



RERCOR: portal de recursos lingüísticos multilingües sobre enfermedades raras

Elena Sánchez Trigo y Tamara Varela Vila¹

Recibido: 25 de septiembre 2018 / Aceptado: 8 de marzo 2019

Resumen. En este artículo se presentan el proceso de creación y las características de RERCOR, un portal de recursos lingüísticos multilingües (francés, español e inglés) sobre enfermedades raras. Este innovador proyecto integra diversas herramientas accesibles vía web sobre un subdominio médico que busca incrementar su visibilidad. RERCOR se estructura en torno a dos aplicaciones que permiten la consulta en corpus y glosarios. Tras las bases teóricas utilizadas, se describen las etapas seguidas para desarrollar el portal, así como las funcionalidades de la interfaz de usuario. Este recurso se dirige fundamentalmente a traductores, intérpretes y redactores para ser utilizado con fines profesionales o investigadores.

Palabras clave: Terminología médica, traducción médica, enfermedades raras, corpus especializados, recursos multilingües.

[en] RERCOR: Portal for Multilingual Linguistic Resources in the Domain of Rare Diseases

Abstract. This article outlines the creative process and the characteristics of RERCOR, a portal for multilingual linguistic resources (French, Spanish and English) in the domain of rare diseases. This innovative project integrates different web-based tools on a medical sub-domain which is looking to raise its profile. RERCOR is structured around two applications which enable consultation in corpora and glossaries. First, the theoretical principles used are described. Then, the stages of the process to develop the portal are described along with the features of the user interface. This resource is essentially designed for translators, interpreters and editors and to be used for professional and research purposes.

Keywords: Medical terminology, medical translation, rare diseases, specialised corpora, multilingual resources.

Sumario. 1. Introducción. 2. Marco teórico. 3. Marco metodológico. 4. Resultados: portal RERCOR e interfaz de consulta. 5. Conclusiones.

Cómo citar: Sánchez Trigo, E. y Varela Vila, T. (2019) RERCOR: portal de recursos lingüísticos multilingües sobre enfermedades raras, en *Estudios de Traducción* 9, 131-150.

¹ Universidade de Vigo (España)
etrigo@uvigo.es / tvarela@uvigo.es

1. Introducción²

RERCOR es un portal de recursos lingüísticos multilingües sobre el subdominio médico de las enfermedades raras (ER). Se trata de un proyecto innovador tanto por el ámbito temático que aborda, de gran actualidad pero poco estudiado, como por la integración de un conjunto de herramientas a través de una interfaz web de libre acceso. El portal presenta una aplicación principal para la consulta de corpus con un analizador de concordancias y otra complementaria de consulta de glosarios. La primera permite realizar búsquedas en tres corpus bilingües (francés y español), de naturaleza comparable y paralela, lematizados y etiquetados. Proporciona líneas de concordancia e información morfosintáctica, colocacional y contextual. Asimismo, posibilita la recuperación selectiva de datos mediante la aplicación de diversos filtros extralingüísticos. La segunda aplicación facilita la consulta de tres glosarios multilingües (francés, español e inglés) que incluyen, además de los términos equivalentes en otra lengua, categoría gramatical, sinónimos o, por ejemplo, notas de uso. De este modo, RERCOR combina dos aproximaciones que se complementan, una de naturaleza analítica, ofrecida por los corpus, y otra sintética, ofrecida por los glosarios.

Las ER constituyen hoy en día —gracias, en gran medida, al trabajo realizado por las asociaciones de pacientes (Carrión et al. 2015; FEDER-OBSER 2017)— un desafío desde el punto de vista médico e investigador. Este amplísimo subdominio incluye más de 6000 enfermedades, generalmente de origen genético, crónicas y degenerativas, que presentan un elemento determinante: su baja prevalencia (cada una de ellas afecta a menos de 1 cada 2000 personas). Si bien de esta última característica se deriva la falta de atención recibida, las ER consideradas de manera conjunta afectan a un número muy importante de personas:

Despite the rarity of each rare disease, it is always surprising for the public to discover that according to a well-accepted estimation, “about 30 million people have a rare disease in the 25 EU countries”, which means that 6% to 8% of the total EU population are rare disease patients. This figure is equivalent to the combined populations of the Netherlands, Belgium and Luxembourg. (EURORDIS 2005: 3).

Padecer una de estas enfermedades es, por tanto, mucho más frecuente de lo que habitualmente se podría considerar. En la actualidad se busca una actuación global que convierta a las ER en una prioridad de las políticas de salud pública mundial. La reciente creación de la plataforma NGO Committee for Rare Diseases en el seno de las Naciones Unidas es una muestra de las iniciativas generadas con esta finalidad: “to elevate rare diseases as a priority within global public health, research, medical care, social care and social integration priority. This, in turn, will serve to advance the cause of people living with rare diseases worldwide, regarding inequalities in access to treatment and care” (NGO Committee for Rare Diseases 2017: s. p.).

² Este trabajo se inscribe parcialmente en el marco de los proyectos ED431C2018/50 de la Xunta de Galicia y TIN2017-85160-C2-2-R del Ministerio de Economía y Competitividad.

Estos datos muestran que RERCOR se centra, como indicábamos, en un ámbito de gran actualidad que busca recibir más atención desde un punto de vista médico y social. El portal tiene como objetivo proporcionar herramientas de información lingüística multilingüe³ para traductores, intérpretes y redactores tanto profesionales como en etapas de formación o bien para ser utilizadas con fines investigadores. Partimos, por un lado, del hecho de que la actividad de estos agentes constituye un elemento fundamental para la difusión del conocimiento sobre este subdominio médico. Por otro, consideramos que la disponibilidad de recursos centrados en ámbitos de conocimiento especializados constituye un elemento fundamental para incrementar su visibilidad y la concienciación pública sobre los mismos.

En este artículo se presentan, en primer lugar, las bases teóricas utilizadas para la creación del portal. A continuación, se describen la configuración y las etapas seguidas para la elaboración de las aplicaciones que lo conforman, así como las principales funcionalidades de la interfaz de usuario diseñada.

2. Marco teórico

Los recursos que integran RERCOR toman como base la lingüística de corpus por su demostrada versatilidad como fuente de datos. En efecto, los corpus son una herramienta clave tanto para realizar análisis empíricos mediante el estudio de muestras reales de la lengua como para la creación de recursos en diferentes lenguas.

Un corpus, de acuerdo con la ya clásica definición de Sinclair (2004), es un conjunto representativo de textos en formato electrónico seleccionados con criterios de naturaleza extralingüística. Las diversas posibilidades de diseño existentes dan lugar a una amplia tipología de corpus que numerosos autores han tratado de organizar en función de su modalidad, tamaño, grado de especialización, lenguas que incluyen o, por ejemplo, período temporal⁴.

Los corpus especializados, que son en los que se centra este trabajo, constituyen una importante fuente de información lingüística de gran ayuda para traducir. No obstante, la disponibilidad de corpus sobre dominios muy especializados o novedosos sigue siendo reducida. A pesar de que en la actualidad existen diferentes herramientas que permiten crear corpus *ad hoc* (Sketch Engine, BootCat o, por ejemplo, Terminus), su compilación exige tiempo y esfuerzo, así como el dominio de las técnicas básicas para la selección y el tratamiento de los textos (Fantinuoli 2016). El portal que presentamos responde a la necesidad de facilitar el acceso a este tipo de recursos para integrarlos en la actividad de traductores y redactores.

En RERCOR se han seguido principalmente las propuestas sobre corpus de EAGLES (1996), realizadas desde la lingüística, así como las formuladas desde el ámbito de los estudios de traducción (Corpas 2001 o Laviosa 2002, por ejemplo). Se ha atendido especialmente a cuestiones como la fiabilidad y la representatividad⁵.

³ Si bien se incluyen recursos en inglés, destacamos la especial atención que se presta en este portal al francés y español, lenguas que necesitan incrementar la disponibilidad de recursos en el ámbito médico.

⁴ En Llisterri (2017) se puede consultar una amplia bibliografía sobre corpus por tipos de publicaciones y temas. Por su parte, en Tognini-Bonelli (2010) se describe la evolución teórica de la lingüística de corpus.

⁵ Entre las diversas aproximaciones cuantitativas para determinar este aspecto, cabe destacar la aplicación informática ReCor, cuyo objetivo es estimar el umbral mínimo de representatividad de un corpus mediante el algo-

Por este motivo, los corpus en los que se basan las herramientas del portal se han compilado atendiendo a los criterios de calidad y cantidad (EAGLES 1996). En relación con este último criterio, debemos señalar que, a pesar de los numerosos debates, todavía no se ha conseguido determinar el número de muestras textuales necesarias para garantizar la representatividad de un corpus. Sin embargo, sí está comúnmente aceptado que los corpus especializados, como los de RERCOR, se caracterizan por ser «small and tidy» (Mair 2006: 356) sin perder representatividad (Bowker y Pearson 2002; Koester 2010). En concreto, se ha confirmado la utilidad de corpus más pequeños y específicos en campos restringidos como sucede, por ejemplo, con numerosos subdominios del ámbito de la medicina (Walsh 2013), que es el que nos ocupa en este trabajo.

La fase de compilación de un corpus es de crucial importancia: “The decisions that are taken about what is to be in the corpus, and how the selection is to be organized, control almost everything that happens subsequently. The results are only as good as the corpus” (Sinclair 1991: 13). Un corpus compilado de forma adecuada puede proporcionar ejemplos fiables y auténticos de diferentes fenómenos lingüísticos, además de información terminológica y colocacional (Zanettin 2002), tan relevante en los textos de especialidad. Por este motivo, la selección de las muestras textuales de los distintos corpus creados estuvo precedida de una etapa de documentación que resulta necesaria para conocer las características del ámbito de estudio y delimitarlo conceptualmente.

Las dos aplicaciones disponibles en el portal RERCOR están basadas en corpus compilados y tratados con finalidades específicas. El objetivo buscado con la primera de ellas (aplicación para consultar corpus) fue crear un analizador de concordancias que permitiese obtener información sobre una palabra o grupo de palabras (contexto, colocaciones, morfosintaxis) en textos del ámbito de las ER. Por este motivo, los corpus compilados se trataron, etiquetaron, lematizaron y, en el caso de los textos paralelos, se alinearon a nivel oracional (Zanettin 2012) para su posterior implementación en el portal. Los corpus constituyen también la base de la segunda aplicación, diseñada para la consulta de glosarios. Se estima que en la elaboración de este tipo de recursos un 70 % del tiempo se dedica al trabajo con corpus (Vargas Sierra 2008).

En la creación de los glosarios se tomó, asimismo, como punto de partida una organización conceptual inicial del ámbito abordado, que posteriormente se completó con la información extraída de los corpus. Seguimos de este modo los postulados de la Teoría Comunicativa de la Terminología, que señalan que las unidades terminológicas poseen un significado preciso en una estructura conceptual (Cabré 2009) y que el conocimiento “solo será especializado si se representa y se transmite respetando escrupulosamente la estructura conceptual del campo en cuestión” (Cabré y Estopà 2005: 69). El vaciado de los corpus en los que basan los glosarios se realizó con el programa WordSmith Tools y para el análisis de la información se prestó especial atención a las propuestas de autores como Nazar y Cabré (2011) o Vargas Sierra (2012). La estructura de cada glosario se estableció teniendo en cuenta principalmente lo indicado por Sager (1990), si bien se adaptaron a los objetivos

ritmo N-Cor que analiza la densidad léxica. Se trata de una medida para determinar *a posteriori* el tamaño mínimo que debe alcanzar un corpus para poder considerarse representativo en términos estadísticos (Corpas y Seghiri Domínguez 2007).

buscados en cada caso. En el apartado siguiente se tratan estas cuestiones de manera más detallada.

La documentación, que es una de las etapas más importantes del proceso traductor, se convierte en el ámbito de la medicina en una fase esencial por el gran número de especialidades y su rápida evolución. En la actualidad existe un volumen ingente de recursos de todo tipo sobre este ámbito. Sin embargo, son muy pocos los que se centran específicamente en las ER. Como excepción, podemos mencionar Orphanet (portal sobre enfermedades raras y medicamentos huérfanos) (2012), que incluye, entre otros, un buscador, una clasificación y una enciclopedia sobre ER.

Los recursos con información de naturaleza lingüística y terminológica son fundamentales en las áreas de especialidad para traductores, intérpretes y redactores. Es precisamente esta información la que proporciona el portal RERCOR. Constituye una aportación innovadora que integra diversas herramientas accesibles vía web en relación con un subdominio caracterizado por la escasez de recursos disponibles.

3. Marco metodológico

El portal se organiza, como se ha indicado, en dos aplicaciones diferenciadas y complementarias, dedicadas respectivamente a la consulta de corpus y glosarios. El proceso de creación de RERCOR se desarrolló fundamentalmente en dos etapas. La primera de ellas, que presentamos en este apartado, se centró en cuestiones metodológicas relativas a la selección y tratamiento de los contenidos que configuran las aplicaciones del portal. La segunda etapa, que se aborda en el apartado siguiente, se centró en el diseño de la interfaz de usuario para permitir el acceso a la información.

3.1. Configuración de la aplicación para consultar corpus

Esta aplicación, que es la principal, se diseñó para permitir la consulta, tanto de manera conjunta como individual, de los corpus que se incluyen. El objetivo era ofrecer asimismo la posibilidad de seleccionar una serie de parámetros para crear subcorpus virtuales adaptados a las necesidades del usuario (p. ej., consultas solo en un género o en un período temporal determinado). El sistema de recuperación de datos desarrollado proporciona concordancias e información morfosintáctica, colocacional o contextual.

La aplicación está conformada por los corpus MYOCOR 2.0, EMCOR y ERCOR. Se trata de corpus bilingües (francés-español), comparables y paralelos, lematizados y etiquetados:

Tabla 1. Datos globales de los corpus

Tamaño	2 161 038 palabras (1 075 158 en francés y 1 085 880 en español)
Textos	712 textos (360 en francés y 352 en español)
Período	1993-2017

En conjunto contienen muestras textuales pertenecientes a 8 géneros y 9 subgéneros diferentes. La mayor parte de los textos (71,95 %) son de nivel especializado o semiespecializado (66,43 % y 15,52 %, respectivamente). Si bien en un porcentaje menor, se incluyen también textos de nivel divulgativo (18,05 %). Se consideró adecuado incluir muestras de diferentes grados de especialización para que estén representados los diferentes tipos de datos lingüísticos (Vargas Sierra 2006).

De manera individualizada, el primero de los corpus indicados, MYOCOR 2.0, está integrado por textos en francés y en español producidos entre 1993 y 2017, un período temporal de 25 años⁶:

Tabla 2. Características del corpus MYOCOR 2.0

Nombre	MYOCOR 2.0
Temática	enfermedades neuromusculares (ENM) ⁷
Tipo de corpus	comparable (70 %) / paralelo (30 %)
Género	multigenérico
Tamaño	1 077 508 palabras (513 872 en francés y 563 636 en español)

Si bien mayoritariamente se trata de un corpus comparable, presenta asimismo un 30 % de textos paralelos. Incluye 8 géneros textuales con distinto grado de especialización:

Tabla 3. Géneros incluidos en el corpus MYOCOR 2.0

artículos científicos	conferencias/comunicaciones en congresos
guías de práctica clínica	guías de información sobre salud
monografías	manuales
convenios	noticias sobre aspectos médicos

El segundo corpus incluido en la aplicación, EMCOR⁸, también de naturaleza multigenérica, se compone de textos producidos entre 2001 y 2012. Está integrado por muestras textuales en francés y español mayoritariamente de nivel especializa-

⁶ Se trata de una segunda versión revisada y ampliada, tanto en número de palabras y géneros como cronológicamente, del corpus MYOCOR, inicialmente creado en el marco de un proyecto de elaboración de un buscador sobre enfermedades neuromusculares (Miquel Verges y Sánchez Trigo 2010).

⁷ Las ENM son un conjunto de más de 200 enfermedades neurológicas, progresivas y en su mayoría de origen genético, cuya principal característica es la pérdida de fuerza muscular. Constituyen el mayor grupo de ER, aproximadamente un 30 % de las mismas, y se estima que afectan a 1 de cada 1000 recién nacidos, por lo que se calcula que en la Unión Europea existen unas 200 000 personas con una ENM.

⁸ EMCOR se creó originalmente para realizar un estudio terminológico sobre enfermedades metabólicas. En su primera versión estaba compuesto por 268 textos y 458 718 palabras (Varela Vila y Sánchez Trigo 2012). Posteriormente se amplió con el fin de llevar a cabo un análisis conceptual y terminológico más profundo (Varela Vila 2015 y Varela Vila y Sánchez Trigo 2020).

do (91,3 %), concretamente artículos científicos, pero también por guías de información sobre salud de nivel semiespecializado (6,1 %) y divulgativo (2,6 %):

Tabla 4. Características del corpus EMCOR

Nombre	EMCOR
Temática	errores innatos del metabolismo (EIM) ⁹
Tipo de corpus	comparable
Género	multigénérico
Tamaño	842 556 palabras (418 092 en francés y 424 464 en español)

Por su parte, ERCOR, el tercero de los corpus, contiene guías de práctica clínica en francés y español publicadas entre 2007 y 2013 (Marta Rodríguez 2015). Es, por lo tanto, monogénérico y con textos de nivel especializado:

Tabla 5. Características del corpus ERCOR

Nombre	ERCOR
Temática	guías de práctica clínica (sobre ER)
Tipo de corpus	comparable
Género	monogénérico
Tamaño	240 974 palabras (143 194 en francés y 97 780 en español)

Todas las muestras textuales se compilaron en formato electrónico, mayoritariamente en PDF o HTML, y se procesaron para convertirlas a texto plano (.txt) con codificación UTF-8. Con la finalidad de obtener una correcta segmentación, que no siempre se consigue automáticamente, el procesamiento se llevó a cabo de forma semiautomática. Asimismo, se efectuó una revisión manual para suprimir encabezados y pies de página repetidos, referencias bibliográficas o tablas con contenido exclusivamente numérico.

Dado que, como se ha indicado, una parte del corpus MYOCOR 2.0 está compuesta por textos paralelos, se empleó el editor InterText (Vondřička 2014) para su alineación a nivel oracional en formato XML. Este programa permite importar los archivos en ambas lenguas y ofrece una opción de alineado automático, que es la que se utilizó en primera instancia. Posteriormente se realizó una revisión manual de dicha alineación para corregir posibles errores derivados de la no correspondencia de segmentos entre el texto original y la traducción (cambio de orden de frases, segmentos no traducidos, etc.).

Con el fin de que en esta aplicación para consultar corpus pudiesen realizarse búsquedas utilizando etiquetas morfosintácticas y lemas, fue necesario etiquetar y lema-

⁹ Los EIM son un grupo muy heterogéneo de ER monogénicas, habitualmente graves, que causan una afectación multiorgánica. Actualmente se conocen más de 700 EIM y se estima que pueden afectar a 1 de cada 500 recién nacidos.

tizar los textos. Para ello se utilizó el programa Freeling (Padró y Stanilovsky 2012), un etiquetador de código abierto que incluye, entre otras muchas, las lenguas de los corpus de este trabajo. Se realizó un etiquetado automático con la versión 4.0 de esta herramienta y se empleó el entrenamiento estándar para el francés y el español.

3.2. Configuración de la aplicación para consultar glosarios

Esta segunda aplicación permite acceder a diferentes recursos terminológicos multilingües (francés, inglés y español) que se elaboraron y trataron para, también en este caso, poder ser consultados tanto de manera conjunta como individual. Recogen información de tipo descriptivo, ya que su objetivo es proporcionar una visión global de la terminología usada en este subdominio.

Si bien en la interfaz de consulta del portal, por cuestiones de espacio, esta aplicación se denomina *Glosarios*, debemos precisar que está constituida por dos vocabularios y un glosario. Todos ellos incluyen terminología del ámbito estudiado junto con sinónimos, equivalentes y otras denominaciones, entre otros datos. Por su parte, el glosario presenta, además, definiciones de los términos que recoge.

Como ya hemos apuntado, la elaboración de los recursos que integran esta segunda aplicación incluyó unas fases previas de trabajo con corpus. Así, una vez compilado, se utilizó para su explotación el programa WordSmith Tools, de especial utilidad para identificar los candidatos a término en sus contextos de aparición. Se analizaron de este modo las listas de palabras más frecuentes y los patrones sintácticos con mayor recurrencia. A continuación, con la terminología extraída en cada lengua se completó la organización conceptual inicialmente creada¹⁰. Finalmente, mediante la combinación de las organizaciones en las diferentes lenguas, se elaboraron los recursos que se incluyen.

El primero de ellos, el *Vocabulario francés-español de errores innatos del metabolismo*, está compuesto por 1420 entradas y toma como punto de partida un vocabulario más reducido y una base de conocimiento terminológico previamente creados¹¹. Dado que la temática de ambos era común y la información contenida complementaria, se unificaron en un único recurso más completo con la siguiente estructura: entrada y sinónimos en francés, equivalente y otras denominaciones en español, así como notas de uso. La extracción terminológica se realizó utilizando el corpus EMCOR, al que ya nos hemos referido en el apartado anterior.

El segundo recurso, el *Vocabulario inglés-español de productos de apoyo*, recoge 212 entradas en inglés con equivalente en español y otras posibles denominaciones en esta misma lengua. Incluye la terminología extraída de un corpus paralelo sobre productos de apoyo utilizados por personas con enfermedades neuromusculares. La integración en RERCOR de este recurso creado con anterioridad (Varela Vila, Sánchez Trigo y Marta Rodríguez 2012), conllevó una revisión del mismo y su introducción en una base de datos conjunta.

¹⁰ En relación con la complejidad de organizar conceptualmente la terminología del ámbito de las ER, remitimos a Varela Vila y Sánchez Trigo (2018).

¹¹ Sobre este vocabulario inicial y sobre la base de conocimiento terminológico ONTERMET, con terminología sobre los errores innatos del metabolismo (EIM), sus síntomas y signos, ver respectivamente Varela Vila et al. (2011), Varela Vila (2015) y Varela Vila y Sánchez Trigo (2020).

Por último, el *Glosario francés-español-inglés de enfermedades neuromusculares* incluye 114 entradas con los siguientes campos: entrada y sinónimos en francés, categoría gramatical, género, número, definición, abreviaciones y equivalentes en español y en inglés. Se basa en la explotación de un corpus de artículos científicos (Varela Vila 2009).

La integración de estos diferentes recursos en la aplicación para poder ser consultados de forma conjunta conllevó además el tratamiento y procesamiento informático de los mismos. En primer lugar, se identificaron los campos comunes y se elaboró un esquema de las relaciones de equivalencia entre ellos. A continuación, a partir de dicho esquema, se diseñó una base de datos en la que se introdujeron las entradas de los distintos recursos y se activaron los campos necesarios para cada uno de ellos.

4. Resultados: portal RERCOR e interfaz de consulta

Las aplicaciones descritas pueden consultarse en el portal de libre acceso RERCOR, disponible en <http://www.rercor.org>. En la barra de navegación de la web figuran las pestañas para acceder a ambas aplicaciones, así como una breve presentación del portal. Se incluyen también tablas con los datos globales más relevantes y tutoriales en vídeo para facilitar las consultas.

4.1. Descripción de la interfaz para consultar corpus

En la pestaña *Corpus* se accede a un formulario de consulta organizado en cinco secciones principales: a) *Búsqueda*, b) *Resultado*, c) *Sensibilidad*, d) *Filtros* y e) *Texto*. Estas secciones contienen, a su vez, una serie de campos con menús desplegables para permitir al usuario una recuperación selectiva de la información aplicando diversos parámetros en función de sus intereses.

Figura 1. Secciones (destacadas con círculo) y campos (subrayados) del formulario de consulta

The screenshot shows the RERCOR search interface with the following sections and fields:

- Section a) Búsqueda:** Corpus: Todos (3) ▾; Tipo: Pal. ortográficas ▾
- Section b) Resultado:** Tipo: Concordanancias ▾; Ordenación: Coincidencia ▾; Tamaño página: 50 ▾
- Section c) Sensibilidad:** Acentos: Sí ▾; Mayúsculas: Sí ▾
- Section d) Filtros:** Género: Cualquiera ▾; Desde: 1993 ▾; Hasta: 2017 ▾; Idioma: Todos ▾; Especialización: Cualquiera ▾
- Section e) Texto:** Cincio palabras máximo

Buttons at the bottom: Volver, Descargar, Limpiar, Buscar

Las consultas pueden hacerse en cualquiera de las lenguas que se incluyen y los resultados que se obtienen son líneas de concordancia que, además, proporcionan: acceso a fragmentos más amplios del texto, información sobre el corpus del que pro-

ceden o, por ejemplo, datos estadísticos. A continuación se presentan de forma más detallada las múltiples posibilidades de consulta que ofrecen las diferentes secciones y campos diseñados, así como algunos ejemplos de los resultados que genera la aplicación.

a) Sección *Búsqueda*: incluye los campos *Corpus* y *Tipo*. El primero permite decidir si la consulta se hará en todos los corpus de la aplicación (MYOCOR 2.0, EMCOR y ERCOR) o bien solo en alguno(s) específicamente. Por su parte, en el campo *Tipo* se puede seleccionar qué elemento buscar entre las cuatro opciones siguientes:

Tabla 6. Opciones disponibles en el campo *Tipo*

<i>Palabras ortográficas</i>	palabra o secuencia de hasta 5 palabras
<i>Elementos gramaticales</i> ¹²	elemento gramatical o secuencia de hasta 5 elementos gramaticales (permite también realizar búsquedas por lema, etiqueta y/o palabra ortográfica)
<i>Palabras ortográficas próximas</i>	distancia de entre 1 y 10 palabras
<i>Elementos gramaticales próximos</i>	distancia de entre 1 y 10 elementos gramaticales

Las dos últimas opciones indicadas son de especial interés para conocer el contexto inmediato de las formas consultadas. Permiten, por ejemplo, buscar las ocurrencias de *déficit* y *progressif* en contextos en los que ambas figuren a no más de 5 palabras de distancia. En este caso, RERCOR ofrece resultados como: *déficit visuel progressif*, *déficit progressif des extenseurs des doigts* o, por ejemplo, *déficit musculaire symétrique et progressif*. Las líneas de concordancia que se obtienen posibilitan analizar las distintas combinaciones de estas formas y las colocaciones más habituales en francés. Se puede hacer la misma búsqueda en español con, en este caso, resultados como: ‘déficit motor progresivo’ o ‘déficit muscular progresivo’. Se trata de una información muy útil para resolver, por ejemplo, problemas de traducción de naturaleza colocacional. Además, pueden completarse con otros datos también ofrecidos por la aplicación y que se describen a lo largo de este apartado.

b) Sección *Resultado*: incluye tres campos (*Tipo*, *Ordenación* y *Tamaño página*) que proporcionan diferentes posibilidades para el acceso y visualización de los datos.

En el primero de ellos, *Tipo*, figuran las siguientes opciones de consulta:

- *Concordancias*: se ofrece por defecto, dada la importancia de su estudio en el análisis de corpus textuales (Fantinuoli 2016). Si se escoge esta opción,

¹² Los elementos gramaticales se crean a partir del análisis del texto y no siempre coinciden con las palabras ortográficas, ya que estas pueden estar formadas por varios elementos gramaticales. Por ejemplo, en español, la palabra ortográfica ‘administrarse’ se descompone en dos elementos gramaticales: ‘administrar’ y ‘se’, cada uno con su correspondiente etiqueta morfosintáctica y lema. En francés, *s'enraidit* se descompone en el pronombre personal *se* y el verbo en presente de indicativo *enraidit*.

los resultados de la consulta realizada por el usuario se presentan en líneas de concordancia en las que cada ocurrencia figura destacada en negrita junto con la siguiente información: fragmento del contexto anterior y posterior, datos del corpus del que procede, idioma y año¹³. Asimismo, un enlace situado al inicio de cada concordancia permite acceder al párrafo completo, párrafos anteriores y posteriores, y, si se trata de un corpus paralelo, al párrafo en la otra lengua. Los resultados obtenidos pueden exportarse a un archivo en formato CSV.

Figura 2. Concordancias de ‘miopatía de’

53	MYOOCR / es / 2004	Se trata de una forma benigna de	miopatía de Duchenne .
54	MYOOCR / es / 2003	Otras denominaciones: miopatía distal de tipo 1, MPD1	miopatía de Gowers .
55	EMOOR / es / 2006	e) Deficiencia del complejo V: Se manifiesta como una	miopatía de inicio congénito y lenta progresión con FRR en la biopala .
56	MYOOCR / es / 2003	La anomalía genética implicada en la	miopatía de Laing se ha localizado en el cromosoma 14, en 14q11.2-q13.
57	MYOOCR / es / 2003	Otras denominaciones: enfermedad de Landouzy-Déjerine, FSH, miopatía facioescapulohumeral	miopatía de Landouzy-Déjerine .
58	MYOOCR / es / 2002	Pueden permitir discriminar entre una	miopatía de las cinturas y una FSH (en la que el glúteo medio y el psoas no están afectados).
59	MYOOCR / es / 2001	■ el 27% con Becker, el 18% con Steinert, el 14% con distrofia muscular facioescapulohumeral, el 10% con	miopatía de las cinturas , el 8% con Duchenne, el 7% con amiotrofia espinal infantil y el 16% con
60	MYOOCR / es / 2003		La miopatía de los bastones es una enfermedad genética hereditaria que presenta diversas formas
61	MYOOCR / es / 2014		La miopatía de Miyoshi es una distrofia muscular alélica del déficit de disferlina, que cursa con deb
62	MYOOCR / es / 2003	El gen de la disferlina (en el cromosoma 2 en 2p12-14) también es determinante en una miopatía distal, la	miopatía de Miyoshi .
63	MYOOCR / es / 2017	• El mismo gen está también relacionado con una miopatía distal, la	miopatía de Miyoshi
64	MYOOCR / es / 2017		• La miopatía de Miyoshi y la LGMD2B pueden estar presentes en una misma familia
65	MYOOCR / es / 2003		La miopatía de Miyoshi se debe a una anomalía del gen de la disferlina, una proteína del sarcolema
66	MYOOCR / es / 2003		La miopatía de Miyoshi comienza a manifestarse por una afectación de los músculos de la pantorri
67	MYOOCR / es / 2017	• Ocurre lo mismo con la	miopatía de Miyoshi y la MDAT.
68	MYOOCR / es / 2003		La miopatía de Nonaka está asociada a la alteración del gen (GNE) [UDD:Nonaka]glucosaminase-2-epitr

- *Frecuencia simple*: proporciona el número de ocurrencias de la búsqueda realizada, la frecuencia normalizada (por millón de palabras) y el número de documentos en los que aparece. Estos datos son de utilidad, por ejemplo, para conocer las diferencias en la frecuencia de uso de una palabra o secuencia en distintas lenguas.
- *Frecuencia completa*: complementa la información de la opción anterior proporcionando tablas con datos estadísticos por género textual, grado de especialización e idioma. En la siguiente figura se muestra un ejemplo de los resultados que se obtienen seleccionando esta opción al realizar la búsqueda del epónimo ‘Duchenne’ como palabra ortográfica en todos los corpus en ambos idiomas. Como se puede observar, las tablas indican la distribución de este epónimo en los corpus, lo que permite valorar si se trata de una forma de uso habitual o no, o, por ejemplo, en qué géneros aparece.

¹³ En el menú contextual que aparece al pasar el cursor sobre esta información se ofrecen más datos sobre el texto en el que figura la ocurrencia: título, idioma, género textual, año y grado de especialización.

Figura 3. Frecuencia completa del epónimo ‘Duchenne’

Géneros textuales				Especialización			
	Coincidencias	Documentos	Freq. norm.		Coincidencias	Documentos	Freq. norm.
Artículo científico	51 / 884.830	8 / 340	58/millón	Divulgativo	132 / 390.118	28 / 237	338/millón
Conferencia/comunicación en congreso	16 / 17.971	2 / 6	890/millón	Especializado	117 / 1.435.524	19 / 394	82/millón
Convenio	2 / 18.446	2 / 2	108/millón	Semiespecializado	100 / 335.396	12 / 81	298/millón
Guía de información sobre salud	163 / 507.649	29 / 291	321/millón	Idiomas			
Guía de práctica clínica	0 / 253.372	0 / 32	0/millón		Coincidencias	Documentos	Freq. norm.
Manual	4 / 147.587	2 / 4	27/millón	español	191 / 1.085.880	30 / 352	176/millón
Monografía	79 / 306.948	8 / 10	257/millón	francés	158 / 1.075.158	29 / 360	147/millón
Noticia sobre aspectos médicos	34 / 24.235	8 / 27	1.403/millón				

Por su parte, el segundo de los campos de esta sección *Resultado*, que como hemos indicado que se denomina *Ordenación*, permite elegir cómo se quieren visualizar las concordancias. En función de sus intereses, el usuario puede ordenarlas atendiendo a una de estas posibilidades:

Tabla 7. Criterios de ordenación de concordancias

<i>Orden alfabético de:</i>	<ul style="list-style-type: none"> - palabra buscada (en desplegable, opción <i>Coincidencia</i>) - primera palabra a la izquierda - segunda palabra a la izquierda - primera palabra a la derecha - segunda palabra a la derecha
<i>Año</i>	
<i>Género textual</i>	
<i>Idioma</i>	
<i>Grado de especialización</i>	

Finalmente, en el campo *Tamaño página* se puede seleccionar el número de concordancias que se mostrarán por página (10, 50, 100 o 500).

c) *Sección Sensibilidad*: permite parametrizar la búsqueda teniendo en cuenta los acentos y las mayúsculas o no, lo que aporta mayor precisión a los resultados. Esta opción es interesante, por ejemplo, para averiguar si una palabra se utiliza mayoritariamente con mayúsculas o sin ellas (‘Apgar’/’APGAR’), o bien para buscar todas las ocurrencias de una palabra, independientemente de si se acentúa o no (‘oftalmoplejía’/’oftalmoplejía’).

d) *Sección Filtros*: sirve para restringir las búsquedas a determinados textos, seleccionando: género textual, subgénero, grado de especialización, año e idioma. En caso de no utilizar estas opciones, la búsqueda se realiza por defecto en todos los textos de los corpus seleccionados. Se trata de una utilidad que permite comprobar, por ejemplo, si un determinado rasgo lingüístico es común a distintos géneros y grados de especialización, así como su comportamiento en diferentes lenguas.

e) *Sección Texto*: tras haber seleccionado las características de la consulta en las secciones anteriores, es en la caja de búsqueda de esta sección donde el usuario introduce la palabra o secuencia concreta que desea consultar. Ofrece la posibilidad de utilizar comodines, lo que constituye una ventaja porque proporcionan flexibilidad. Por otra parte, si la consulta se hace por elemento gramatical (como ya hemos indicado, es una de las opciones en el campo *Tipo* de la sección *Búsqueda*), la interfaz permite, además, refinar las búsquedas combinando los cuatro parámetros que se recogen en la figura 4 (*Elemento gramatical*, *Etiqueta*¹⁴, *Lema* y *Palabra ortográfica*). De este modo se puede obtener información precisa sobre un determinado aspecto lingüístico (cómo se construye un término compuesto, qué preposiciones se emplean en una construcción determinada, qué verbos suelen acompañar a un nombre, etc.).

Figura 4. Búsqueda por varios elementos gramaticales y muestra parcial de concordancias

Elementos gramaticales

Elemento gramatical

Elemento gramatical

Elemento gramatical

V* ?

SP* ?

N* ?

traiter

Lema

Lema

Palab. ortográfica

Palab. ortográfica

Palab. ortográfica

Elemento gramatical

Etiqueta

Lema

Palabra ortográfica

Volver Descargar Limpiar Buscar

Resultados 1 a 50 de 80

Lorsque l'enfant est traité à domicile par AINS seuls, sa famille doit être en mesure de pouvoir joindre un médecin.

Le patient 6 dont l'observation a été publiée, a été traité avec succès par l'administration de mannose.

ple, un cours pour dirigeants handicapés, organisé en 1997 par le Centre_pour_la_Vie_Indépendante_de_Dublin (Irlande) a traité de sujets généraux tels que Les_Différents_Types_de_Handicap ou Les_Droits de les Pe

Lors_d'un programme de formation, le coût doit être traité en termes de conception de stratégies de financement et d'influence exercée sur le m

et en imagerie cérébrale d'un patient atteint de maladie de Fabry avec atteinte de le système nerveux central. traité par alpha-galactosidase A recombinante (agalactase bêta, 1_mg/kg/14 jours) depuis

Il sera confirmé par un examen tomodensitométrique et traité par chirurgie pour éviter toute récurrence.

Le patient était traité par corticoïdes et rééducation musculaire.

Un cas comparable a été rapporté par Haikio et al. chez un patient de 25 ans. traité par enclouage centromédullaire.

Il était traité par greffes de cellules de foie fœtales entre 1985 et 1987, sans amélioration clinique r

de 17-hydroxyprogestérone 20 ng / ml (60,6 nmol / l) et de Δ4-androstènedione_1 ng / ml est insuffisamment traité par hydrocortisone.

En la figura anterior se recoge una muestra de los resultados obtenidos al buscar el verbo *traiter* (en cualquiera de sus formas) seguido de una preposición y de un nombre ('*traiter* + PREP + NOM'). El objetivo es determinar la preposición que se emplea en francés con este verbo para indicar el medicamento o técnica utilizados al tratar a un paciente o una enfermedad. Un análisis de los resultados permite concluir que las ocurrencias que registran una mayor frecuencia en los corpus de RER-COR se construyen con la preposición *par* (p. ej., *traité par placebo*, *traités par corticoïdes et/ou immunosuppresseurs* o *traités par thérapie enzymatique*). Esta consulta se completó buscando el patrón '*traiter* + PREP + DET + NOM' para comprobar si al incluir un determinante aparecían otras construcciones. Los resultados obtenidos, aunque en conjunto son mucho menos numerosos, presentan de nuevo la preposición *par* (p. ej., *traitée par un lavage pulmonaire total* o *traitées par l'ajustement de l'alimentation*), pero se recogen también ocurrencias con la preposición

¹⁴ Para facilitar las consultas, este parámetro incluye un selector con las categorías de palabras y otras características vinculadas a estas (p. ej., verbo, tipo de verbo, modo, tiempo, persona, número o género). De esta forma se evita que el usuario tenga que recordar el etiquetario del corpus.

avec (p. ej., *traités avec l'alpha-glucosidase humaine recombinante* o *traitées avec l'allopurinol*). La misma búsqueda en español revela que, frente a lo que sucede en francés, todas las ocurrencias que coinciden con el significado estudiado se construyen con la preposición 'con' (p. ej., 'tratada con corticoides', 'tratado con óxido nitroso' o 'tratadas con estatinas'). Este tipo de comparaciones intra e interlingüísticas pueden ayudar a solucionar dificultades puntuales surgidas en el proceso de redacción y traducción.

Además de combinados, como en el ejemplo anterior, los cuatro parámetros disponibles en la sección *Texto* (*Elemento gramatical*, *Etiqueta*, *Lema* y *Palabra ortográfica*) también pueden emplearse de forma independiente. Así, para analizar los patrones sintácticos más recurrentes es especialmente útil el parámetro *Etiqueta*. Estos patrones facilitan la recuperación de los términos compuestos y de la fraseología del corpus, cuyo conocimiento es indispensable para redactar textos especializados con corrección y precisión. En la siguiente figura se muestran como ejemplo los resultados obtenidos al consultar el patrón 'NOM + ADJ + ADJ + ADJ' en español. Se localizaron 214 coincidencias, que incluyen términos como 'afectación selectiva tibial anterior', 'alteración genética autosómica recesiva' o 'alteración neurológica orgánica irreversible'.

Figura 5. Búsqueda de patrones sintácticos

Elementos gramaticales

Elemento gramatical	N*	?
Elemento gramatical	A*	?
Elemento gramatical	A*	?
Elemento gramatical	A*	?

parámetro *Etiqueta*

Lema	Palab. ortográfica

Volver Descargar Limpiar Buscar

Resultados 1 a 50 de 214

Anterior 1 2 3 4 5 Siguinte

1 Ir a la página

La actividad inflamatoria articular persistente produce daño en el desarrollo muscular con retraso en el crecimiento, así como lazos radiográficos : radiografía PA-L de tórax (Figura 1 A-B) : en el parénquima pulmonar se observa una afectación intersticial reticular difusa de predominio en campos medios e inferiores ; TCAR (Aquillon_Toshiba de 64 de

* Inicio infantil con una afectación selectiva tibial anterior : debilidad de los extensores de los dedos de la mano (signo característico de " [])

Se trata de un trastorno de almacenamiento lisosómico autosómico recesivo .

La enfermedad de Farber es una alteración genética autosómica recesiva de almacenamiento lisosomal causada por deficiencia de ceramidasa y asociada a [] años de la vida , de tal modo que antes de los 10 años de edad la hiperferritinemia da lugar a alteraciones neurológicas orgánicas irreversibles , mientras que después de la pubertad las alteraciones neurológicas consisten en buscar la anomalía específica de la FSHD1 en el brazo largo de el cromosoma 4 a partir de un análisis sanguíneo estándar previo (análisis rutinario)

Estos son solo algunos de los múltiples tipos de búsquedas que se pueden realizar en esta aplicación de consulta de corpus. Su diseño permite que sea fácil de usar para un usuario sin conocimientos sobre programas de análisis de corpus textuales.

4.2. Descripción de la interfaz para consultar glosarios

La pestaña *Glosarios* del menú principal permite acceder a esta aplicación específica en la que se incluyen los recursos descritos más arriba (ver 3.2): *Vocabulario francés-español de errores innatos del metabolismo*, *Vocabulario inglés-español de productos de apoyo* y *Glosario francés-español-inglés de enfermedades neuromusculares*.

Figura 6. Formulario de consulta de glosarios

abductor pommel	Es	bloque de abducción	
	Glosario	Vocabulario inglés-español de productos de apoyo	
absence de leucocytes	Es	ausencia de leucocitos	
	Glosario	Vocabulario francés-español de errores innatos del metabolismo	resultados
absence de puberté → impubérisme → absence de développement pubertaire	Es	infantilismo sexual	Otras denominaciones en español ausencia de pubertad, impuberismo
	Glosario	Vocabulario francés-español de errores innatos del metabolismo	

En el menú desplegable que figura en la parte superior de la interfaz se pueden seleccionar búsquedas en un recurso concreto o en todos a la vez. Se proporcionan, igualmente, dos posibles formas de consulta: clicando en el listado alfabético que se incluye o introduciendo un término en la caja de búsqueda. Los resultados se presentan siempre organizados alfabéticamente, y se indica de qué recurso concreto procede cada entrada.

En relación con la primera forma de consulta, si se selecciona en el desplegable solo uno de los recursos disponibles, por ejemplo, el *Vocabulario inglés-español de productos de apoyo*, y se marca la letra ‘p’ en el listado alfabético, se obtienen todos los términos incluidos en el vocabulario que comienzan por esta letra: *padded sling*, *page-up*, *paper stand*, *P-Bag*, *pelvic strap*, etc. Esta opción resulta útil cuando no se conoce con exactitud el término que se está buscando o cuando se desea obtener una visión global del contenido del glosario.

Si se prefiere optar por la segunda opción de consulta indicada, introducir un término, y ahora, frente al caso precedente, realizar la búsqueda en el conjunto de los recursos, la aplicación hace un rastreo global en todos los campos que se incluyen (entrada, sinónimos, equivalentes, otras denominaciones, etc.), con independencia de las lenguas y de la direccionalidad, y destaca las coincidencias en color verde. De este modo se facilita la consulta, ya que no es necesario indicar en qué lengua se desea buscar.

Como ejemplo de esta opción de consulta, en la figura anterior se recogen los resultados de la búsqueda ‘anemia’ en el conjunto de glosarios. En los dos primeros resultados (1 y 2), esta forma se localizó en el campo de equivalentes en español (‘anemia sideroblástica’ y ‘anemia sideroblástica refractaria’). En el tercero (3), sin embargo, figura en el campo de otras denominaciones (‘anemia hipocelular’), que como hemos indicado se recogen también en las entradas. De este modo se han obtenido diferentes términos poliléxicos en español en los que aparece la forma ‘anemia’. Asimismo, los resultados incluyen las denominaciones equivalentes en francés (4), *anémie sidéroblastique*, *anémie sidéroblastique réfractaire* y *oligocythémie*, dado que en este caso los resultados proceden del *Vocabulario francés-español de*

errores innatos del metabolismo (5). Un análisis detallado permite comprobar, además, la existencia del término ‘oligocitemia’ en español.

Figura 7. Búsqueda en glosarios

Termo	Traducción	Notas
anémie sidérolastique	Es anemia sideroblástica	equivalentes en español (1 y 2)
anémie sidérolastique réfractaire	Es anemia sideroblástica refractaria	recurso de procedencia (5)
oligocythémie	Es oligocitemia	Otras denominaciones en español: anemia hipocelular, otras denominaciones (3)

Como se puede comprobar, esta opción de consulta es especialmente útil para localizar términos poliléxicos que contienen una misma forma. Podemos ilustrarlo con otro ejemplo ahora en inglés y español, y realizando búsquedas complementarias. Si se introduce el término *sling* (‘arnés’), se obtienen todas aquellas entradas en las que figura, como *hammock sling*, *mesh sling*, *non-slip sling*, *padded sling*, *sheepskin sling*, etc. Si, a continuación, se busca en español ‘arnés’, dado que, como se ha indicado, la aplicación rastrea todos los campos de los glosarios, se obtienen términos como *bib harness*, *diagonal harness*, *hammock sling*, etc. porque en sus equivalentes en español figura la palabra ‘arnés’: ‘arnés de tronco’, ‘arnés diagonal’ y ‘arnés tipo hamaca’. La comparación de los resultados permite comprobar que los términos ingleses *sling* y *harness* suelen traducirse al español en este subdominio como ‘arnés’.

Por otra parte, este modo de búsqueda por término facilita también la localización de entradas relacionadas. Por ejemplo, si se busca *motoneurone*, no solo se obtendrá la entrada *maladie du motoneurone*, sino también otras como *amyotrophie spinale* o *paralyse bulbaire progressive*, que incluyen la palabra *motoneurone* en su definición.

Las características de la interfaz confieren rapidez y versatilidad a las consultas de los usuarios, por lo que resulta de utilidad en el proceso de documentación terminológica sobre el campo de las ER. El hecho de que se recoja un número elevado de sinónimos y equivalentes, así como la inclusión de notas de uso, hace que esta aplicación, combinada con la de consulta de corpus, facilite un mejor conocimiento y uso de la terminología propia de este campo, imprescindible para redactar textos sobre ER.

5. Conclusiones

En este trabajo se han descrito las características y las etapas seguidas para la creación de RERCOR, un portal de recursos sobre ER, un subdominio novedoso y de actualidad. Su objetivo es contribuir al desarrollo de recursos multilingües para la

traducción especializada en inglés, francés y español. El portal, integrado por dos aplicaciones, constituye, además, una aportación de especial interés en relación con estas dos últimas lenguas por el bajo número de recursos disponibles sobre las mismas en el ámbito médico.

La aplicación de consulta de corpus, que incluye más de dos millones de palabras: a) permite el acceso a corpus especializados y la creación de subcorpus virtuales de acuerdo con los intereses de los usuarios; b) favorece una consulta intuitiva y próxima a la de los motores de búsqueda; c) proporciona datos estadísticos que permiten valorar cuantitativa y cualitativamente los resultados obtenidos; d) posibilita por medio del etiquetado el análisis morfosintáctico y la recuperación selectiva de la información; e) da acceso a los textos originales de manera parcial para examinar el contexto de las ocurrencias; f) facilita el estudio de las coocurrencias mediante el análisis de las concordancias y las opciones de búsqueda por proximidad. Esta aplicación reúne, por tanto, los principales criterios que, de acuerdo con Fantinuoli (2016), deben cumplir los programas de corpus específicamente diseñados para traductores, que son los principales destinatarios del portal.

Por su parte, la aplicación de consulta de glosarios, con cerca de dos mil entradas y concebida como complementaria, aporta información terminológica a la proporcionada por el análisis de concordancias. La utilización combinada de ambas aplicaciones permite obtener unos resultados más detallados.

Para finalizar, es importante destacar que la interfaz de consulta de ambas aplicaciones es clara y no requiere conocimientos informáticos avanzados por parte del usuario. Del mismo modo, el diseño del portal simplifica el proceso de actualización y ampliación. En efecto, nuestro objetivo es dotar a RERCOR de nuevos corpus que permitan una mejor descripción de los distintos subdominios que conforman el campo de las ER, crear nuevos glosarios y ampliar los existentes con nuevas entradas y más información de utilidad para los destinatarios a los que se dirige.

Referencias bibliográficas

- Bowker, L. y Pearson, J., *Working with specialized language: a practical guide to using corpora*. Londres: Routledge 2002.
- Cabré, M. T., «La Teoría Comunicativa de la Terminología, una aproximación lingüística a los términos», *Revue Française de Linguistique Appliquée* 14 (2) (2009), 9-15. Disponible en: <https://goo.gl/QjL1mF> [Último acceso: 24/09/2018].
- Cabré, M. T. y Estopá, R., «Unidades de conocimiento especializado, caracterización y tipología», en: Cabré, M. T. y Bach, C. (eds.), *Coneixement, llenguatge i discurs especialitzat*. Barcelona: IULA, Documenta Universitaria, 2005, 69-94.
- Carrión, J. et al., *FEDER, la fuerza del movimiento asociativo*. Diego Marín Editores: Murcia 2015. Disponible en: <http://bit.ly/2GgTIHt> [Último acceso: 16/04/2019].
- Corpas Pastor, G., «Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada», *TRANS. Revista de Traductología* 5 (2001), 155-184. Disponible en: <https://goo.gl/2GjMD7> [Último acceso: 24/09/2018].
- Corpas Pastor, G. y Seghiri Domínguez, M., «Determinación del umbral de representatividad de un corpus mediante el algoritmo N-Cor», *Procesamiento del Len-*

- guaje Natural* 39 (2007), 165-172. Disponible: <http://bit.ly/2IByBT8> [Último acceso: 16/04/2019].
- EAGLES, *Preliminary Recommendations on Corpus Typology*, EAG-TCWG-CTYP/P. 1996. Disponible en: <https://bit.ly/2scX1sS> [Último acceso: 24/09/2018].
- EURORDIS, *Rare Diseases: understanding this Public Health Priority*. 2005. Disponible en: <https://goo.gl/zN9RDW> [Último acceso: 24/09/2018].
- Fantinuoli, C., «Revisiting corpus creation and analysis tools for translation tasks», *Cadernos de Tradução* 36, n.º especial (2016), 62-87. Disponible en: <https://goo.gl/SdsFHW> [Último acceso: 24/09/2018].
- FEDER-OBSER, *Estado de situación del movimiento asociativo de enfermedades raras en España*. 2017. Disponible en: <https://goo.gl/aHtU1h> [Último acceso: 24/09/2018].
- Koester, A., «Building small specialised corpora», en: O’Keeffe, A. y McCarthy, M. (eds.), *The Routledge Handbook of Corpus Linguistics*. Londres: Routledge, 2010, 66-79.
- Laviosa, S., *Corpus-based translations studies: theory findings applications*. Ámsterdam: Rodopi 2002.
- Llisterri, J., *Corpus Linguistics and Written Language Resources - Bibliography*. 2017. Disponible en: <https://goo.gl/soKfYz> [Último acceso: 24/09/2018].
- Mair, C., «Tracking ongoing grammatical change and recent diversification in present-day standard English: the complementary role of small and large corpora», *Language and Computers* 55 (2006), 355-376. Disponible en: <https://goo.gl/BLjoVP> [Último acceso: 24/09/2018].
- Marta Rodríguez, V., *Géneros textuales biomédicos y traducción: análisis de guías de práctica clínica en francés y en español en el ámbito de las enfermedades raras*. Tesis doctoral. Vigo: Universidade de Vigo 2015. Disponible en: <https://goo.gl/GJ5Bry> [Último acceso: 24/09/2018].
- Miquel Verges, J. y Sánchez Trigo, E., «The social model of translation and its application to internet search engines specialized in health: the ASEM search engine for neuromuscular diseases», *Meta* 55 (2) (2010), 374-386. Disponible en: <https://goo.gl/WyR4b9> [Último acceso: 24/09/2018].
- Nazar, R. y Cabré, M. T., «Un experimento de extracción de terminología utilizando algoritmos estadísticos supervisados», *Debate terminológico* 7 (2011), 36-55. Disponible en: <https://goo.gl/rLJBXu> [Último acceso: 24/09/2018].
- NGO Committee for Rare Diseases, *Our actions*. 2017. Disponible en: <https://goo.gl/pDo8NJ> [Último acceso: 24/09/2018].
- ORPHANET, *About Rare Diseases*. París: INSERM 2012. Disponible en: <https://goo.gl/rRTNP2> [Último acceso: 24/09/2018].
- Padró, L. y Stanilovsky, E., «FreeLing 3.0: Towards Wider Multilinguality», en: *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. ELRA: Estambul, 2012. Disponible en: <https://goo.gl/dBrSx8> [Último acceso: 24/09/2018].
- Sager, J., *A Practical Course in Terminology Processing*. Ámsterdam/Filadelfia: John Benjamins Publishing Company 1990.
- Sinclair, J., «Corpus and Text: Basic Principles», en: Wynne, M. (ed.), *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books, 2004. Disponible en: <https://goo.gl/e1d2Wo> [Último acceso: 24/09/2018].

- Sinclair, J., *Corpus, Concordance, Collocation*. Oxford: Oxford University Press 1991.
- Tognini-Bonelli, E., «Theoretical overview of the evolution of corpus linguistics», en: O’Keefe, A. y McCarthy, M. (eds.), *The Routledge Handbook of Corpus Linguistics*. Londres: Routledge, 2010, 14-27.
- Varela Vila, T., «Creation of a Trilingual Glossary of Neuromuscular Diseases», *International Journal of Translation* 21 (1-2) (2009), 107-118.
- Varela Vila, T., *Ontoloxías e tradución biomédica: creación dunha base de coñecemento terminolóxico sobre os erros innatos do metabolismo en francés e español*. Tesis doctoral. Vigo: Universidade de Vigo 2015. Disponible en: <https://goo.gl/YchVnk> [Último acceso: 24/09/2018].
- Varela Vila, T. et al., «Vocabulario francés-español de Enfermedades Raras: Errores Innatos del Metabolismo», *Panace@* 12 (33) (2011), 35-78. Disponible en: <https://goo.gl/BFXSFh> [Último acceso: 24/09/2018].
- Varela Vila, T. y Sánchez Trigo, E., «EMCOR: a corpus for terminological medical purposes», *JoSTrans: Journal of Specialised Translation (Special Issue on Terminology, Phraseology and Translation)* 18 (2012), 139-159. Disponible en: http://www.jostrans.org/issue18/art_varela.php [Último acceso: 24/09/2018].
- Varela Vila, T. y Sánchez Trigo, E., «Análisis terminológico de la categoría conceptual ‘enfermedad’ en artículos de investigación originales sobre enfermedades raras en francés y español», *Anales de Filología Francesa* 26 (2018), 255-275. Disponible en: <https://bit.ly/2XaLkX8> [Último acceso: 16/04/2019].
- Varela Vila, T. y Sánchez Trigo, E., «Diseño e implementación de una base de conocimiento terminológico sobre enfermedades raras», *Onomázein* 49 (2020), en prensa.
- Varela Vila, T., Sánchez Trigo, E. y Marta Rodríguez, V., «Vocabulario inglés-español de productos de apoyo utilizados en terapia ocupacional para personas con enfermedades neuromusculares», *Panace@* 13 (35) (2012), 69-90. Disponible en: <https://goo.gl/goPAzw> [Último acceso: 24/09/2018].
- Vargas Sierra, C., «Diseño de un corpus especializado con fines terminográficos: el Corpus de la Piedra Natural», *Debate Terminológico* 2 (7) (2006). París: RITERM (Red Iberoamericana de Terminología).
- Vargas Sierra, C., «La sistematización terminográfica: una propuesta metodológica para la elaboración de diccionarios traductológicos», en: *Actas del X Simposio Iberoamericano de Terminología*. Montevideo, Uruguay, 2008. Disponible en: <https://goo.gl/cmQPHJ> [Último acceso: 24/09/2018].
- Vargas Sierra, C., «Caracterización contrastiva (inglés-español) de la combinatoria léxica especializada: estudio basado en un corpus especializado», en: Mogorrón Huerta, P. y Mejri, S. (dirs.), *Langues Spécialisées, figement et traduction*. Alicante: Publicaciones de la Universidad de Alicante, 2012, 179-200. Disponible en: <http://goo.gl/6BtQ4b> [Último acceso: 24/09/2018].
- Vondříčka, P., «Aligning parallel texts with InterText», en: Calzolari, N. et al. (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*. Islandia: European Language Resources Association (ELRA), 2014, 1875-1879.
- Walsh, S., «Corpus Linguistics and Conversation Analysis at the Interface: Theoretical Perspectives, Practical Outcomes», en: Romero-Trillo, J. (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013. New Domains and Methodologies*. Dordrecht (Países Bajos): Springer, 2013, 37-51.

- Zanettin, F., «DIY Corpora: The WWW and the Translator», en: Maia, B., Haller, J. y Urlrych, M. (eds.), *Training the Language Services Provider for the New Millennium*. Porto: Faculdade de Letras, Universidade do Porto, 2002, 239-248.
- Zanettin, F., *Translation-driven corpora*. Manchester y Kinderhook (Nueva York): St. Jerome Publishing 2012.