a sua vinculação com um projecto real de tradução permitem definir linhas de avanço num espaço que, como explica Habert (1999), reclama um esforço de especificidade aos pesquisadores.

Palavras-Chave:

Corpora electrónicos; Tradução textos médicos; Doenças neuromusculares; Francês; Espanhol.

Abstract:

The evolution of information processing technologies has led to a series of projects related to text corpora. In the scientific sphere, the omnipresent supremacy of English accounts for the fact that this language has been paid more attention than others in the elaboration of projects which, generally speaking, adopt a terminological standpoint.

In this paper I present the criteria we are currently applying for the elaboration and exploitation of an electronic bilingual corpus (French–Spanish) related to a medical translation subfield (neuromuscular diseases). The availability of highly contemporary French texts allows us to define an approach that is closer to the needs of professional translation from French into Spanish.

The choice of a communicative standpoint, the multidisciplinary nature of the subfield selected, the creation of specific tools and their connection with a real translation project allow us to define progress pathways in a space which, as explained by Habert (1999), claims a specific effort on behalf of researchers.

Keywords:

Electronic corpora; Translation of medical texts; Neuromuscular diseases; French; Spanish.

INTRODUCCIÓN¹

Si bien en un primer momento la creación y análisis de corpus no se orientaba de manera específica a la traducción, poco a poco ha ido reclamando la atención de los principales investigadores de este ámbito.

Los resultados que se fueron obteniendo a partir de los años noventa del siglo pasado han demostrado que se trata de una línea de trabajo que aporta resultados de especial interés para los profesionales de la mediación comunicativa². En 1999, Mona Baker señalaba que aún estábamos en una fase inicial pero, desde entonces, la evolución ha sido muy rápida. Cada vez son más los trabajos que se adscriben a esta línea y en la actualidad ya puede considerarse como uno de los principales paradigmas en la investigación sobre traducción. Se ha acuñado, incluso, una denominación específica, *Corpus Translation Studies* (CTS), para los estudios que adoptan estos enfoques.

Entre las causas que podrían justificar el interés suscitado por esta línea de investigación en traducción, es fundamental disponer de conocimientos textuales que permitan a traductores e intérpretes la producción de textos adecuados e inteligibles en las diferentes lenguas y culturas con las que trabajan.

En la actualidad, la posibilidad de acceder a numerosos textos a través de Internet y el desarrollo de herramientas informáticas específicas han facilitado la creación y análisis de corpus electrónicos.

Sin embargo, la utilidad real de un corpus depende «de los criterios que se aplican en el momento de selección y compilación de los textos» (Abaitua 2000, 2). Es determinante, asimismo, la metodología seguida para la explotación del mismo (análisis de los textos que lo integran, herramientas a las que puede dar lugar, etc.). No hay que olvidar que los estudios cuantitativos no son un fin en sí mismos, sino que deben integrarse en una investigación bien enfocada y de interés sustancial (Tymoczko 1998, 7).

En las páginas que siguen abordamos los criterios que nos han guiado en la creación y explotación de MYOCOR. Un corpus sobre enfermedades neuromusculares en el que estamos trabajando en estos momentos³.

¹ Este trabajo se integra en el marco de los siguientes proyectos de investigación: PGIDIGT04SIN065E (Xunta de Galicia), TIN2004-07246-C03-01 (Ministerio de Educación y Ciencia) y PGIDIGT05PXIC30501PN (Xunta de Galicia).

² Remitimos sobre este punto a Johansson y Oksefjell (1998) y Laviosa (1998).

³ Aspectos relacionados con las características de MYOCOR han sido abordados en algunas otras de nuestras publicaciones como, por ejemplo, Sánchez Trigo 2005a y 2005b.

1. CRITERIOS SEGUIDOS PARA LA ELABORACIÓN DE MYOCOR

Los autores que han analizado los problemas que afectan a la creación de los corpus de especialidad⁴, como es el caso que nos ocupa, coinciden en señalar que los principales aspectos a los que se debe atender son los siguientes: ámbito de especialidad, tamaño del corpus, cronología de los textos, modo de producción (escrito/oral), factualidad (textos de ficción o no ficción), calidad de los textos, autor y características del mismo (autoría individual o colectiva,...), elementos relativos a la publicación (tipo de publicación...), grado de tecnicidad, tipo de texto y receptores.

Veamos qué enfoque hemos ido adoptando para todos estos aspectos en el caso de MYOCOR. Empezaremos por abordar la delimitación del ámbito de especialidad seleccionado.

a) Un corpus especializado: enfermedades neuromusculares

El subdominio médico de las enfermedades neuromusculares está integrado por un amplio número de afecciones; los especialistas destacan hasta 150 formas de expresión diferentes. La mayor parte de ellas presenta un origen genético, aunque también existen algunas de origen idiopático.

A pesar de la variabilidad que las caracteriza, el efecto común en todas estas dolencias es la pérdida de capacidad funcional muscular.

Se trata de enfermedades crónicas, generalmente progresivas y de gravedad muy diversa. Pueden llegar a ser muy invalidantes e incluso provocar la muerte en cualquier momento de la vida, incluido el periodo neonatal o la infancia.

En relación con su incidencia, podemos indicar que está estimada en 1 de cada 1000 nacimientos. De acuerdo con estos datos, y a modo de ejemplo, podemos señalar que en el Estado español se calcula que existen unas 50 000 personas con enfermedades neuromusculares.

Si bien el número total de afectados es importante, la gran diversidad de expresión y de localización genética de estas afecciones hace que las múltiples entidades bien individualizadas que se han definido se incluyan en el ámbito de las conocidas como 'enfermedades raras'.

⁴ Ver, por ejemplo, Meyer y Mackintosh (1996), Browker (1996) y Pearson (1998). Asimismo, sobre los problemas generales para la constitución de un corpus remitimos a: Atkins, Clear y Oster (1992), Quirk (1992), Biber (1993), Hockey y Walker (1993), EAGLES (1994, 1996a y 1996b) o Sinclair (1996b).

Esta situación ha tenido consecuencias muy negativas para los enfermos (retraso en los diagnósticos, problemas de adaptación, etc.), que en estos momentos pretenden ser superados mediante la intervención desde diferentes ámbitos.

Precisamente, las variadas consecuencias que pueden tener estas enfermedades hacen que en este ámbito confluya un buen número de especialidades médicas (como la neurología, la miología, la cardiología, etc.). Se trata, además, de un ámbito en el que están presentes muchas otras áreas relacionadas con la salud en su sentido más amplio (como la fisioterapia, la ortopedia, la psicología, la nutrición, aspectos relacionados con la calidad de vida de los enfermos, etc.).

Esta multidisciplinariedad propia del subdominio seleccionado es, lógicamente, un elemento también característico del corpus. Los textos que integran MYOCOR presentan una interesante diversidad tanto desde el punto de vista temático (derivada de la convergencia de áreas diversas), como en relación con los géneros textuales representados (a causa de los diferentes tipos de situaciones comunicativas que se pueden producir). Volveremos sobre estos aspectos más adelante.

En nuestro caso, las tareas de traducción y revisión de textos que realizamos desde hace años en este ámbito⁵ nos permitieron conocer la riqueza y variedad del mismo. Este hecho es de especial interés para traductores, intérpretes y redactores científicos, al constituir esta área una muestra representativa de diversos subdominios de la medicina.

El corpus que presentamos se inscribe en un ámbito en el que existe, como hemos indicado, una demanda social, lo que permite establecer una conexión, que consideramos fundamental, entre la actividad académica e investigadora y la realidad social.

b) Un corpus bilingüe: francés-español

Si bien el inglés es la lengua que predomina en la comunicación científica, y por lo tanto también en medicina, el corpus que presentamos está integrado por un conjunto de textos en francés y en español⁶. Varias son las razones que justifican esta elección.

⁵ Desde el año 1997, nuestra actividad en esta área incluye la traducción y revisión de textos pertenecientes a géneros diversos, como artículos dirigidos a diferentes tipos de receptores, convenios o monografías. Esta actividad ha dado lugar, por ejemplo, a la publicación de las monografías: *Distrofia facioescapulohumeral* (Díaz y Rivière 2002), *Enfermedades neuromusculares* (V. AA. 2003) y *Distrofia miotónica de Steinert* (Jaeger y Rivière 2004). Hemos dirigido, asimismo, numerosos trabajos de fin de carrera de la Licenciatura de Traducción e Interpretación centrados en este ámbito en la Universidad de Vigo.

⁶ Si bien no se descarta la introducción de otras lenguas en el futuro.

El corpus que presentamos se inscribe en un ámbito en el que existe, como hemos indicado, una demanda social, lo que permite establecer una conexión, que consideramos fundamental, entre la actividad académica e investigadora y la realidad social

En relación con el francés, hemos encontrado un ámbito científico en el que existe un importante volumen de textos producidos en esta lengua. Este hecho es el resultado de varios factores como son: el trabajo de primera línea realizado por los investigadores franceses en este campo; la existencia, en Francia, de asociaciones de enfermos con un desarrollo muy importante y un amplio reconocimiento social (como es el caso de la AFM 'Association Française contre les Myopathies')⁷; y el establecimiento por parte del Estado francés de una serie de medidas específicas de apoyo a estos colectivos, medidas que constituyen un modelo a imitar por colectivos de otros países.

En relación con la selección del castellano, los motivos tienen que ver, en parte, con los que acabamos de indicar. En España, la situación es diferente a la descrita en Francia, pero, precisamente por ello, en estos momentos existe una demanda por parte de los enfermos neuromusculares para difundir el conocimiento sobre estas enfermedades. Se considera necesario favorecer un diagnóstico precoz y abordar medidas que mejoren la calidad de vida.

Dar respuesta a esta demanda es uno de los objetivos prioritarios de ASEM (Federación Española de Enfermedades Neuromusculares)⁸. Entre las actividades que esta asociación lleva a cabo para dar a conocer estas enfermedades, se encuentra la publicación y traducción de textos del francés al castellano, gracias a un acuerdo que mantiene con la AFM.

Entendemos que la selección de lenguas de MYOCOR confiere al corpus un elemento de interés. Es muy significativo, por ejemplo, en relación con el francés, la escasa representación que en los corpus presentan áreas como el francés científico (Habert 1999, 6). Desde este punto de vista MYOCOR es innovador, puesto que reúne textos representativos de áreas consideradas como deficitarias.

c) Un corpus representativo: variedad de géneros textuales y actualidad

La selección textual de MYOCOR se hace utilizando como punto de partida un área temática, y tiene como objetivo incluir textos representativos desde el punto de vista de la actualidad científica y social del ámbito delimitado.

⁷ Se trata de una importante y conocida asociación, creada en 1958 (<<http://www.afm-france.org/>>). Cuenta con un gran presupuesto que le permite financiar de manera total o parcial una serie de centros de investigación de gran prestigio en el ámbito científico. La AFM publica, de manera directa o a través de sus centros de investigación, un número importante de textos caracterizados por su interés, sometidos a revisiones de calidad y actualizados de forma permanente.

⁸ Se trata de una asociación homóloga de la AFM, aunque más joven y con menos recursos (<<http://www.asem-esp.org/>>).

En un corpus esta selección debe realizarse de acuerdo con unos criterios externos explícitos, y por ello hemos indicado más arriba qué son estas enfermedades neuromusculares, con la finalidad de delimitar el subdominio objeto de estudio y destacar el gran número de áreas con las que están directamente relacionadas.

Los criterios que han guiado la selección de los textos que se incluyen en MYOCOR tienen como finalidad reflejar las prácticas propias del ámbito en el que se centra y las diferentes situaciones de comunicación particulares.

Se busca representar una comunidad de discurso (Swales 1990), es decir, un grupo formado en torno a temas de interés común y caracterizado por: a) existencia de una finalidad común aceptada por todos los miembros de la comunidad, b) mecanismos de interacción entre los miembros, c) mecanismos de participación que garantizan la circulación de la información, d) existencia y uso de uno o varios géneros, e) adquisición de un lenguaje específico y f) existencia de un grupo de miembros.

El criterio de selección textual ha tenido en cuenta, asimismo, el nuevo enfoque sobre el concepto de salud introducido por la OMS en 2001 (OMS 2001, 21)⁹. Por este motivo, en el corpus se incluyen textos producidos por diferentes tipos de sujetos e instituciones implicadas en el ámbito de las enfermedades neuromusculares. Se incluyen, por tanto, textos elaborados por médicos o profesionales sanitarios con diferentes niveles de especialización (biólogos, enfermeros, fisioterapeutas, etc.), por asociaciones de enfermos, por empresas (farmacéuticas, ortopédicas, etc.), por investigadores y por diversas instituciones.

En consecuencia, los textos seleccionados constituyen una muestra representativa de diferentes géneros. Se incluyen tanto artículos procedentes de revistas científicas, actas de congresos, pósteres o monografías, como textos sobre mejoras en las condiciones de vida, maneras de ayudar a aceptar el diagnóstico o, por ejemplo, textos de naturaleza administrativa relativos a este subdominio específico¹⁰.

De forma análoga, los receptores a quienes se dirigen son también de naturaleza diversa, médicos, investigadores, enfermos, familias o, por ejemplo, instituciones¹¹. Los

⁹ Se trata de una concepción sobre la enfermedad que se aleja de algunos planteamientos tradicionales. Esta nueva visión se basa en la integración de dos modelos antagónicos: el «modelo médico» y el «modelo social». El primero, derivado de una visión más tradicional, concibe la salud como un problema individual en relación directa con una enfermedad que requiere un tratamiento médico concreto. El modelo social aporta un enfoque más amplio, y remite a situaciones que, al menos parcialmente, están originadas por el entorno social, por lo que implican a toda la sociedad. Ambos modelos tienen un carácter complementario.

¹⁰ Para asegurar el criterio de calidad de los textos se han excluido páginas personales.

¹¹ La propia AFM clasifica sus textos en función de los receptores a los que se destinan. Así, por ejemplo, bajo la denominación genérica de *Myoline*, realiza una serie de publicaciones que se clasifican en función de sus destinatarios: a) profesionales del ámbito médico y social (médicos, personal de enfermería, fisioterapeutas, psicólogos, etc.): *Bulletin Myoline*, *Les*

destinatarios se corresponden con un amplio abanico que incluye no sólo a los que acabamos de citar sino también a la sociedad en su conjunto.

Desde el punto de vista cronológico, MYOCOR está integrado por textos producidos entre 1995 y 2006. Los límites temporales vienen determinados por la voluntad de trabajar con textos actuales, puesto que la investigación y la situación de los enfermos evolucionan con rapidez. Este criterio de actualidad nos lleva a pretender crear un corpus dinámico en el que se integren de manera continuada nuevos textos.

Este criterio de representatividad nos ha llevado a incluir textos de diversas áreas geográficas. Francia, Canadá, Bélgica y Suiza¹², para los textos en francés, y España, Hispanoamérica y Estados Unidos¹³, para los textos en español. Por lo tanto, si bien los textos que integran el corpus proceden en gran medida de las asociaciones mencionadas en apartados precedentes (AFM y ASEM), éstas no son, sin embargo, las únicas fuentes utilizadas¹⁴.

MYOCOR está integrado por textos producidos entre 1995 y 2006 (...)

El objetivo es elaborar un corpus bilingüe con textos redactados originalmente en francés y español sobre enfermedades neuromusculares. Sin embargo, en el corpus se incluye también un conjunto de textos traducidos del francés al español

El objetivo es elaborar un corpus bilingüe con textos redactados originalmente en francés y español sobre enfermedades neuromusculares. Sin embargo, en el corpus se incluye también un conjunto de textos traducidos del francés al español, lo que permite la creación de un subcorpus paralelo y otro comparable, con textos en español traducidos de otras lenguas, fundamentalmente el inglés.

Se crea, así, un corpus central y dos corpus complementarios, con la finalidad de poder analizar la denominada 'lengua de la traducción' (Baker 1998). Es decir, estudiar las soluciones adoptadas en relación con aspectos neológicos, la influencia del texto original en la traducción, o estudiar si existen patrones utilizados de manera más habitual en estos textos.

Los textos que se incluyen en MYOCOR son escritos y completos (no fragmentos textuales, por no presentar interés desde el punto de vista de la traducción¹⁵), proceden de fuentes diversas y se encuentran tanto en papel como en formato electrónico.

Fiches Techniques, Les Comptes Rendus Flash, Les Monographies; b) entorno familiar de los enfermos: *Les Repères, Les tracts maladies neuromusculaires, Les avancées de la recherche;* c) Para ambos colectivos: *VLM*.

¹² El diferente desarrollo que caracteriza a los países francófonos hace que no se hayan encontrado fuentes en otros países que forman parte del área de la francofonía.

¹³ Se han incluido textos de la MDA (*Muscular Dystrophy Association*) en español, por su gran influencia en el ámbito hispanófono en EE. UU.

¹⁴ Se incluyen también textos de la *Revista de Neurología*, de la Societat Catalana de Neurologia, de asociaciones como CMT (*Charcot-Marie-Tooth*) o *Duchenne Parent Projet*, entre otros.

¹⁵ Por este mismo motivo, y siguiendo los criterios de Dubuc y Lauriston (1997, 85), tampoco se incluyen diccionarios o bases de datos terminológicas.

Para que los textos estén disponibles y puedan ser estudiados de manera eficaz es necesario seguir una serie de pasos. Se lleva a cabo su incorporación en formato gráfico íntegro, lo que permite conservar el formato original y todos los elementos paratextuales tan frecuentes en estos géneros textuales. Asimismo, se obtiene un fichero de texto plano que permite su posterior tratamiento. Los textos se integran siguiendo las recomendaciones de TEI (*Text Encoding Initiative*), por lo que se anotan en XML¹⁶. Esto facilita su intercambio, el tratamiento automatizado y una posible reutilización del corpus para otros fines. Además, los textos también se archivan en soporte físico para su consulta inmediata.

Una vez establecido el corpus, su explotación permite llevar a cabo una serie de análisis de tipo cuantitativo y cualitativo, así como la elaboración de una serie de recursos. En el apartado que sigue nos ocupamos de estos aspectos.

2. CRITERIOS SEGUIDOS PARA LA EXPLOTACIÓN DEL CORPUS

La explotación de MYOCOR se está desarrollando en torno a dos líneas principales. Por un lado, la creación de una herramienta de consulta documental y, por otro, el análisis de las convenciones textuales de los textos que integran el corpus.

a) Creación de una herramienta de consulta documental

MYOCOR ha permitido la elaboración de una herramienta de consulta en línea para obtener información sobre las enfermedades neuromusculares. Dicha herramienta realiza una búsqueda directa sobre el corpus, tanto en castellano como en francés.

La aplicación se ha desarrollado a partir del uso de software libre¹⁷ y en estos momentos se encuentra en fase de prueba en su versión en castellano. Se ha diseñado de manera que ofrezca la posibilidad de realizar dos tipos de búsqueda.

¹⁶ MYOCOR no va a presentar anotaciones lingüísticas sistemáticas. Sí presentará anotaciones estructurales y, probablemente, lematización y alineación, así como anotaciones extralingüísticas (datos de la cabecera del corpus).

¹⁷ Se ha recurrido, con adaptaciones, al sistema Apache (<<http://www.apache.org/>>). El esquema de la aplicación es el siguiente: *Lucene* indexa los documentos y realiza las búsquedas solicitadas por los usuarios; *Xerces*: recupera la información de los documentos; *Tomcat* y las clases que contienen la lógica de la aplicación. Opcionalmente se puede utilizar Apache conjuntamente con *Tomcat* como servidor de aplicaciones para optimizar el trabajo del servidor *web*. Un navegador *web* permite el acceso a la aplicación y la interacción del usuario con la aplicación para realizar las búsquedas que desee.

Buscador de ASEM Galicia

Palabras clave:

Opciones:

<input type="checkbox"/> Introducción	<input type="checkbox"/> Enfermedad	<input type="checkbox"/> Descripción enfermedad
<input type="checkbox"/> Causas	<input type="checkbox"/> Síntomas	<input type="checkbox"/> Evolución
<input type="checkbox"/> Transmisión	<input type="checkbox"/> Diagnóstico	<input type="checkbox"/> Tratamiento
<input type="checkbox"/> Investigación	<input type="checkbox"/> Vida diaria	<input type="checkbox"/> Estudio
<input type="checkbox"/> Contacto		

Enfermedad:

Gráfico I. Pantalla inicial del buscador

La primera de estas búsquedas (Gráfico I: parte superior de la pantalla, 'Palabras clave' y 'Opciones') permite acceder a la información a través de palabras clave, con la finalidad de obtener una información organizada. El usuario introduce la denominación de una enfermedad y realiza una selección de la información que quiere obtener en relación con la totalidad o parte de aspectos específicos como: descripción, causas, síntomas, evolución, transmisión, diagnóstico, tratamiento, investigación y vida diaria.

Así, si un usuario desea obtener información sobre el diagnóstico de la Miastenia gravis, tras introducir el nombre de la enfermedad en 'palabras clave' y seleccionar 'diagnóstico' en 'opciones', obtendrá el siguiente resultado:



Resultados de la búsqueda

Palabras clave:	Miastenia gravis
Diagnóstico:	Miastenia grave Los Hechos Acerca de la Miastenia Grave Los Hechos Sobre la Esclerosis Lateral Amiotrófica (ALS) Los Hechos Acerca de las Miopatias
Documentos:	 Los Hechos Acerca de la Miastenia Grave  Miastenia grave  Los Hechos Sobre la Esclerosis Lateral Amiotrófica (ALS)  Los Hechos Acerca de las Miopatias

Gráfico II. Ejemplo de una búsqueda por palabras clave

En la pantalla aparecen enlaces para acceder a los fragmentos de los textos en los que se presentan la información seleccionada. La aplicación permite acceder, asimismo, a los textos en formato gráfico (apartado 'Documentos'), por lo que ofrece la posibilidad de consultar esquemas o gráficos, tan frecuentes en estos textos.

La segunda forma de búsqueda (Gráfico I: parte inferior de la pantalla, 'Enfermedad') sirve para obtener toda la información existente en el corpus sobre una enfermedad determinada. El usuario introduce la denominación de una enfermedad neuromuscular, o la selecciona de entre las que se ofrecen en un listado desplegable.

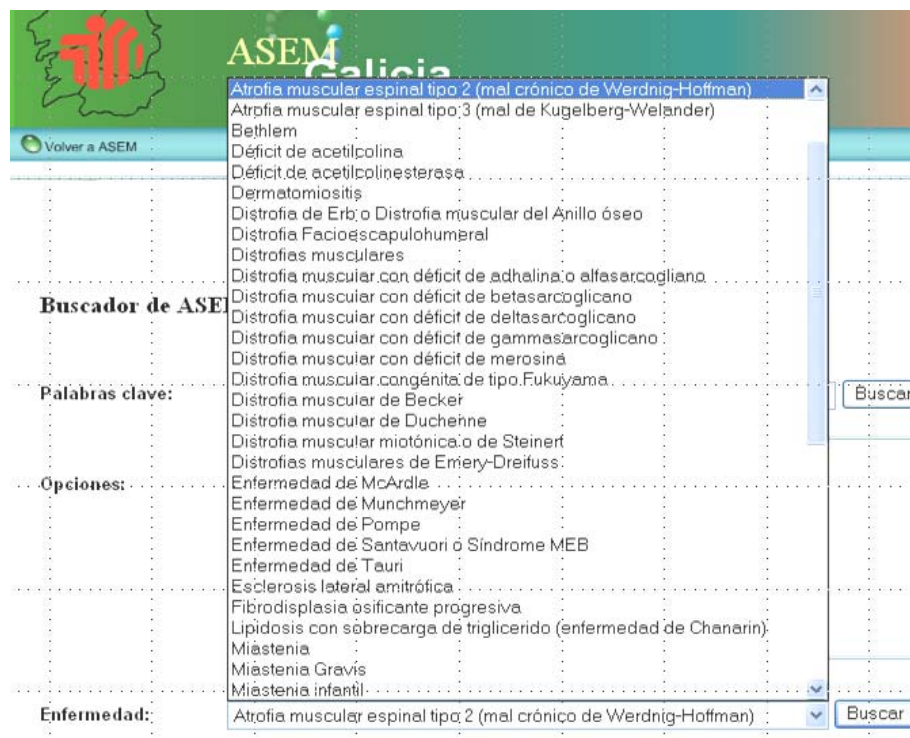


Gráfico III. Vista parcial del listado de enfermedades

En la pantalla aparece el título de todos aquellos textos en los que se hace referencia a dicha enfermedad. Pinchando sobre cada uno de ellos se accede al contenido del documento.

Como apuntábamos más arriba la herramienta está en estos momentos en fase de prueba interna. Estimamos que en el último trimestre del 2006 estará a disposición de los usuarios a través de la página de ASEM Galicia (<<http://www.asemgalicia.com/>>) en su versión en castellano, y a lo largo del 2007 también se podrá acceder a su versión en línea en francés.

Consideramos que MYOCOR será de utilidad para traductores, intérpretes o redactores científicos dado que facilita las tareas de obtención de documentación fiable y organizada sobre enfermedades neuromusculares. Queremos destacar que esta herramienta será, asimismo, útil para otros colectivos, como profesionales de la salud, afectados por estas enfermedades y personas que conforman su entorno.

b) Análisis de las convenciones textuales

Como indicábamos al inicio de este artículo, la competencia textual es una de las que conforman la denominada competencia traductora. Si bien sabemos que cada texto es único, por ser el resultado de una situación comunicativa determinada e irreplicable,

también conocemos que los textos producidos en situaciones comunicativas similares presentan una serie de características compartidas y reiteradas.

Los textos que integran MYOCOR están siendo analizados, asimismo, desde un punto de vista microtextual, con la finalidad de obtener datos sobre las convenciones que los caracterizan.

Para ello se utiliza, con adaptaciones, el modelo comunicativo propuesto por la Estilística Funcional, que establece las siguientes categorías o parámetros: categorías lineales (cadena temática y cadena lógica), de campo (modalidad, tiempo y espacio) y estructurales. Dichos parámetros se expresan en los textos mediante una serie de marcadores caracterizados por: a) no identificarse con una categoría lingüística concreta; b) estar determinados por su función semántico-lógica; c) presentar una serie de cambios de naturaleza estilístico-funcional derivados de la situación comunicativa, de la intención y del tipo de texto¹⁸.

El estudio de las diferentes categorías aporta datos sobre las características de los textos y las estrategias comunicativas propias de cada situación, lo que forma parte de la competencia textual que los traductores profesionales necesitan. Veamos algunos ejemplos.

Los marcadores de la cadena temática nos permiten identificar las diferentes denominaciones utilizadas para el tratamiento de un tema. Entre las cadenas temáticas más frecuentes podemos citar las relativas a: *paciente, patología, terapia* o, por ejemplo, *consecuencias de la terapia*.

Con el análisis de los marcadores se obtiene una relación de las formas a las que se recurre en los textos para hacer referencia a las cadenas indicadas. Además, el hecho de que la identificación de dichas cadenas venga determinada por aspectos tanto lingüísticos como extralingüísticos nos permite obtener resultados que otros enfoques no tomarían en cuenta.

Por ejemplo, la delimitación de los marcadores de la cadena temática *paciente* obliga a conocer las especiales características de las enfermedades neuromusculares, puesto que la naturaleza de los pacientes en estos casos es diferente de la que podría ser en otros casos.

Así, el origen genético de las afecciones que nos ocupan hace que la existencia de un enfermo en la familia lleve a convertir en pacientes a personas no afectadas por la enfermedad de manera evidente, ya que se trata de enfermedades que afectan al grupo familiar (pueden ser transmisores y desconocer ellos mismos su afección).

¹⁸ Sobre las propuestas de la Estilística funcional remitimos a Sánchez Trigo (2003) en donde se presentan de manera más detallada estas cuestiones y se ofrece bibliografía específica.

Términos o construcciones que en otros textos no se considerarían como propios de la cadena temática *paciente* aquí sí lo serían, e incluso podrían generar subcadenas específicas. Esto nos lleva a incluir como marcadores de dicha cadena las referencias a enfermos con síntomas graves, enfermos con síntomas leves, enfermos con manifestación clínica, transmisores de la enfermedad o familiares no afectados (cónyuge, hermanos, parientes...), que se expresan en los textos por medio de diferentes marcadores.

El análisis de la modalidad nos lleva a analizar cómo se expresa en los textos la seguridad, la inseguridad, la credibilidad, el interés, la convicción, el desacuerdo o la ratificación.

En castellano encontramos marcadores como: 'el tratamiento que veníamos utilizando como idóneo', 'la modificación básica realizada en el protocolo', 'el resultado es satisfactorio', o 'el tratamiento más adecuado'.

El análisis textual
propuesto busca profundizar
en todos aquellos aspectos
que son relevantes para
la traducción.

En francés, podemos citar como ejemplo: *'l'imbroglia des maladies distales, plus intéressante encore, l'effet est particulièrement efficace, ce sont des données fiables, o il est difficile de ne pas être d'accord avec cette étude.*

De este modo, podemos identificar la subjetividad del autor en el entramado textual, lo que constituye un índice sobre su actitud y valoraciones.

El estudio de las categorías de tiempo y espacio real y conceptual permite obtener conclusiones sobre la manera en que estas coordenadas están presentes en los textos y cómo estos se imbrican en el universo textual del que forman parte.

Así, junto a los marcadores directos del tiempo real ('desde los años 70', 'en la actualidad', *depuis deux ans, des nos jours*), encontramos otros, como los datos culturales o nombres propios, que implícitamente aluden a un momento determinado, como es el caso de: 'la investigación de Grüber', 'de acuerdo con Griffino', *'l'intervention du type Bristw-Latarjet, Casady et Nielsen o les travaux de Mac Neill.*

La importancia de los aspectos comunicativos a los que permite prestar atención el modelo propuesto por la Estilística Funcional, y que he esbozado en los casos a los que acabo de referirme, se pone de manifiesto también en las demás categorías textuales.

El estudio de los marcadores se lleva a cabo, en un primer momento, por géneros, orígenes geográficos y lenguas. En una fase posterior, se contrastarán los datos obtenidos, con el objetivo de proporcionar una caracterización genérica que sea de utilidad desde el punto de vista de la traducción del francés al español.

De las diferentes las herramientas informáticas a las que se puede recurrir para realizar estos análisis¹⁹ hemos decidido utilizar el *WordSmith Tools*²⁰ por su actual difusión en el ámbito académico y su flexibilidad.

CONCLUSIONES

En este trabajo hemos presentado una de las líneas de investigación que estamos desarrollando en estos momentos. Se trata de una investigación innovadora por varias razones²¹.

Tanto el área de las enfermedades neuromusculares, como las lenguas consideradas para el corpus MYOCOR (francés y español), constituyen elementos novedosos, en un ámbito en el que el inglés, por sí sólo o como centro de estudios multilingües, ha centrado la mayor parte de los esfuerzos de los investigadores.

Ambos aspectos (el subdominio seleccionado y las lenguas de trabajo) responden, además, a un ámbito en el que existe actualmente una demanda de traducción, lo que es especialmente interesante para contrastar la relevancia de las conclusiones obtenidas para el trabajo profesional.

Los recursos que se están creando constituyen una fuente de documentación multilingüe sobre enfermedades neuromusculares, así como sobre las diferentes especialidades médicas y otras áreas de la salud implicadas en el diagnóstico o tratamiento de estas afecciones.

¹⁹ Otros programas comerciales también muy difundidos son *WordCruncher* (*Wordcruncher Publishing Technologies*), *Oxford Concordancing Program* o *MicroConcord* (estos dos últimos son versiones anteriores de *WordSmith Tools*). Frente a estos, existen algunos que se distribuyen de manera gratuita como *TACT* (Departamento de Computing in the Humanities and Social Sciences, Universidad de Toronto), *FreeText Browser* (Universidad de Michigan) o *Conc* (Summer Institute of Linguistics). Podemos citar, asimismo, *O Corpógrafo*, una herramienta de análisis y tratamiento creado por Linguateca (<<http://poloclup.linguateca.pt/cgi-bin/corpografo>>).

²⁰ Se trata de un conjunto integrado de programas diseñado por Mike Scott en Liverpool, con la finalidad de poder analizar el comportamiento de las palabras en los textos. Dispone de una serie de utilidades que sirven para generar y manipular listas de palabras, concordancias y colocaciones. Scott define las *WordSmith Tools* como herramientas de uso general, es decir, como un conjunto de utilidades probadas para realizar de modo eficiente un conjunto de tareas que pueden ser útiles para fines muy diversos.

²¹ En nuestro país existen otros trabajos sobre corpus pero desde aproximaciones diferentes al enfoque presentado en este artículo. Se pueden destacar, entre otras, las propuestas de aplicación para la formación de traductores especializados realizadas por Corpas (2004).

El análisis textual propuesto busca profundizar en todos aquellos aspectos que son relevantes para la traducción. Por ello se aborda un análisis múltiple que no sólo se ocupa de aspectos terminológicos, sino que, partiendo del modelo de análisis de Estilística Funcional, se exploran los rasgos textuales que definen su carácter convencional y los factores determinantes de las variaciones relevantes (situación comunicativa, ámbito geográfico, etc.).

Junto a las tareas de explotación del corpus que en estos momentos estamos desarrollando, en el futuro pretendemos crear otras herramientas para la traducción especializada (elaboración de una base de datos terminológica, glosarios, trabajo en programas de traducción asistida que utilicen memorias de traducción, elaboración de programas didácticos basados en documentos reales...), así como mantener la actualización y representatividad del corpus que hemos creado. ■

5. Referencias bibliográficas

- Abaitua, J. 2000. "Tratamiento de corpora bilingües". [En línea]. Fundación Duques de Soria, Seminario «La ingeniería lingüística», Soria, 17-21 de julio de 2000. Consultado el 12 de enero de 2005. Disponible en: <<http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/soria00.htm>>.
- Atkins, B., Clear, J. y N. Ostler. "Corpus Design Criteria", *Literary and Linguistic Computing* 7, n.º 1 (1992): 1-16.
- Baker, M. "The role of corpora in investigating the linguistic behaviour of professional translators", *International Journal of Corpus Linguistics* 4, n.º 2 (1999): 281-298.
- Bowker, L. "Towards a Corpus-Based Approach to Terminography", *Terminography* 3, n.º1 (1996): 27-52.
- Biber, D. "Representativeness in Corpus Design", *Literary and Linguistic Computing* 8, n.º 4 (1993): 243-257.
- Corpas, GL. "Localización de recursos y compilación de corpus via Internet: aplicaciones para la didáctica de la traducción médica". En *Manual de documentación y terminología para la traducción especializada*, eds. C. Gonzalo García y V. García Yebra, 223-258. Madrid: Arco libros, 2004.
- Díaz Cl. y A. Rivière. *Distrofia Muscular Facioescapulohumeral* (trad. esp. de E. Sánchez Trigo), Vigo: ASEM, 2002.
- Dubuc, R. y A. Lauriston. "Terms and Contexts". In *Basic Aspects of Terminology Management*, Amsterdam: John Benjamins, 1997.
- EAGLES. *Corpus Typology: A Framework for Classification*. Informe interno N. 2.1. preparado por J. M. Sinclair, Universidad de Birmingham, Corpus Linguistic Group, 1994.
- EAGLES. *Text Corpora Working Group Reading Guide*. Documento EAGLES, EAG-TCWG-FR-2, 1996a.
- EAGLES. 1996b, ver Sinclair 1996.

Habert, B. 1999. "Un corpus clé pour le français actuel". [En línea]. Consultado el 7 de marzo de 2004. Disponible en <<http://www.biomath.jussieu.fr/CLEF/PresentationCorpusClef.pdf>>. [En la actualidad es imposible la consulta].

Hockey, S. y D. Walker. "Developing effective resources for research on texts: collecting texts, tagging texts, cataloguing texts, using texts, and putting texts in context", *Literary and Linguistic Computing* 8, n.º 4 (1993): 235-242.

Jaeger, C. y A. Rivière. *Distrofia miotónica de Steinert* (trad. esp. de E. Sánchez Trigo), Vigo: ASEM, 2004.

Johansson, S. y S. Oksefjell. *Corpora and Cross-Linguistic Research. Theory, Method and Case Studies*, Amsterdam/Atlanta: Rodopi, 1998.

Laviosa, S. (ed). 1998. "L'Approche basée sur le corpus/The Corpus-based Approach", *Meta* 43, n.º 4, Special Issue. [En línea]. Consultado el 26 de abril de 2006. Disponible en <<http://www.erudit.org/revue/meta/1998/v43/n4/index.html>>.

Meyer, I. y K. Mackintosh. "The Corpus from a Terminographer's Viewpoint", *International Journal of Corpus Linguistics* 1, n.º 2, (1996): 257-285.

OMS. 2001. CIF. Introduction. [En línea]. Consultado el 26 de abril de 2006. Disponible en <<http://whqlibdoc.who.int/publications/2001/9243545426.pdf>>.

Pearson, J. *Terms in Context*, Amsterdam/Filadelfia: John Benjamins, 1998.

Quirk, R. "On Corpus Principles and Design". En *Directions in Corpus Linguistics*, ed. J. Svartvik, 457-469, Berlín/Nueva York: Mouton de Gruyter, 1992.

Sánchez Trigo, E. "Aproximación traductológica al análisis de corpus para el estudio de las convenciones textuales". En *Panorama actual de la investigación en traducción e interpretación*, ed. Emilio Arjonilla, vol. I, 113-127, Granada: Atrio, 2003. (2.ª ed. 2004).

Sánchez Trigo, E. "Le corpus électronique Myocor : Application des nouvelles technologies à la traduction des textes médicaux". En *Actes du XVIIe Congrès mondial Fédération Internationale des Traducteurs*, ed. Leena Salmi & Kaisa Koskinen, 155-157, Tampere: FIT, 2005a.

Sánchez Trigo, E. "Creación y explotación de recursos documentales sobre enfermedades neuromusculares". [En línea]. En *Libro de ponencias XXII Congreso de la Federación ASEM. Presente y Futuro de las enfermedades neuromusculares*, 85-94, Vigo: ASEM, 2005b. Consultado el 26 de abril de 2006. Disponible en <http://www.asemgalicia.com/documentos/XXII_Congreso_ASEM_2005_Libro_de_PONENCIAS-1.pdf>.

Sinclair, J. *Preliminary Recommendations on Corpus Typology*, Documento EAGLES, EAG-TCWG-FR-2, 1996.

Swales, J. M. *Genre Analysis*, Cambridge: Cambridge University Press, 1990.

Tymoczko, M. "Computerized Corpora and the Future of Translation Studies", *Meta* 43, n.º. 4, (1998), Special Issue. [En línea]. Consultado el 26 de noviembre de 2005. Disponible en <<http://www.erudit.org/revue/meta/1998/v43/n4/index.html>>.

W. AA. *Enfermedades Neuromusculares* (trad. esp. de E. Sánchez Trigo), Vigo: ASEM, 2003.