

Corpus especializados como recurso para la traducción: análisis de los marcadores de la cadena temática en artículos científicos sobre enfermedades neuromusculares en pediatría

*Specialized corpora as a translation resource: analysis of
the thematic chains and its textual markers in scientific
papers on neuromuscular diseases in pediatrics*

María Magdalena Vila Barbosa

Universidad de Vigo
España

ONOMÁZEIN 27 (junio de 2013): 78-100



María Magdalena Vila Barbosa:

Departamento de Traducción y Lingüística - Grupo de Investigación COLE (Compiladores y Lenguaje),
Universidad de Vigo

Correo electrónico: mariamagdavilabarbosa@uvigo.es / Avenida de Santiago, 61, 6°C. 3200, Ourense. España

Fecha de recepción: diciembre de 2011

Fecha de aceptación: abril de 2013

Resumen

El presente trabajo gira en torno a dos ejes principales de interés para la traducción: la explotación de corpus comparables multilingües y el análisis de la relación entre una categoría textual y sus manifestaciones lingüísticas en los textos especializados. En la primera parte, se realiza un repaso teórico de algunas cuestiones que giran en torno al uso de corpus especializados en la traducción, no solo para satisfacer las necesidades terminográficas puntuales, sino para explorar nuevas posibilidades como las

propuestas desde la estilística funcional para el estudio de los convencionalismos propios de los textos especializados. En la segunda parte, repasaremos las principales fases diseñadas para la recopilación y explotación del corpus sobre enfermedades neuromusculares en pediatría: características y análisis del corpus, extracción automática de los contextos definitorios y metodología empleada para el análisis de las cadenas temáticas y sus marcadores.

Palabras clave: traducción; corpus especializado; estilística funcional; cadena temática.

Abstract

This article deals with two main topics of interest for Translation Studies: the exploitation of multilingual comparable corpora and the analysis of the relationship between a text category and its linguistic manifestation in specialized texts. In the first part of this study, we make a review of some theoretical issues related to the use of specialized corpora in translation practice. The aim in this part is not only to satisfy punctual terminological needs, but also

to explore new possibilities, like those proposed by Functional Stylistics for the study of conventionalisms in specialized texts. In the second part, we describe some of the steps involved in the process of compilation and exploitation of a particular specialized corpus, that of neuromuscular diseases in pediatrics. We thereby study corpus characteristics and analysis, automatic extraction of definitional contexts, and analysis of the thematic chains and its textual markers.

Keywords: translation; specialized corpora; functional stylistic; thematic chain.

1. Introducción¹

La lingüística de corpus se ha conformado como un paradigma de investigación flexible y adaptable, con una metodología de análisis que se puede aplicar a cualquier disciplina lingüística (Teubert, 2005: 49). De hecho, actualmente casi la totalidad de los trabajos de investigación en terminología apuestan por un diálogo con la lingüística de corpus, de tal forma que la terminología basada en corpus hoy en día no es la excepción, sino la regla en los estudios dentro de esta área (Tagnin y Guimarães, 2004: 316).

En una primera fase, esos trabajos tenían como principal objeto el estudio de los términos en su entorno natural, el texto, para la elaboración de glosarios. Sin embargo, hoy en día se ha abierto una nueva senda, propuesta por la llamada terminología textual, en la que los investigadores se preocupan por la descripción del lenguaje utilizado en un área de especialidad determinada. Con la observación extensiva en grandes corpus de textos especializados, es posible identificar patrones de uso, fraseologías, variaciones terminológicas, frecuencia de palabras conceptualmente relevantes, características cohesivas de los textos especializados, macro- y microestructuras, entre otros rasgos.

Superando los enfoques tradicionales, estas nuevas propuestas invitan a traspasar la frontera de lo rigurosamente terminológico para ocuparse también de la estructura textual y el *modus dicendi* de una determinada área del conocimiento, considerando, además de los términos, otros elementos integrantes que pueden actuar como mecanismos cohesivos. Leičik (1993: 97) apunta que es posible observar un texto especializado desde dos perspectivas: partir del término para observar el texto, o bien partir

del texto para observar los términos. Dichos términos no son sino una de las características del texto especializado; así siendo, el listado de los términos de un área del conocimiento no es suficiente para realizar un reconocimiento terminológico profundo de esa área de especialidad. En el estudio terminológico merecen atención no solamente los términos y los elementos que los rodean, sino también todos aquellos elementos que componen el texto y que nos permiten abordarlo como un todo (Leičik, 2002) y como un objeto multifacético y multidimensional (Finatto, 2007).

Los resultados de este tipo de estudios que observan el texto especializado en su globalidad son de especial interés para el traductor, quien tiene unas necesidades específicas que no se satisfacen con conocer el sentido de un término. Este necesita informaciones pragmáticas y datos sobre frecuencia de uso, variantes, coocurrencias, colocaciones, etc. Es importante conocer los contextos en los que ese término es utilizado para dilucidar si es o no el más adecuado en el correspondiente contexto del texto traducido.

Por otro lado, en la misma medida en que existe un cierto convencionalismo en la lengua común que caracteriza la forma peculiar de expresión de una comunidad lingüística, en los lenguajes de especialidad se reconocen diversos convencionalismos que caracterizan la forma de decir las cosas y de organizar la información en una determinada área del conocimiento (Tagnin, 2005). El acceso a esta información le permitirá al traductor producir textos de calidad, es decir, textos que no les resulten extraños a los lectores profesionales y que encajen perfectamente con los textos originales producidos en esa área. En este sentido, el uso de corpus especializados de-

1 Proyecto dirigido por la Dra. Elena Sánchez Trigo (Universidad de Vigo), cofinanciado por la Xunta de Galicia a través del Plan I2C (2011-2015), por la *Consellería de Cultura, Educación e Ordenación Universitaria* (Xunta de Galicia - España) a través de las ayudas para la *Consolidación y Estructuración de Unidades de Investigación Competitivas del Sistema Universitario de Galicia* (ref. CN2012/317) y por la Universidad de Vigo a través de las subvenciones para estadias en centros de investigación (2012).

sempeñará un papel fundamental:

... a Lingüística de Corpus tem exercido um papel incontestável na evolução dos estudos tradutológicos e terminológicos, por insistir na noção de linguagem enquanto objeto de pesquisa cujos fenômenos só podem ser observados em textos autênticos, produzidos naturalmente, sejam eles escritos ou falados² (Tagnin y Guimarães, 2004: 315).

En una primera fase de aplicación de la lingüística de corpus a la traducción todo el centro de interés estuvo dirigido por las preocupaciones en cuanto a los diferentes criterios de selección y compilación de las muestras textuales que mejor se adecuasen a los objetivos del trabajo de investigación traductológica (Abaitua, 2000: 2). Sin embargo, poco se hablaba acerca de cómo tratar y analizar la información que nos aportaba la gestión de esos textos, qué parámetros de estudio y qué enfoques adoptar para desarrollar modelos de análisis orientados específicamente a la traducción.

Para un estudio de los textos de un corpus adaptado a las necesidades de los traductores, la estilística funcional propone una concepción del texto como unidad comunicativa dependiente, como todo acto de comunicación, de la situación en la que se produce. Lvovskaya (2002) ha puesto los cimientos de una propuesta que, desde las bases teóricas de la estilística funcional, pretende dar las pautas de una tipología textual y de un estudio comparativo de las convenciones textuales dirigidos a las necesidades del traductor.

De acuerdo con los planteamientos expuestos anteriormente, el objetivo de este artículo es

presentar un ejemplo de las posibilidades que la explotación de un corpus de especialidad ofrece a la práctica terminográfica y a los estudios de traducción. Tras la exposición del marco teórico, abordamos los pasos seguidos para la creación de un corpus comparable³ bilingüe (francés y español) especializado en un subdominio de la biomedicina (enfermedades neuromusculares pediátricas) y en un género textual específico (artículo científico). A continuación, describimos la metodología y herramientas utilizadas para llevar a cabo su explotación, con especial interés en algunas herramientas específicas como el extractor de contextos definitorios ECODE. Por último, explicamos cómo se ha llevado a cabo el estudio de cada subcorpus a través del análisis de las cadenas temáticas y de sus marcadores. Tomamos como hipótesis de partida que este análisis permitirá identificar las áreas que presentan un mayor desarrollo terminológico y referencial en el mismo.

Este trabajo forma parte de un proyecto de investigación en el área de la traducción de textos médicos, en concreto textos sobre enfermedades raras, ámbito en el que se integran las enfermedades neuromusculares pediátricas. Se trata de una línea de investigación que toma como base la elaboración de corpus multilingües para la creación de recursos para traductores, intérpretes o redactores científicos y técnicos, pero también para otros destinatarios, como los profesionales del ámbito sanitario o los propios pacientes.

Queremos subrayar que tanto el área de las enfermedades neuromusculares como las lenguas implicadas (francés-español) responden a

2 Traducción propia (adaptada) al español: "... la Lingüística de Corpus ha desempeñado un papel indiscutible en la evolución de los estudios traductológicos y terminológicos, por insistir en la noción de lenguaje en tanto que objeto de investigación cuyos fenómenos solo pueden observarse en textos auténticos, producidos naturalmente, ya sean escritos o hablados".

3 Rabadán y Fernández (2002: 54) destacan el potencial extraordinario que presentan estos corpus en la investigación en traducción, ya que estos presentan la ventaja de contener textos originales (no traducciones) que tratan sobre temas similares y que funcionan de forma semejante en el plano de la comunicación, sin la contaminación o influencia de otras lenguas —como es el caso de los corpus paralelos.

un ámbito en el que existe actualmente una demanda real de traducciones. Francia es uno de los países europeos que lideran la investigación en enfermedades neuromusculares y desde su Gobierno se han impulsado una serie de medidas específicas de apoyo a la lucha contra estas afecciones⁴. Destacada también es la labor que desempeña desde hace más de 50 años la *Association Française contre les Myopathies* (AFM), uno de los organismos con mayor prestigio en territorio francés y un ejemplo de gestión para asociaciones de otros países. Las distintas publicaciones de la AFM abordan temas de la actualidad sanitaria, científica y social en torno a las enfermedades neuromusculares y constituyen una fuente de documentación a la que recurren otras asociaciones, entre ellas la federación española contra las enfermedades neuromusculares (ASEM). La asociación española, aún en fase de consolidación, cuenta con el apoyo de la AFM, de la cual procede la mayoría de sus publicaciones. Y es en ese terreno donde los traductores pueden aportar su grano de arena, vertiendo al español los documentos de la AFM, con el fin de facilitar el acceso a esta información a la comunidad hispanohablante.

2. Marco de referencia

En los siguientes apartados analizaremos las posibilidades que la explotación de corpus ofrece a terminólogos y traductores, centrándonos en un tipo de corpus concreto, el corpus comparable especializado, que mayor interés tiene para el estudio de las lenguas de especialidad. El uso de corpus especializados satisface no solo las necesidades lingüísticas sino también conceptuales del investigador de un área del conocimiento:

A corpus can be a useful resource for learning about the linguistic features of an LSP, such as knowledge about terms, collocations, grammar and style. It can also provide conceptual information, such as knowledge about the characteristics of the concepts behind the terms and about the relationships concepts have with one another⁵ (Bowker y Pearson, 2002: 39).

Por otro lado, hemos querido combinar, en un mismo estudio, dos metodologías de interrogación de corpus —aquella propia de los trabajos terminográficos y la propuesta desde el enfoque de la estilística funcional— con el objetivo de responder a los intereses del traductor especializado. En su práctica profesional, este requiere no solamente información sobre los términos y conceptos de un área del conocimiento concreta, sino también sobre los textos donde *habitan* dichos términos. Como explicaremos en el apartado 2.3, la tipología y los parámetros textuales propuestos por Lvovskaya (2002) permiten un estudio contrastivo, sistemático y objetivo, de las convenciones propias de los textos producidos en diferentes culturas en determinadas situaciones comunicativas a través del análisis de las categorías o parámetros textuales.

2.1. La lingüística de corpus al servicio de la terminología

La metodología denominada *lingüística de corpus* ha tenido y sigue teniendo una importante repercusión en el estudio del lenguaje y en el desarrollo de programas informáticos para el procesamiento del lenguaje natural (Butler, 1992; McEnery y Wilson, 1996). Desde sus comienzos, esta metodología ha tenido la pretensión empiricista de estudiar el funcionamiento real de la lengua mediante el estudio de ingentes cantidades de textos producidos por los hablantes de una

4 Ejemplos de su actuación son el *1er Plan Maladies Rares* (2005-2008) y el *2ème Plan Maladies Rares* (2011-2014) llevados a cabo por el Ministère de la Santé et de la Recherche.

5 Traducción propia (adaptada) al español: “Un corpus puede constituir un recurso útil para estudiar las características lingüísticas de un lenguaje de especialidad, proporcionando información sobre los términos, colocaciones, gramática y estilo. Asimismo, el estudio de corpus puede proporcionar información conceptual como el reconocimiento de las características de los conceptos que encierra cada uno de los términos y las relaciones que los conceptos establecen entre sí”.

lengua. Haciendo un símil con conceptos propios de la *ecología* (Walker y Amsler, 1991), podemos decir que, en el estudio de un área del conocimiento, los textos profesionales y académicos específicos de ese ámbito conforman el *ecosistema* en el que habitan las diferentes *especies*, entendidas aquí como unidades léxicas (Vargas, 2006: 2). De este modo, los términos son observados en su entorno natural, de forma descriptiva y no prescriptiva.

El texto, como *hábitat* de los términos, es objeto de estudio de la llamada terminología textual (Finatto, 2007). Este enfoque, de cuño comunicativo-textual, dedica especial atención a los términos en el texto, en los contextos de ocurrencia, pero también a la caracterización de los textos de especialidad en relación a los términos y a otras estructuras sintáctico-semánticas y pragmáticas que vehiculan conocimiento especializado (Krieger y Finatto, 2004: 192).

El texto ha pasado a ser una parte imprescindible del análisis terminológico. En la contextualización de los términos se puede extraer información lingüística, semántica y conceptual. Gracias a la importante cantidad de textos accesibles en la red y el acelerado desarrollo de la sociedad de la información y el conocimiento, que ha dado lugar a un buen número de herramientas de análisis textual, el estudio de amplias colecciones de textos se ha revelado como un procedimiento útil para la creación de recursos terminográficos multilingües (Sánchez, 2006: 69). La extracción de conocimiento mediante corpus multilingües permite estudiar las lenguas de especialidad en el uso que de ellas hacen sus productores naturales o quienes las emplean para comunicarse (Pérez, 2002).

2.2. La lingüística de corpus al servicio de la traducción

La investigación sobre corpus aplicada a la traducción, denominada en inglés *Corpus Translations Studies* (CTS), se remonta a principios de la

década de los noventa. Rebatida en un primer momento por su vinculación con la lingüística, la línea de investigación centrada en la explotación de corpus está en sintonía con las aproximaciones descriptivas y los estudios centrados en el producto impulsados por Toury y Lefevere. Baker y otros (1993) busca una convergencia entre los partidarios de estudios descriptivos dentro de la traducción y las propuestas de la lingüística de corpus de extraer inferencias sobre el funcionamiento de la lengua basadas en datos reales. De hecho, Baker sostiene que las técnicas estadísticas y la metodología puestas en funcionamiento por Sinclair (1991) pueden contribuir al paso de estudios prescriptivos a descriptivos. La autora egipcia apunta asimismo que el cambio de orientación, desde una perspectiva conceptual a un enfoque situacional, y del significado al uso, es la base del desarrollo de una nueva manera de trabajar que, sin duda, toma el texto como eje central de estudio.

Siguiendo la inclinación hacia la interdisciplinariedad de la que se nutre la traducción, numerosos investigadores en el área de los estudios de traducción han defendido la utilización de corpus informatizados (Baker y otros, 1993; Hartmann, 1994; Barlow, 1996). La utilización de los diferentes tipos de corpus permite el estudio *in vivo* de la fraseología, estructuras retóricas, prosodia semántica, patrones colocacionales, usos metafóricos de un vocablo, equivalencia entre términos de dos lenguas, convenciones textuales, entre otras variables.

Más recientemente, Zanettin, Bernardini y Stewart (2003: 1) bautizaron como *Applied Corpus-Based Translation Studies* a la rama resultante de la convergencia de CTS y *Applied Translation Studies*. Los tres investigadores resaltaron tres áreas relacionadas con la formación y la práctica profesional en las que el uso de corpus electrónicos resulta de interés: la enseñanza y aprendizaje de idiomas, la traducción asistida por ordenador y los estudios traductológicos de corpus.

2.3. Aplicaciones traductológicas de la estilística funcional en la interrogación de corpus

Mona Baker (1998) reconoce que el trabajo de corpus en traducción abre una perspectiva interesante, pero advierte que es necesario dar un paso más, establecer una metodología de interrogación de corpus y de análisis de los datos y resultados extraídos que responda a las necesidades del traductor. El estudio de los corpus deberá proporcionar a los teóricos de la traducción una metodología para observar su objeto de estudio y para explorar lo que lo hace diferente de otros objetos de estudio (Baker, 1993: 235). Es precisamente en este aspecto que las propuestas de la estilística funcional pueden aportar una visión complementaria. De hecho, los resultados de estudios estilísticos textuales que adopten una perspectiva contrastiva intercultural tendrían “una inestimable utilidad instrumental tanto para la formación de traductores e intérpretes como para los profesionales” (Lvovskaya, 2002: 11).

Todo texto posee unas características precisas que comparte con otros, generados en una situación comunicativa similar y que desempeñan un conjunto de funciones definidas también análogas. Estas características, sólidamente enraizadas en cada cultura, reciben el nombre de convenciones textuales y hacen posible el reconocimiento de un género textual por parte del receptor, desencadenando una serie de expectativas que facilitarán su comprensión. Podemos esperar que existan diferencias y/o similitudes entre las convenciones propias de textos aparecidos en situaciones comunicativas similares pero en el seno de diferentes culturas. El traductor deberá saber reconocer las analogías y diferencias existentes entre las convenciones de los textos propios de una y otra cultura, para poder producir en la cultura receptora un texto que

funcione pragmáticamente y sea aceptado por el público meta: “translated texts must speak ‘the language’ of the target audience and should resemble other texts produced within that particular language community and subject domain⁶” (Wright, 1993: 70).

Aunando factores lingüísticos y extralingüísticos (comunicativos), Lvovskaya (2002) aborda el estudio comparativo de las convenciones de tipos de texto a partir de categorías textuales universales, perceptibles en cualquier producto textual. Dichos parámetros no pertenecen a una categoría lingüística concreta, están determinados por su función semántico-lógica y presentan una serie de variaciones estilístico-funcionales según la situación e intención comunicativas y el tipo de textos.

Estas categorías textuales se subdividen, según su valor textual y el carácter de sus manifestaciones lingüísticas, en tres grandes grupos: lineales —cadenas temáticas y lógicas—, de campo y estructurales. A continuación, explicamos de forma sucinta en qué consiste cada categoría, aportando ejemplos extraídos de nuestro subcorpus en español.

La cadena temática compone, junto a la cadena lógica, las categorías lineales que se manifiestan en el texto mediante una serie de marcadores que cumplen un mismo papel semántico-funcional (Piñero y García, 2002: 164). Esta permite desarrollar el tema del texto a través de designaciones sucesivas que, mediante el principio de la correferencia, contraen entre sí relaciones anafóricas y catafóricas fundadas en principios semánticos —comparten igual sema— y pragmáticos —forman parte de un mismo ámbito de la experiencia humana (Piñero y García, 2003: 37). Esos marcadores constituyen un mecanismo primordial de cohesión textual creando un entramado de relaciones significativas dentro del discurso.

6 Traducción propia (adaptada) al español: “Los textos traducidos deben hablar ‘la lengua’ del público meta y deberían parecerse a otros textos producidos en una determinada comunidad lingüística y en un ámbito concreto”.

La cadena lógica está relacionada con el esquema lógico mental de todo texto ideado por su autor. Se diferencian tres tipos: marcadores lógico-objetivos, marcadores lógico-subjetivos y marcadores lógico-estructurales. Estos marcadores están relacionados con la coherencia y cohesión textuales y pueden intervenir como marcadores de otras categorías.

Los marcadores de carácter lógico-objetivo indican una demarcación denotativa e introducen un conjunto referencial en relación con el origen de la información facilitada —fuentes de autoridad a las que se remite en las citas y bibliografía— y en relación con una red de conocimiento compartido por los miembros de una determinada comunidad científica. Son marcadores de esta categoría aquellos que introducen información ilustrativa —*por ejemplo, como, tales como*— y datos procedentes de otras fuentes.

Los marcadores lógico-subjetivos están relacionados con las opiniones propias del autor —por ejemplo, verbos de entendimiento y de lengua, como *creo, me parece, me refiero, consideramos*, y estructuras que marcan el punto de vista del autor, como *en mi opinión, a nuestro juicio*— y el énfasis y/o matiz que deposita en determinados aspectos del texto. Estas fórmulas responden a lo que se ha denominado la ética del investigador (Díaz, 2003: 61). Tanto los marcadores lógico-objetivos como los lógico-subjetivos son polifuncionales y pueden actuar como marcadores de modalidad de la categoría campo.

En fin, los marcadores lógico-estructurales reflejan la lógica secuencial del discurso, encadenando un fragmento textual con su pretexto y postexto (Díaz, 2002: 138).

Las categorías de campo abarcan las modalidades objetivas (valorativas) y subjetivas (emotivo-expresivas), el tiempo y el espacio textuales, y sus marcadores pertenecen a diferentes niveles de la lengua.

La modalidad objetiva implica una evaluación racional o juicio de valor a través del uso de sinónimos textuales de bien *versus* mal, bueno *versus* malo (Lvovskaya, 2002: 49-50). Esta modalidad valorativa se corresponde con los juicios positivos o negativos que se emiten acerca de lo que se expone. En nuestro subcorpus en español hemos encontrado marcadores de esta modalidad sobre todo en los apartados de planteamiento del problema y conclusiones. Citamos, por ejemplo, *el mayor inconveniente (de este tipo de tratamiento); (tratamiento) adecuado; escasos (resultados); la eficacia (de este tipo de terapias) es inferior a*; etc.

Frente a la pretendida objetividad de los textos científicos, la práctica de los análisis textuales ha demostrado que la completa neutralidad y la emotividad cero no existen. En otras palabras, el autor del texto no solo se refiere en su discurso a un tema dado, sino que refleja, conscientemente o no, su propio yo, creando la llamada *tonalidad del texto* (Lvovskaya, 2002: 14). Los marcadores de la modalidad subjetiva reflejan el grado de convicción expresado por el emisor acerca de lo que dice —seguridad / falta de seguridad, interés / falta de interés, importancia / falta de importancia, suposición/duda, posibilidad/imposibilidad (Lvovskaya, 2002: 46). Su estudio permite analizar la intencionalidad del autor, reflejada en una estrategia discursiva concreta y en la selección reflexiva de las herramientas y recursos lingüísticos a su alcance para expresar su grado de convencimiento. De nuestro subcorpus en español citamos, por ejemplo, *según parece; eficacia dudosa; con escepticismo; no es fácil; paradójicamente; quizás insuficientemente*; etc.

La sucesión de bloques comunicativos o segmentos textuales conforma las categorías estructurales. Su configuración se relaciona con el programa conceptual del autor, pero debe atender a ciertas normas más o menos fijas. Todo discurso escrito posee unas características distintivas de organización discursiva. En el caso de los

artículos científicos como los aquí estudiados, la estructura suele responder al siguiente esquema: título, resumen / *abstract*, palabras clave / *keywords*, introducción, parte principal, conclusiones, notas y bibliografía. Dichos rasgos convencionales están sujetos a las normas de publicación de los editores.

Por su íntima relación con la carga terminológica, las connotaciones estilísticas que poseen, la consideración de aspectos lingüísticos y extralingüísticos y su importancia en la garantía de la coherencia textual, hemos optado por aplicar a nuestro corpus un análisis global de las cadenas temáticas.

Un término clave y sus sucesivas denominaciones formalizan la cadena temática principal. Al mismo tiempo se generan cadenas temáticas secundarias o adicionales subordinadas a la principal (García y Piñero, 2002: 164). En conjunto, unas y otras conforman las peculiaridades de la tipología textual en que se integran. Recíprocamente, un tipo textual determinado exigirá la inclusión de cadenas temáticas específicas.

El estudio de los marcadores de la cadena temática permite identificar las diferentes deno-

minaciones utilizadas para el tratamiento de un tema; designaciones y combinaciones referidas a un determinado tópico que no están recogidas en ningún diccionario monolingüe o bilingüe. En ocasiones, el análisis detallado revela términos o construcciones que en otros textos o circunstancias no se tomarían como propios de una determinada cadena temática.

La catalogación de los marcadores de la cadena temática reflejará los núcleos de interés de los textos, es decir, las áreas que tienen un mayor peso terminológico y referencial. En el apartado 4.2 presentamos y analizamos los marcadores de la cadena temática extraídos de nuestro corpus.

3. Recopilación del corpus sobre enfermedades neuromusculares en pediatría

3.1. Composición del corpus

Nuestro corpus está constituido por dos subcorpus especializados en el subdominio de las enfermedades neuromusculares en pediatría, en francés y en español, cada uno compuesto por veinticinco textos. El tamaño del corpus se fijó *a priori* por cuestiones relativas a las dimensiones y objetivos de nuestro estudio.

TABLA 1

Datos generales del corpus

CRITERIOS	DESCRIPCIÓN DEL CORPUS
Canal de producción	Textos escritos
Cantidad de palabras	<i>Tokens</i> : 72.316 (fr) y 102.790 (es) <i>Types</i> : 9.228 (fr) y 10.149 (es)
Contenido	Especializado tanto por el tema (contenido específico sobre enfermedades neuromusculares en pediatría) como por el tipo de género (artículos científicos)
Tamaño de las muestras	Textos íntegros
Anotación	Corpus no anotado, muestras analizadas en .txt.
Período de tiempo	Textos actuales (1995-2009)
Finalidad	Recopilación de textos con fines terminológicos
Idiomas	Bilingüe (francés y español)
Capacidad de actualización	Colección finita de textos

Los criterios que utilizamos en la compilación de los cincuenta textos de nuestro corpus bilingüe están inspirados en las propuestas de Sinclair (1996), Biber (1993), Meyer y Mackintosh (1996: 267), Bowker (1996: 42) y Pearson (1998: 56), constituyendo, al mismo tiempo, una integración híbrida de las diferentes directrices y una adaptación al caso concreto del corpus de enfermedades neuromusculares en pediatría.

En la tabla 1 recogemos las características principales de nuestro corpus atendiendo a los parámetros empleados para establecer la tipología de corpus.

3.1.1. Subcorpus en español

Este subcorpus está compuesto por veinticinco artículos científicos del área de la neuro-pediatría y del subdominio de las enfermedades neuromusculares, extraídos de revistas especializadas⁷ accesibles desde Internet. Todas las revistas cuentan con un índice de impacto elevado dentro de su ámbito de conocimiento. El período de publicación de estos artículos engloba los años comprendidos entre 1995 y 2009, siendo la mayoría de ellos fechados a partir del año 2000.

Por lugar de procedencia, los textos fueron publicados no solamente en revistas españolas, sino de países de Hispanoamérica, lo cual justifica la aparición de algunos americanismos como *tomografía axial computada* en vez de la forma más común en España, *tomografía axial computadorizada*.

Los textos abordan diferentes aspectos de las enfermedades neuromusculares en la infancia: diagnóstico, signos y síntomas, y terapia. Algunos tratan en concreto casos clínicos de una determinada afección mientras otros hablan en general de un grupo de dolencias.

En cuanto al número de palabras de cada muestra textual, el intervalo se sitúa entre 1.948

y 10.160 palabras, aunque la mayoría se sitúa en torno a las 4.000. Hemos utilizado el Translator's Abacus versión 3.10 (<http://www.globalrendering.com/download.html>) para realizar el recuento. Se han contabilizado sin tener en cuenta resúmenes en otras lenguas y referencias bibliográficas.

Queremos subrayar el hecho de que ha habido dos etapas de selección. En una primera etapa, fuimos nosotros los responsables de elegir una muestra de 25 textos. En un segundo momento, el Dr. Madruga, pediatra del Hospital de Sevilla, revisó nuestra propuesta y nos indicó la pertinencia de incluir algunos artículos de relevancia en el área.

A continuación presentamos los datos estadísticos del subcorpus en español:

TABLA 2

Datos estadísticos del corpus en español

<i>Tokens</i>	102.790
<i>Types</i>	10.149
<i>Ratio Type/Token</i>	9,87
<i>Estandarizado Type/Token</i>	41,68
<i>Longitud media de palabras</i>	5,45
<i>Oraciones</i>	3.045
<i>Párrafo</i>	799

3.1.2. Subcorpus en francés

Los criterios de compilación de este subcorpus son los mismos que los seguidos en español, si bien presenta algunas especificidades. La muestra textual está constituida, también en este caso, por veinticinco artículos científicos extraídos de revistas especializadas accesibles desde Internet, la inmensa mayoría de la publicación *Archives de pédiatrie*, prestigiosa revista de la *Société Française de Pédiatrie* y de la *Association des Pédiatres de Langue Française*. La te-

7 Las publicaciones seleccionadas han sido, por ejemplo, *Anales de Pediatría*, *Revista de Neurología*, *Revista Cubana de Pediatría*, *Revista de la Sociedad Boliviana de Pediatría*, *Archivos de Pediatría del Uruguay*, *Revista Chilena de Pediatría*, entre otras. La mayoría de las publicaciones están indexadas en Scopus.

mática así como el período de publicación coinciden con el del corpus en español, el 80% fechado en los últimos diez años. En esta ocasión, el 100% de los textos proceden únicamente de Francia, ya que nuestra universidad no está suscrita a publicaciones de características similares procedentes de otros países francófonos.

En cuanto al número de palabras de cada muestra textual, el intervalo se sitúa entre 1.236 y 9.892 palabras, aunque la mayoría se sitúa en torno a las 3.000 palabras. En la selección del corpus en lengua francesa el auxilio del Dr. Madruga fue menor ya que él no suele trabajar con bibliografía en esa lengua. Procuramos incluir artículos de contenido similar a los del corpus en español.

Los datos estadísticos del subcorpus en francés quedan reflejados en la tabla 3:

TABLA 3

Datos estadísticos del corpus en francés

<i>Tokens</i>	72.316
<i>Types</i>	9.228
Ratio <i>Type/Token</i>	12,76
Estandarizado <i>Type/Token</i>	42,8
Longitud media de palabras	5,34
Oraciones	2.406
Párrafo	543

3.2. Herramientas de análisis de corpus

En este apartado intentaremos mostrar la forma como hemos iniciado el tratamiento del corpus mediante el uso de la herramienta informática *WordSmith Tools v.3.0*⁸.

La primera operación que llevamos a cabo en la fase de procesamiento del corpus fue la aplicación *WordList* que genera la lista de posibles candidatos a términos simples (monoléxi-

cos) ordenados por frecuencia de aparición y por orden alfabético. Para evitar el ruido de palabras vacías de contenido se utilizó una *stopword list*. Una primera observación nos permite identificar palabras que, por su frecuencia de aparición en el ámbito de las enfermedades neuromusculares, podrían considerarse, de antemano, candidatos a términos.

Al cotejar la lista de las cincuenta palabras de aparición más frecuente en cada uno de los subcorpus, observamos que el 75% de los 50 *tokens* más frecuentes en el subcorpus en español tienen su equivalente en la lista de los 50 *tokens* más frecuentes en francés. Al seleccionar una colección de textos pertenecientes a un mismo tema, campo, género y motivación situacional y comunicativa, hemos podido obtener un corpus bilingüe comparable cualitativamente equilibrado y homogéneo. Se demuestra así que “words in specialized lexical fields and their equivalent ones in other languages share similarities in terms of context⁹” (Fung, 1995).

El análisis de las listas de frecuencia ha revelado que las unidades monoléxicas más frecuentes pertenecen a diferentes campos léxicos. En primer lugar, encontramos, en ambos subcorpus, una importante presencia de palabras de uso corriente en la lengua general, que adquieren un valor especializado en los textos del corpus, como *paciente/s*, *enfermedad/es*, *diagnóstico*, *tratamiento*, *terapia*, *síndrome*, *trastornos*, *alteraciones*, *déficit*, *anomalías*, *troubles*, *atteinte*, *signes*, entre otros.

En segundo lugar, encontramos vocabulario más especializado, como, por ejemplo, *miopatía*, *distrofia*, *hipotonía*, *miotonia* y *atrofia* —pertenecientes al área de la neurología—; *inmunohistoquímica* y *biopsia* —referidos a pruebas diagnósticas—; *gen*, *ADN*, *mutación*, *mitochondrial* o *ADNmt* —propios de la genética—, y, además,

8 Existen versiones posteriores con más funcionalidades, pero solo pudimos acceder a esta versión.

9 Traducción propia (adaptada) al español: “Los vocablos pertenecientes al léxico especializado y sus equivalentes en otras lenguas comparten similitudes en lo que se refiere al contexto de aparición”.

neonato, neonatal, lactante, recién (nacido), adolescente, nourrisson y nouveau-(né) —propios de la franja de edad pediátrica.

El listado de las unidades monoléxicas puso también en evidencia el fenómeno de *truncación*, muy frecuente en el área de las enfermedades neuromusculares. Así, tenemos:

- **Siglas:** *DMD (Distrofia muscular de Duchenne), DMB (Distrofia muscular de Becker), RM (resonancia magnética), MG (miastenia gravis), VCN (velocidad de conducción nerviosa), AME (atrofia muscular espinal), TC o TAC (tomografía computarizada o tomografía axial computarizada), DME (dermatomyosite de l'enfant), SGB (síndrome de Guillain-Barré)*... Tanto en español como en francés se usan habitualmente siglas en inglés, como *LGMD (Limb Girdle Muscular Dystrophy)* o *HNMS (Hereditary Motor Sensory Neuropathy)*.

- **Formas abreviadas:** *quimio (quimioterapia), fisio (fisioterapia), maladie de Becker (dystrophie musculaire de Becker)*...

- **Acrónimos:** *Agrimed (agricultura mediterránea), Insalud (Instituto Nacional de Salud)*...

- **Abreviaturas:** *b. i. d. (bis in die o dos veces al día), q. h. (cuando cada hora), s. op. s. (si opus sit o si es necesario)*...

Los resultados de este primer análisis nos permitieron, tras llevar a cabo una selección, obtener esa relación de posibles candidatos a términos¹⁰ monoléxicos a la que hicimos referencia. Se ha tomado esta lista como punto de partida para realizar búsquedas más precisas y llevar a cabo el análisis exploratorio de cada una de estas formas en su respectivo contexto de aparición.

Para esta operación empleamos el programa *Concord* que nos permite observar cómo funcionan los posibles candidatos en el interior del discurso, así como ver sus combinaciones más habi-

tuales y sus derivados. Con ello, pudimos apreciar cómo algunas de las unidades simples se convierten en la base de un término compuesto o acompañan dicha base. Por otro lado, los contextos ofrecen pistas al terminógrafo para intuir que se trata de una unidad terminológica, dado que, en determinados casos, dichos contextos facilitan información conceptual mediante la incorporación de definiciones.

El análisis de los *clusters* se realiza a partir de un fichero de concordancias de una determinada palabra. Tomando como referencia el horizonte colocacional fijado, es posible obtener secuencias de palabras repetidas con una determinada extensión, *clusters* de dos, tres, cuatro o más palabras.

Los patrones sintácticos más comunes en nuestro corpus responden a las siguientes estructuras:

- **nombre + adjetivo:** *miopatías inflamatorias, pie cavo, myopathies métaboliques, réflexes ostéotendineux*.

- **nombre + adjetivo + adjetivo:** *distrofias musculares congénitas, syndromes myasthéniques congénitaux*.

- **nombre + preposición + nombre:** *biopsia de nervio, marcha de pato, myopathie à bâtonnets*.

- **nombre + adjetivo + preposición + nombre:** *distrofias musculares de cinturas, dystrophies musculaires des ceintures*.

- **nombre + preposición + nombre propio:** *enfermedad de McArdle, miopatía de Bethlem, maladie de Steinert, syndrome de Walker-Werburg*.

- **nombre + adjetivo + preposición + nombre propio:** *distrofia miotónica de Steinert, distrofia muscular de Duchenne, dystrophie myotonique de Steinert*.

Las secuencias *enfermedad de, síndrome de*

10 Siguiendo los preceptos de la teoría comunicativa de la terminología (TCT), concebimos el término como una unidad poliédrica construida por tres dimensiones indisociables: la conceptual, la lingüística y la comunicativa (Cabré, 2003).

y *distrofia muscular* de seguidas de un nombre propio son particularmente frecuentes ya que la mayoría de estas patologías se designan por un epónimo correspondiente a un subtipo de una determinada enfermedad.

Al hablar de los términos compuestos o unidades polilexemáticas o sintagmáticas presentes en los discursos especializados, Montero y Faber (2008) proponen la siguiente tipología (se incluyen ejemplos del corpus):

- **Unidades lexicalizadas:** no permiten alteraciones en el orden de los elementos.

› Unidades fraseológicas lexicalizadas de la lengua general.

› Unidades lexicalizadas del dominio especializado: términos compuestos sin unión gráfica que transmiten un concepto especializado a través de dos o más palabras que no varían y cuyo orden es inalterable. Por ejemplo, *distrofia facioescapulohumeral*, *Western blot*...

- **Unidades no lexicalizadas:** permiten alteraciones en el orden de los elementos.

› Unidades fraseológicas usadas para parafrasear términos especializados o para cambiar de un registro formal a uno informal. Por ejemplo, *dolores musculares generalizados* y *mialgias*, *dolor de cabeza* y *cefalea*...

› Unidades fraseológicas de tipo metafórico. Por ejemplo, las metáforas empleadas para caracterizar a las distintas afecciones o malformaciones de los pies: *pie de trinchera*, *pie de elefante*, *pie de atleta*... O también las utilizadas para caracterizar los tipos de marcha: *marcha equina*, *marcha de pato*, *marcha laberíntica*, *marcha de gallo*...

› Unidades fraseológicas transmisoras de un solo concepto, equivalentes a términos compuestos, pero cuyos componentes carecen de fijación, admitiendo cambio de orden y variación. Por ejemplo: *biopsia muscular* y *biopsia del músculo*, *cintura pelviana* y *cintura pélvica*...

- **Unidades semilexicalizadas:** colocaciones.

3.3. La extracción automática de contextos definatorios

El avance tecnológico en el desarrollo de herramientas que agilizan el trabajo terminográfico y traductor los ha provisto, por un lado, de corpus lingüísticos especializados donde se almacena digitalmente una gran cantidad de documentos técnicos y, por otro lado, de sistemas para la explotación de dichos corpus, como son los sistemas para la extracción automática de términos.

Actualmente, existe un creciente interés por el desarrollo de sistemas que ayuden en la identificación automática de información relevante sobre términos que sea útil para describir su significado. Una herramienta para la extracción automática de contextos definatorios sería de gran utilidad tanto en la práctica terminográfica, como ayuda para la elaboración de diccionarios especializados, glosarios y ontologías, como en otros ámbitos como la traducción especializada, el aprendizaje de una lengua para fines específicos o la enseñanza de una especialidad.

Dentro del área del procesamiento del lenguaje natural distinguimos dos tendencias, la primera volcada en la recuperación de información, y la segunda, en la extracción de información. En terminología, las investigaciones en torno a la extracción de información se han orientado hacia el desarrollo de sistemas para extracción de términos y, más actualmente, hacia la implementación de sistemas de extracción de conocimiento definatorio. Esta línea de investigación ha supuesto un giro en la elaboración de recursos léxicos y terminológicos para la organización conceptual de las unidades de conocimiento especializado y en la descripción de sus significados.

En esta segunda línea se sitúan los trabajos del grupo de investigación del área de ingeniería lingüística de la Universidad Nacional Autónoma de México (UNAM), encabezado por Gerardo

Sierra y Rodrigo Alarcón¹¹. Estos investigadores han desarrollado un programa extractor automático de contextos definitorios basado en reglas lingüísticas (ECODE) y el sistema Describe® para la búsqueda, clasificación y agrupamiento de definiciones en la web.

En plena fase de experimentación, el ECODE aún no estaba disponible en línea, pero tuvimos la oportunidad de formar parte de este proyecto ya que nuestro corpus en español ha sido analizado por ECODE. El uso de esta herramienta supuso un importante ahorro de tiempo en cuanto a la localización de información pertinente para la redacción de las definiciones.

A continuación presentamos algunos de los contextos definitorios obtenidos al final del tratamiento realizado por el ECODE, contextos que nos permitieron obtener informaciones conceptuales importantes para redactar las definiciones y diseñar el árbol conceptual. Hemos limpiado las etiquetas del ECODE, pero en cursiva están señalados los verbos y nexos que permitieron identificar los contextos definitorios:

- Los trastornos neuromusculares *constituyen* un grupo de enfermedades que afectan cualquiera de los componentes de la unidad motora, *es decir*, la unidad...
- La distrofia miotónica de Steinert *se caracteriza por* miotonía (dificultad o retardo en la relajación del músculo luego de su contracción), y debilidad muscular...
- La hipotonía neonatal generalizada *se define como* la disminución patológica del tono postural en las cuatro extremidades, el tronco y el cuello...
- El síndrome de Walker-Warburg *consiste en* distrofia muscular congénita con anomalías de las circunvoluciones cerebrales y cerebelosas, hipodensidad...
- *Como* amiotrofias musculares proximales se

conoce a un grupo de enfermedades neuromusculares hereditarias de transmisión autosómica recesiva...

4. Análisis textual del corpus desde la estilística funcional: estudio de la cadena temática

4.1. Consideraciones previas

Pasamos ahora a detallar la metodología empleada para el análisis de las cadenas temáticas y sus marcadores en cada subcorpus. El listado de frecuencias, los *clusters* y concordancias obtenidos automáticamente con el uso de *WordSmith Tools* nos sirvieron de punto de partida para saber cuáles eran los temas principales y secundarios de los textos, así como para agrupar las diferentes denominaciones relacionadas, directa o indirectamente, con cada cadena temática. Esta herramienta permite elaborar datos estadísticos que reflejan la incidencia de las diversas cadenas por su frecuencia de aparición en el corpus analizado.

Pero el análisis estadístico no es suficiente en sí mismo, sino que es necesario estudiar el tipo de relación semántica (sinonimia, hiponimia, hiperonimia, derivación y cognación) o pragmática (metonimia) en virtud de la que los marcadores se integran en una u otra cadena temática. Dichos marcadores no se identifican con una categoría lingüística concreta; la correferencia se expresa con unidades textuales autosemánticas (sustantivos, verbos, adjetivos, adverbios) o unidades textuales fóricas (proformas).

Además de identificar y analizar las cadenas temáticas, los tipos de relaciones semánticas o pragmáticas más frecuentes, la naturaleza gramatical de los marcadores y las estructuras de los marcadores complejos, estudiaremos las posibles similitudes y/o diferencias entre los marcadores de las dos lenguas de trabajo.

11 Remitimos a los trabajos de Sierra y otros (2003, 2008) y Sierra (2009).

4.2. La naturaleza de la cadena temática en nuestro corpus

Para la realización del estudio analítico y clasificatorio de las cadenas temáticas nos hemos guiado por los trabajos de Sánchez (2003, 2006) y Piñero y García (2002, 2003).

El primer problema que se nos planteó fue la identificación de las cadenas temáticas más importantes y cómo agruparlas. Sánchez (2006: 79), en su trabajo con un corpus de textos sobre enfermedades neuromusculares, identificaba las siguientes cadenas temáticas más frecuentes: *paciente, patología, terapia, consecuencias de la terapia*. Piñero y García (2003: 39) en su trabajo con textos en el área de fisioterapia identificaban dos supracadenas: *tratamiento y paciente*. Estas autoras hacían hincapié en la conveniencia de reducir el número de cadenas temáticas y agruparlas en torno a las dos supracadenas según relaciones metonímicas, como las que se establecen, por ejemplo, entre el paciente y las enfermedades que padece, o entre el paciente y las partes de su organismo. Así siendo, las *patologías* se integraban, en virtud de una relación pragmática, como marcador metonímico en la cadena temática *paciente*.

Sin embargo, Sánchez (2003: 26), en el mismo tipo de texto, reconocía tres cadenas temáticas principales: *tratamiento, patología y paciente*, sin incluir *patología* dentro de *paciente*. La autora subrayaba que la combinación de las tres cadenas temáticas permitía recomponer los títulos de los artículos analizados. Preferimos decantarnos por la propuesta de Sánchez (2003) y optar por considerar *patología* y *paciente* como dos cadenas separadas ya que en nuestros artículos ambas cadenas tienen una gran relevancia.

Una segunda dificultad fue la de identificar los marcadores de esta categoría. Estos pertenecen a diversas categorías gramaticales, se producen diversas posibilidades combinatorias y entre ellos se verifican relaciones de naturaleza

semántica pero también pragmática, de tal forma que dos vocablos aparentemente inconexos pueden formar parte de una misma cadena temática y funcionar como marcadores indirectos.

Veamos cómo hemos ido abordando estos problemas para la identificación de las diferentes cadenas temáticas y de sus marcadores presentes en nuestro corpus.

4.2.1. Análisis de los marcadores de la cadena temática en el subcorpus en español

Para establecer las cadenas temáticas principales de los textos del corpus nos hemos basado en dos criterios: análisis de los términos de mayor frecuencia de aparición y análisis de los títulos de los artículos.

Si analizamos los títulos de nuestros artículos podemos extraer cuatro cadenas.

1. *Patología / enfermedad / síndrome / enfermedad en pediatría / patología en pediatría / enfermedades en niños / enfermedades pediátricas / variedad neonatal...*
2. *Paciente / paciente en pediatría / paciente crítico en pediatría...*
3. *Diagnóstico/evaluación.*
4. *Tratamiento.*

En la lista de frecuencia que presentamos a continuación, las palabras de mayor frecuencia de aparición coinciden con los temas anteriores.

TABLA 4

Lista de frecuencias del subcorpus en español

Word	Frecuencia	%
PACIENTES	497	0,48
DIAGNÓSTICO	356	0,35
CASOS	329	0,32
ENFERMEDAD	312	0,3
TRATAMIENTO	254	0,25

Así siendo, optamos por reconocer cuatro cadenas temáticas: *paciente*, *patología*, *diagnóstico* y *tratamiento*. Estas cuatro cadenas están representadas en el título de nuestro proyecto: *Enfermedades neuromusculares en pediatría: diagnóstico y tratamiento*, de tal forma que los textos elegidos intencionadamente respondían a esos cuatro focos de interés.

Establecidas las cadenas temáticas principales, vamos a estudiar los marcadores de cada cadena. En primer lugar la cadena *paciente*. Destacamos aquí que el origen genético de la mayoría de las enfermedades neuromusculares influirá en la variedad de denominaciones incluidas en la cadena *paciente*. Así lo ponía de relieve Sánchez (2006) al indicar que debemos incluir como marcadores de la cadena *paciente* términos y construcciones que en otros textos no se considerarían como propios de esta cadena, como, por ejemplo, *transmisores de la enfermedad* o *familiares no afectados*. Dentro de esta misma cadena temática aparecerán los términos *recién nacido* y *niño*, debido al hecho de que los artículos hacen referencia a las enfermedades neuromusculares en edad pediátrica.

Como se puede observar en la siguiente tabla, dentro de la cadena temática *paciente* consideramos los siguientes marcadores¹².

TABLA 5

Marcadores cadena temática *paciente*

Paciente/s
<p>casos; niño/s; recién nacido; músculo; vida (calidad de); edad; neonatal; madre; familia; familiares; grupo étnico; grupo familiar; dolor; neonatos.</p>

Agrupamos ahora los marcadores que comparten un mismo núcleo y que se difieren en la adjetivación especificativa. El adjetivo que los acompaña acota la extensión semántica del sustantivo y le proporciona una referencia unívoca. Otros marcadores complejos están formados por sustantivo seguido de una frase preposicional.

- **Paciente/s:** *paciente crítico*, *paciente con diagnóstico de* (nombre de la enfermedad), *paciente afectado de* (nombre de la enfermedad), *paciente afecto de* (nombre de la enfermedad), *paciente con enfermedad* (neuromuscular), *paciente con enfermedad de* (nombre de la enfermedad), *calidad de vida*...

- **Recién nacido y niño:** *recién nacidos hipotónicos*, *recién nacidos vivos*, *recién nacidos con hipotonía*, *recién nacidos afectados*, *niño con antecedentes*, *niña con antecedentes*, *niña asintomática*, *niño con* (nombre de la enfermedad)...

- **Madre y familiares:** *madre afectada*, *madre no afectada*, *madre asintomática*, *mujeres portadoras*, *madre portadora*, *estado portador de la madre*, *madre del paciente*, *madre transmisora*, *madre miasténica*, *mujeres en riesgo*, *familiares afectados*, *familias con deleciones en el gen*, *familias estudiadas*, *hermanas portadoras*, *antecedentes familiares*, *antecedentes personales*...

- **Casos:** *casos neonatales*, *casos clínicos*, *casos sin diagnóstico*, *casos estudiados*, *casos más graves*, *casos evaluados*...

- **Músculo:** *músculo liso*, *músculo esquelético*, *músculo estriado*, *músculos*, *músculos respiratorios*, *músculos faciales*, *músculos intercostales*, *músculos distales*, *músculos proximales*, *músculos de la cintura*, *debilidad muscular*...

- **Vida:** *calidad de vida*, *expectativa de vida*, *su vida*, *muerte*...

12 En los estudios de los marcadores de cada cadena temática, incluimos solo los de mayor frecuencia, ya que, dada la cantidad de los textos, un análisis detallado de todos y cada uno de los marcadores exigiría una dedicación temporal que excede a las características de este trabajo.

La siguiente tabla recoge los marcadores de la cadena *patología*.

TABLA 6Marcadores cadena temática *patología*

Patología
<p>↳ enfermedad; afectación; alteraciones; forma; trastornos; distrofia; tipo; síndrome; grupo; inicio; miopatía; síntomas; signos; complicaciones; transmisión; su (diagnóstico).</p>

Destacamos aquí algunas combinaciones más frecuentes de los marcadores de la cadena *patología*.

- **Enfermedad:** enfermedad subyacente, enfermedad hereditaria, enfermedad neuromuscular, enfermedad infantil, enfermedad autoinmune, grupo de enfermedades, enfermedades neuromusculares, enfermedades raras...

- **Formas:** forma típica, forma infantil, forma recesiva, forma congénita, forma autosómica, distintas formas, formas generalizadas...

- **Alteraciones y trastornos:** trastorno neuromuscular, trastornos genéticos, alteraciones histológicas, alteraciones neuromusculares...

Las dos cadenas de mayor peso en los textos eran la de *paciente* y *patología*. Pero incluimos el estudio de otras dos, *diagnóstico* y *tratamiento*, porque algunos textos del corpus tenían como foco de interés los métodos diagnósticos y los avances en el tratamiento de las enfermedades neuromusculares. Artículos como estos: "Importancia diagnóstica de los signos de hipocinesia fetal en la atrofia muscular espinal de presentación neonatal", "Distrofia muscular de Duchenne: actualidad y perspectivas de tratamiento" y "Efecto de los corticoides en el tratamiento de la distrofia muscular de Duchenne" presentan también como centro de interés las formas de tratamiento y los métodos diagnósticos.

La cadena *diagnóstico* presenta los siguientes marcadores.

TABLA 7Marcadores cadena temática *diagnóstico*

Diagnóstico
<p>↳ análisis; estudio; evaluación; resultados; datos; test; detección.</p>

Las combinaciones más frecuentes de los marcadores de *diagnóstico* son:

- **Diagnóstico:** diagnóstico clínico, diagnóstico diferencial, diagnóstico prenatal, diagnóstico precoz, diagnóstico genético, diagnóstico molecular, diagnóstico sindromático, diagnóstico preciso, avances diagnósticos...

- **Estudios, análisis, evaluación:** estudio genético, estudio electrofisiológico, estudio retrospectivo, análisis directo, análisis morfológico, estudio inmunohistoquímico, estudio previo, estudio controlado, estudio bioquímico, evaluación diagnóstica...

Cadena temática *tratamiento*:

TABLA 8Marcadores cadena temática *tratamiento*

Tratamiento
<p>↳ terapia; procesos; respuesta; estimulación; repetitiva.</p>

Las combinaciones más frecuentes de los marcadores de *tratamiento* son:

- **Tratamiento/s:** tratamiento con, tratamiento corticoideo, tratamiento crónico, tratamiento quirúrgico, tratamiento farmacológico, tratamiento específico, tratamiento esteroideo, tratamiento oportuno, tratamiento rehabilitador...

- **Terapia:** terapia física, terapia de reemplazo, terapia génica, terapia farmacológica, fisioterapia...

4.2.2. Análisis de los marcadores de la cadena temática en el corpus en francés

Repetimos las mismas operaciones para el corpus en lengua francesa. De los títulos de los 25 artículos obtenemos las siguientes cadenas temáticas:

1. *Enfant/nourrisson.*
2. *Maladie.*
3. *Étude/diagnostic/stratégie diagnostique.*
4. *Essai thérapeutique / prise en charge.*

Tras la elaboración de la lista de frecuencia se obtuvieron los resultados recogidos en la siguiente tabla.

TABLA 9

Lista de frecuencias del subcorpus en francés

Word	Frecuencia	%
MALADIE	270	0,37
ENFANT	185	0,26
DIAGNOSTIC	161	0,22
PATIENTS	144	0,2
TRAITEMENT	52	0,07

Destacamos que el término *tratamiento* en el corpus de francés aparece con mucho menos frecuencia que en el corpus en español.

Los marcadores más frecuentes de la cadena *patient* se recogen en esta tabla.

TABLA 10

Marcadores cadena temática *patient*

CCC	Pathologie
→	<i>enfant; nourrisson; adolescent; famille; adulte; malade; parents; âge; décès; fratrie; nouveau-né; porteurs; atteints; muscle.</i>

Las combinaciones más frecuentes de los marcadores de esta cadena son:

- **Enfant:** *enfant malade, enfant porteur, enfant atteint, jeune enfant...*

- **Nourrisson:** *nourrisson atteint, nourrisson immobile, nourrisson hypotonique, nourrisson traité...*

La cadena temática *pathologie* se manifiesta con los siguientes marcadores:

TABLA 11

Marcadores cadena temática *pathologie*

CCC	Pathologie
→	<i>maladie; troubles; atteinte; forme; signe; syndrome; transmission; anomalies; début; effet; origine; apparition; types.</i>

En función de los patrones sintácticos más frecuentes podemos destacar las siguientes combinaciones:

- **Maladies:** *maladies neuromusculaires de l'enfant, maladies musculaires, maladies héréditaires, maladies mitochondriales, maladies de surcharge...*

- **Atteinte:** *atteinte musculaire, atteinte motoneurale, atteinte motrice, atteinte respiratoire, atteinte cardiaque, atteinte neurologique, atteinte cognitive...*

- **Troubles:** *troubles de la déglutition, troubles respiratoires, troubles sensitifs, troubles de conduction, l'apparition de troubles...*

- **Transmission:** *transmission maternelle, transmission paternelle, transmission autosomique récessive, transmission autosomique dominante, transmission autosomique récessive liée à, mode de transmission...*

La cadena temática *diagnostic* se recoge en la siguiente tabla:

TABLA 12

Marcadores cadena temática *diagnostic*

	Diagnostic
	→ <i>étude; analyse; résultat.</i>

Las estructuras más frecuentes son, por ejemplo:

- **Étude**: *étude familiale, étude génétique, étude du gène, étude des protéines...*

- **Analyse**: *analyse clinique, analyse moléculaire, analyse sémiologique, analyse histologique...*

- **Diagnostic**: *diagnostic clinique, diagnostic précoce, diagnostic étiologique, diagnostic différentiel, diagnostic définitif, diagnostic prénatal, diagnostic anténatal, diagnostic précis, l'annonce du diagnostic...*

Por último, la cadena *traitement* presenta los siguientes marcadores:

TABLA 13

Marcadores cadena temática *traitement*

	Traitement
	→ <i>thérapie; essais; prise en charge; soins.</i>

Dentro de la cadena *traitement* destacamos estas combinaciones:

- **Thérapie**: *thérapie génique, ergothérapie, enzymothérapie, enzymothérapie substitutive, kinésithérapie, essais thérapeutique...*

- **Traitement**: *traitement efficace, traitement par transfert de gène, traitement génétique...*

4.3. Consideraciones finales acerca de los marcadores de la cadena temática

Hemos querido presentar algunas consideraciones generales acerca de la naturaleza de los marcadores de la categoría cadena temática. Se

haría necesario un estudio más profundo para poder ahondar en estas conclusiones de carácter provisional. No obstante, queremos destacar algunos aspectos que, a nuestro juicio, presentan cierto interés para estudios posteriores.

En un análisis que tenga como objeto indagar en lo más común y sistemático, tendríamos que destacar las dos supracadenas más evidentes en los 25 textos de cada idioma, que son la de *paciente* y *patología*. Prácticamente la totalidad de los artículos dan protagonismo a los dos; por ello no podríamos incluir *patología* dentro de *paciente*, como sí lo hicieron Piñero y García (2003) en su estudio, concibiendo la *patología* como la enfermedad que sufre el *paciente*. Efectivamente, en artículos como los analizados por las dos investigadoras, centrados en el área de la fisioterapia, será más probable que sean el *tratamiento* y el *paciente* los focos de atención. Sin embargo, nuestro estudio se hizo de artículos sobre enfermedades neuromusculares, de ahí que en los artículos gran cantidad de espacio se dedique a todo lo relacionado con estas afecciones, siendo *patología* una cadena de mayor peso. La lista de frecuencias nos mostraba que *paciente* era otro foco de atención, con muchos marcadores sinónimos propios de la franja de edad pediátrica (*recién nacido, niño, adolescente*).

Las otras dos cadenas elegidas, *diagnóstico* y *tratamiento*, las podríamos haber incluido como subcadenas de *patología*, ya que, en la mayoría de los casos, los autores hacían referencia a pruebas diagnósticas y terapias específicas en relación a una enfermedad o grupo de enfermedades. Salvo en aquellos artículos en los que el tema central era la investigación de una nueva terapia o la eficacia de una prueba diagnóstica para el diagnóstico preciso de una enfermedad neuromuscular, las cadenas *diagnóstico* y *tratamiento* quizás podrían haberse incluido como subcadenas con referencia a *patología* con una relación pragmática.

Los marcadores temáticos del corpus, tanto en español como en francés, se caracterizan por una escasa diversidad léxica. Creemos que esto se deriva del interés en anteponer la eficacia, efectividad y claridad expositiva a la preocupación estilística. La mayoría de los marcadores que integran las cadenas temáticas se forja a partir del mecanismo de recurrencia léxica. Por este motivo, marcadores como *paciente*, *enfermedad*, *diagnóstico* y *tratamiento* aparecen con tanta frecuencia en el corpus en español, sin que otros marcadores de esas cadenas tengan la misma periodicidad. Al mismo tiempo, esos marcadores constituyen el núcleo de diferentes marcadores complejos. Lo mismo es aplicable para el corpus en francés, en el que se repiten constantemente los marcadores *enfant*, *maladie* y *diagnostic* con sus muchas combinaciones.

Bien que teóricamente los marcadores de la cadena temática pueden ser sustantivos, adjetivos, verbos o adverbios, así como pronombres, determinantes o pronombres adverbiales, el lexema sustantivo es el marcador por excelencia de los textos de nuestro corpus. La alta proporción de sustantivos es un rasgo muy característico del discurso científico en general. Hemos podido ver cómo la gran mayoría de los marcadores pertenecen a la categoría de sustantivos, acompañados de adjetivaciones, de construcciones prepositivas o una combinación de ambas estructuras que precisan su significado.

Sin embargo, debemos mencionar también el caso de otros elementos, como, por ejemplo, de las proformas que constituyen una categoría particular en las cadenas *paciente* y *patient/s*. En el subcorpus en español, la proforma más reiterada es el posesivo *su*, tanto en singular como en plural, cuya referencia anafórica facilita la relación entre una serie de sustantivos y el *paciente*, y cuyo ejemplo más usado es *su (diagnóstico)*. En el subcorpus en francés, también son frecuentes los posesivos acompañando a denominaciones de miembros de la familia (*sa tante, mère, famille*

/ son fils, grand-père) o el cuerpo o parte de este (*son corps, leur bras, leurs mains, leurs doigts...*).

Por otro lado, hemos visto algunos ejemplos de relaciones metonímicas, como las que se establecen entre *músculo* (u otras partes del cuerpo) y *paciente*, relaciones pragmáticas como las que relacionan *dolor* o *debilidad* con *paciente*. Encontramos igualmente ejemplos de deixis retrospectiva, con el uso de determinantes demostrativos, como, por ejemplo, *estos trastornos* (como correferencia a los *trastornos neuromusculares*), *esta terapia* (como correferencia a *terapias génicas*)... Dichas construcciones remiten a un marcador anterior, evitando así la reproducción reiterada de una misma combinación.

5. Conclusiones

En este trabajo hemos presentado las fases seguidas para la elaboración del corpus bilingüe comparable y su explotación con la finalidad de crear recursos terminológicos de utilidad para el trabajo de traductores en el área de las enfermedades neuromusculares.

Para la compilación del corpus hemos tenido en cuenta los criterios de representatividad, cantidad y calidad, así como de equilibrio entre los subcorpus, con la finalidad de poder obtener unos resultados fiables.

Habíamos planteado la utilidad de los programas, como el ECODE, que permiten reconocer y extraer contextos definitorios, recuperar relaciones semánticas específicas y descripciones generales acerca del significado de los términos. La posibilidad de participar en el proceso de evaluación del ECODE, gracias al equipo de Sierra y Alarcón de la UNAM, nos sirvió para corroborar que esta herramienta agiliza el proceso de elaboración de recursos terminológicos.

Hemos identificado y analizado las cadenas temáticas de los dos subcorpus, los tipos de relaciones semánticas o pragmáticas más frecuentes, la naturaleza gramatical de los marcadores

y los patrones sintácticos de los marcadores complejos. El estudio de las cadenas temáticas se convierte en una especie de radiografía de la columna vertebral terminológica de los textos analizados, a través de la que se han observado las áreas con mayor peso terminológico y referencial.

La metodología que adoptamos en el análisis de los marcadores de la cadena temática constituye una visión novedosa cuyos hallazgos necesitan aún cierto tiempo de maduración. Por lo tanto, nuestra propuesta es susceptible de posteriores revisiones para establecer unos criterios homogéneos de aplicación universal.

Gracias a este proyecto nos hemos podido acercar a una de las líneas de investigación que más atención e interés ha generado en los últimos años, tanto en las disciplinas lingüísticas como en áreas de la ingeniería computacional. El uso de corpus ha cambiado definitivamente el paradigma de investigación en traducción, marcando un hito solo comparable a la formulación touriana de normas y la redefinición del concepto de equivalencia (Malmkjaer, 2003: 119).

6. Bibliografía citada

ABAITUA, Joseba, 2000: *Tratamiento de córpora bilingües*, Fundación Duques de Soria, Seminario "La ingeniería lingüística", Soria, 17-21 de julio de 2000. Documento disponible en línea [<http://paginaspersonales.deusto.es/abaitua/konzeptu/ta/soria00.htm>, fecha de consulta: 29 de marzo de 2011].

BAKER, Mona, 1993: "Corpus Linguistics and Translation Studies. Implications and Applications" en Mona BAKER y otros (eds.): *Text and Technology: In Honour of John Sinclair*, Amsterdam / Philadelphia: John Benjamins, 233-250.

BAKER, Mona y otros, 1993: *Text and Technology*, Amsterdam / Philadelphia: John Benjamins.

BAKER, Mona, 1998: "Réexplorer la langue de la tra-

duction: une approche par corpus", *Meta* 43 (4).

BARLOW, Michael, 1996: "Parallel texts in language teaching" en Simon Philip BOTLEY y otros (eds.): *Proceedings of Teaching and Language Corpora* (UCREL Technical Papers, vol. 9), Lancaster: University of Lancaster, 45-56.

BIBER, Douglas, 1993: "Representativeness in corpus design", *Literary and Linguistic Computing*, vol. 8 (4), 243-257.

BOWKER, Lynne, 1996: "Towards a Corpus-based approach to terminography", *Terminography* vol. 3 (1), 27-52.

BOWKER, Lynne y Jennifer PEARSON, 2002: *Working with Specialized Language: a practical guide to using corpora*, London / New York: Routledge.

BUTLER, Christopher (ed.), 1992: *Computers and Written Texts*, Oxford: Blackwell.

CABRÉ, María Teresa, 2003: "Investigar en terminología: posibilidades y línea de trabajo" en Emilio ORTEGA (dir.): *Panorama actual de la investigación en traducción e interpretación* (volumen I), Granada: Atrio, 495-512.

DÍAZ, Marina, 2002: "Marcadores de la cadena textual cadena lógica en un texto científico" en Zinaida LVOVSKAYA (coord.): *La estilística textual: visión traductológica del texto*, Málaga: Encasa, 137-162.

DÍAZ, Marina, 2003: "La cadena lógica en textos de fisioterapia" en Zinaida LVOVSKAYA (dir.): *Convenciones textuales. Estudio traductológico*, Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, 51-66.

FINATTO, Maria José, 2007: "Exploração terminológica com apoio informatizado: perspectivas, desafios e limites" en *Ciências do Léxico* vol. III, Uberlândia: Publicação da Universidade Federal de Uberlândia (UFU).

- FUNG, Pascale, 1995: "Compiling bilingual lexicon entries from a non-parallel English-Chinese corpus", *14th Annual Meeting of Very Large Corpora*, Boston, Massachusetts, 173-183.
- HARTMANN, Reinhard, 1994: "The Use of Parallel Text Corpora in the Generation of Translation Equivalents for Bilingual Lexicography" en *Proceedings of the Euralex Congress*, Amsterdam: VLL, 291-298.
- KRIEGER, Maria da Graça y Maria José FINATTO, 2004: *Introdução à terminologia: teoria e prática*, São Paulo: Contexto.
- LEIČIK, Vladimir, 1993: "Peculiarities of term function in texts" en Jennifer DRASKAU (ed.): *Selected Readings in Russian Terminology Research / International Network for Terminology*, Viena: Termnet, Internat. Network for Terminology, 91-101.
- LEIČIK, Vladimir, 2002: "Elements of Terminological Text Theory" en Serguey SHELOV y Vladimir LEIČIK (eds.): *Russian Terminology Science (1992-2002)*, Viena: Termnet Publisher, 368-386.
- LVOVSKAYA, Zinaida (dir.), 2002: *La estilística textual: visión traductológica del texto*, Málaga: Encasa.
- LVOVSKAYA, Zinaida (coord.), 2003: *Convenciones textuales en textos científicos sobre fisioterapia: estudio textual orientado a la traducción*, Las Palmas de Gran Canaria: Servicio de Publicaciones de la Universidad de Las Palmas de Gran Canaria.
- MALMKJAER, Kirsten, 2003: "On a pseudo-subversive use of corpora in translation training" en Federico ZANETTIN, Silvia BERNARDINI y Dominic STEWART (eds.): *Corpora in Translator Education*, Manchester: St. Jerome, 119-134
- McENERY, Tony y Andrew WILSON, 1996: *Corpus Linguistics*, Edinburgh: Edinburgh University Press.
- MEYER, Ingrid y Kristen MACKINTOSH, 1996: "The Corpus from a Terminographer's Viewpoint", *International Journal of Corpus Linguistics* vol. 1(2), 257-285.
- MONTERO, Silvia y Pamela FABER, 2008: *Terminología para traductores e intérpretes*, Granada: Traga-cantos.
- PEARSON, Jennifer, 1998: *Terms in Context*, Amsterdam / Filadelfia: John Benjamins.
- PÉREZ, Chantal, 2002: *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*, Madrid: CSIC / Elies. Documento disponible en línea [<http://elies.rediris.es/elies18/>], fecha de consulta: 29 de marzo de 2011].
- PIÑERO, Gracia y María Jesús GARCÍA, 2002: "Marcadores de la cadena textual cadena temática en un texto jurídico" en Zinaida LVOVSKAYA (coord.): *La stilística textual: visión traductológica del texto*, Málaga: Encasa, 163-193.
- PIÑERO, Gracia y María Jesús GARCÍA, 2003: "La naturaleza de los marcadores de la cadena temática en los artículos científicos sobre fisioterapia" en Zinaida LVOVSKAYA (coord.): *Convenciones textuales en textos científicos sobre fisioterapia. Estudio textual orientado a la traducción*, Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, 37-49.
- RABADÁN, Rosa y Purificación FERNÁNDEZ, 2002: *La traducción inglés-español: fundamentos, herramientas, aplicaciones*, León: Universidad de León.
- SÁNCHEZ, Elena, 2003: "El texto habla de sí mismo: marcadores textuales en artículos sobre fisioterapia" en Zinaida LVOVSKAYA (dir.): *Convenciones textuales. Estudio traductológico*, Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, 21-36.
- SÁNCHEZ, Elena, 2006: "MYOCOR: creación y explotación de un corpus bilingüe (fr-es) sobre enfermedades neuromusculares", *Confluências. Revista de Tradução Científica e Técnica* 4, 67-83.

SIERRA, Gerardo, 2009: "Extracción de contextos definitorios en textos de especialidad a partir del reconocimiento de patrones lingüísticos", *Linguamática* 2, 13-38.

SIERRA, Gerardo y otros, 2003: "Definitional Contexts Extraction from Specialised Texts" en *PALC 2003 Proceedings: Language, Corpora and E-Learning*, Frankfurt: Peter Lang Publish, 21-31.

SIERRA, Gerardo y otros, 2008: "Definitional Verbal Patterns for Semantic Relation Extraction", *Terminology* 14 (1), 74-98.

SINCLAIR, John McHardy, 1991: *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

SINCLAIR, John McHardy, 1996: *Preliminary recommendation on Corpus Typology*. EAG-TCWG-ATYP / P. Pisa: EAGLES.

TAGNIN, Stella, 2005: *O jeito que a gente diz: expressões convencionais e idiomáticas*, São Paulo: Disal.

TAGNIN, Stella y Elisa GUIMARÃES, 2004: "Lingüística de Corpus e tradução técnica: relato da montagem de um corpus multivarietal de culinária", *Tradterm* 10, 313-358.

TEUBERT, Wolfgang, 2005: "My version of Corpus Linguistics", *International Journal of Corpus Linguistics* 10(1), 1-13.

VARGAS, Chelo, 2006: "Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural", *Revista Debate Terminológico*, 2, París: RITERM (Red Iberoamericana de Terminología).

WALKER, Donald y Robert AMSLER, 1991: "The ecology of language" en *Proceedings of the International Workshop on Electronic Dictionaries*, Tokio: Japan Electronic Dictionary Research Institute, 10-22.

WRIGHT, Sue Ellen, 1993: "The inappropriateness of the merely correct: stylistic considerations in

scientific and technical-translation" en Sue Ellen WRIGHT (ed.): *Scientific and Technical Translation* (American Translators Association Series, Volume VI), Amsterdam: John Benjamins, 69-86.

ZANETTIN, Federico, Silvia BERNARDINI y Dominic STEWART (eds.), 2003: "Corpora in Translator Education: An Introduction" en Federico ZANETTIN, Silvia BERNARDINI y Dominic STEWART (eds.): *Corpora in Translator Education*, Manchester: St. Jerome, 1-13.