

INVESTIGAR CON CORPUS EN EL ÁMBITO DE LA TRADUCCIÓN: CREACIÓN DE RECURSOS MULTILINGÜES SOBRE ENFERMEDADES RARAS

ELENA SÁNCHEZ TRIGO

Universidad de Vigo, España

Resumen. En este trabajo se presentan un conjunto de herramientas web de libre acceso y de naturaleza multilingüe centradas en el ámbito médico de las enfermedades raras que han sido creadas por nuestro grupo de investigación. Se describen las características del motor de búsqueda *MYOCOR*, de la base de conocimiento terminológico *ONTERMET* y, en especial, del reciente portal de recursos lingüísticos sobre enfermedades raras *RERCOR*. El objetivo de estos recursos es proporcionar a traductores, intérpretes y redactores, tanto profesionales como en formación, herramientas de consulta que permitan profundizar en el conocimiento de las características textuales y la terminología utilizada en el ámbito de la biomedicina. La novedad de nuestras aportaciones viene dada por: la metodología empleada (*corpus-based approach*), el área temática abordada (si bien prioritaria en las políticas de salud de la UE busca incrementar la atención médica y su visibilidad) y las lenguas seleccionadas (fundamentalmente el francés y español, en un ámbito en el que parece que solo existe el inglés). El interés de estos recursos reside, asimismo, en otros dos aspectos que consideramos relevantes. Por un lado, permite la vinculación entre nuestra investigación y la actividad traductora que realizamos. Por otro, conecta con las demandas sociales actuales, ya que contribuye a la difusión de la información sobre enfermedades raras, aspecto demandado por las asociaciones de pacientes, proporcionando herramientas que favorezcan la calidad de las traducciones.

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

INTRODUCCIÓN

La línea de investigación que presentamos se centra de manera específica en las enfermedades raras (ER), también denominadas enfermedades ‘minoritarias’, ‘huérfanas’ o ‘poco frecuentes’, por tratarse de un subdominio de gran actualidad y en el que traducimos habitualmente textos al español dirigidos, principalmente, a pacientes y su entorno más próximo. Las ER son un numeroso y diverso grupo

de enfermedades, entre 6000 y 8000, que se caracterizan por su frecuencia reducida. Como consecuencia han sido objeto de poca atención tanto por parte de la investigación médica como de la industria farmacéutica. Sin embargo, estas enfermedades consideradas de manera conjunta afectan, por ejemplo, en la UE a unos 24-36 millones de personas. No son, por tanto, tan infrecuentes. En estos momentos, los afectados, a través de las asociaciones de pacientes, reclaman la provisión de una asistencia sanitaria y social adecuada que convierta a las ER en una prioridad en las políticas de salud pública a nivel mundial.

Nuestra finalidad es la creación de herramientas que faciliten la mediación comunicativa en el ámbito de la biomedicina. En las páginas que siguen presentamos *MYOCOR* (En línea 1), *ONTERMET* (En línea 2) y, en especial, *RECOR* (En línea 3), la más reciente. Si bien estos recursos se centran en el ámbito de las ER, son de naturaleza multilingüe y de acceso libre online, presentan también características diferenciadas en función de los objetivos establecidos en los proyectos en el marco de los que se crearon. Este trabajo se inscribe en los proyectos ED431C2018/50 Xunta de Galicia y TIN2017-85160-C2-2-R del Ministerio de Economía y Competitividad (España).

1 METODOLOGÍA

Tomamos como punto de partida el trabajo con corpus por permitir esta aproximación un análisis empírico y cuantitativo de los textos. Metodológicamente la creación de los recursos se ha realizado en varias fases: compilación y tratamiento de los diferentes corpus multilingües, configuración de las aplicaciones y diseño de las interfaces de consulta. Se han seguido los criterios de los principales expertos en corpus (*EAGLES*, 1996; Laviosa, 2002; McEnery, 2003; Sinclair, 2004; o Tognini-Bonelli, 2010), con especial atención a cuestiones de calidad y representatividad en corpus especializados (Bowker y Pearson, 2002; Sardinha, 2002; Walsh, 2013). Tipológicamente, además de especializados en ER, los corpus compilados son multilingües (español-francés y/o inglés), multigenéricos (excepto uno de *RECOR*: *ERCOR*) (En línea 3) y comparables y/o paralelos. Los textos, inicialmente en formato gráfico, se pasaron a formato plano (Unicode/UTF-8) para poder ser tratados. Se etiquetaron (estructura y/o morfosintaxis) y, si las utilidades buscadas lo requerían, se lematizaron. Se realizó una alineación oracional semiautomática de los corpus paralelos. Junto a estos elementos comunes, cada recurso tiene, a su vez, aspectos y aplicaciones propios por ser resultado de proyectos distintos.

2 ANÁLISIS DE LAS HERRAMIENTAS

Se sintetizan los elementos específicos de cada recurso: finalidad, características de los corpus en los que se basan (tipo, géneros textuales, tamaño, período temporal, etc.) y análisis de las funcionalidades de las interfaces. Las referencias a las que se remite permiten ampliar la información que se proporciona.

2.1 MYOCOR: MOTOR DE BÚSQUEDA SOBRE ENFERMEDADES NEUROMUSCULARES

Recurso documental accesible online creado en el marco de un proyecto de investigación realizado para ASEM Galicia (Asoc. contra las Enfermedades Neuromusculares) entre 2004 y 2007. Su finalidad fue facilitar el acceso a una información de calidad sobre enfermedades neuromusculares (ENM), el mayor grupo de ER. Cuando se construyó la información sobre ENM era difícil de localizar por su dispersión y la inexistencia de fuentes en español para pacientes. La única excepción eran las traducciones de textos de la AFM (Association française contre les myopathies) promovidas por ASEM.

Primero se compiló un corpus bilingüe paralelo (francés-español) multi-genérico, con textos especializados y de divulgación (artículos, monografías, manuales, conferencias o folletos), producidos de 1995 a 2005. Después, tras el etiquetado estructural, se elaboró una herramienta de búsqueda documental para la extracción de información del mismo (Miquel Verges y Sánchez Trigo, 2010).

Buscador de MYOCOR de ASEM versión 1.0 ■ ■
(Buscador sobre enfermedades neuromusculares)

Palabras clave:

Para búsquedas textuales, insertar entre comillas. Ej.: "miopatía miotubular". [MÁS...](#)

Idioma:

Opciones:

- Introducción
- Enfermedad
- Descripción enfermedad
- Causas
- Síntomas
- Evolución
- Transmisión
- Diagnóstico
- Tratamiento
- Investigación
- Vida diaria
- Estudio
- Contacto

Enfermedad:

Idioma:

- Español
- Francés
- Español->Francés
- Francés->Español

Copyright © 2005, ASEM Galicia [HOME](#)

Figura 1 MYOCOR: interfaz consulta

La interfaz de consulta de la aplicación permite (Figura 1): a) realizar búsquedas a partir de una o varias palabras clave (caja de búsqueda superior); b) buscar información sobre una ENM concreta (desplegable sección 'Enfermedad'). En la primera de estas posibilidades, se puede, además, seleccionar aspectos

concretos o partes determinadas de los textos de acuerdo con criterios temático-estructurales ('Opciones'). Así se obtiene información solo sobre algunos aspectos específicos.

Los resultados ofrecen enlaces a los textos en los que aparece la información buscada. Están ordenados por número de apariciones de la búsqueda realizada. En caso de que se haya optado por buscar una parte concreta del documento se muestran una serie de enlaces a los respectivos fragmentos textuales.

2.2 ONTERMET: BASE DE CONOCIMIENTO TERMINOLÓGICO SOBRE ERRORES INNATOS DEL METABOLISMO

Se trata de un recurso terminográfico para acceder de forma conceptualmente estructurada a la terminología del ámbito de los errores innatos del metabolismo (EIM). Incluye: denominación de las enfermedades, síntomas y signos (Varela Vila y Sánchez Trigo, aceptado, en prensa). Su finalidad es describir el subdominio y servir como fuente de información para traducir. Se han seleccionado los EIM por ser un ejemplo representativo del conjunto de ER, por su número (más de 500) y heterogeneidad clínica.

The screenshot shows the ONTERMET website interface. At the top, there is a search bar with the text "Buscar" and navigation links for "Acerca de ONTERMET", "Ayuda", and "Español". Below the search bar, the text "Base de conocimiento terminológico" is displayed. On the left side, there is a list of categories with expandable sub-items:

- Enfermedades
 - error innato del metabolismo C. EMCOR
- Síntomas o signos
 - ciertas enfermedades infecciosas y parasitarias
 - enfermedades de la sangre y de los órganos hematopoyéticos, y ciertos trastornos que afectan el mecanismo de la inmunidad
 - enfermedades endocrinas, nutricionales y metabólicas
 - trastornos mentales y del comportamiento
 - enfermedades del sistema nervioso
 - enfermedades del ojo y sus anexos
 - enfermedades del oído y de la apófisis mastoides

On the right side, there is a section titled "Acerca de ONTERMET" with the following text:

ONTERMET es una base de conocimiento terminológico bilingüe sobre el campo biomédico de los errores innatos del metabolismo (EIM), un numeroso y complejo grupo de enfermedades raras (ER), caracterizado por su alta heterogeneidad clínica. Este recurso, que combina un diseño conceptual y terminológico, permite acceder a la terminología del subdominio de los EIM, en español y en francés.

ONTERMET forma parte del trabajo llevado a cabo por Tamara Varela Vila para su tesis doctoral, titulada «Ontoloxías e tradución biomédica: creación dunha base de coñecemento terminolóxico sobre os erros innatos do metabolismo en francés e español» (Ontologías y traducción biomédica: creación de una base de conocimiento terminológico sobre los errores innatos del metabolismo en francés y español), bajo la dirección de la Dra. Elena Sánchez Trigo (Universidade de Vigo). Desde el punto de vista metodológico, la creación de la base de conocimiento ha incluido tres fases previas:

Figura 2 ONTERMET: interfaz consulta

Para la creación de esta base de conocimiento de acceso libre online, en primer lugar, se compiló, anotó y etiquetó (morfosintaxis) EMCOR, un corpus bilingüe (español y francés), comparable y multigenérico (90% especializado: caso clínico, artículos, resumen y 10% semiespecializado/divulgación: información pacientes), con textos publicados entre 2001 y 2013. A continuación, se etiquetó morfosintácticamente con *TreeTagger*, se hizo la organización conceptual del ámbito y la extracción terminológica. Posteriormente, se construyó la ontología que almacena de manera jerárquica los EIM y sus síntomas o signos. Asimismo, en ella

se formalizan las relaciones existentes entre los conceptos: qué síntomas o signos afectan a cada enfermedad y qué enfermedades pueden presentar determinado síntoma o signo. Constituye el núcleo que estructura la base de conocimiento.

La interfaz permite consultas en francés y en español. Pueden realizarse a través de la organización conceptual (Figura 2, columna izda.) o de la caja de búsqueda superior con palabras clave. Por ejemplo, si se busca un síntoma/signo como ‘dificultad auditiva’ la información obtenida ofrece: datos sobre el término preferido (‘hipoacusia’), código y enlace con la CIE-10 (Clasificación Internacional de Enfermedades), denominaciones alternativas (‘dificultad para la audición’, ‘pérdida auditiva’, ‘sordera’, etc.), enfermedades asociadas y síntomas/signos superordinados o subordinados. Si se cambia el idioma al francés, ONTERMET (En línea 2) ofrece el término preferido equivalente, *surdité*, su código en la CIM-10 (Classification internationale des maladies) e hiperenlace, así como el resto de datos indicados en español ahora en francés. La aplicación ofrece, también, notas específicas sobre los términos utilizados para denominar el concepto. Por ejemplo, en relación con el ‘trastorno del ciclo de la urea’ se indica: ‘Debe evitarse el uso de “desorden” con el sentido de “trastorno”, puesto que constituye un calco del inglés *disorder*’.

Se recogen 2124 conceptos que representan enfermedades, síntomas o signos y 1188 términos preferidos en español y francés. Se incluyen, también, 1357 denominaciones alternativas en español frente a 901 en francés. Esta diferencia indica un mayor grado de variación terminológica en español. ONTERMET (En línea 2) ofrece, igualmente, la posibilidad de acceder a la estructura conceptual del ámbito de los EIM en forma de mapa conceptual. Así, la información se presenta de un modo gráfico.

2.3 RERCOR: PORTAL DE RECURSOS LINGÜÍSTICOS SOBRE ENFERMEDADES RARAS

Creado recientemente reúne en una misma plataforma diversos recursos multilingües de acceso libre. Se organizan en torno a una aplicación de consulta de corpus con un analizador de concordancias y a otra de consulta de glosarios (Sánchez Trigo y Varela Vila, 2019).

La primera incluye tres corpus con más de dos millones de palabras (MYOCOR 2.0, EMCOR y ERCOR). Se trata de corpus bilingües (francés y español), multigenéricos los dos primeros (7 géneros con 9 subgéneros) y monogenérico el tercero (guías de práctica clínica). Son comparables y paralelos y presentan etiquetación (Freeling) y lematización. Permite búsquedas seleccionando un conjunto de parámetros disponibles en el formulario de consulta. Se incluyen cinco secciones principales: *Búsqueda*, *Resultado*, *Sensibilidad*, *Filtros* y *Texto* (enmarcadas en rojo Figura 3). Dentro de estas, hay diferentes campos con menús desplegables: *corpus*, *tipo*, *género*, *nivel de especialización*, *año*, *idioma*, entre otros (subrayadas en azul Figura 3).



Figura 3 RERCOR: interfaz aplicación consulta de corpus

De este modo el usuario puede seleccionar diversas opciones, por ejemplo: a) consultar los tres corpus que integran la aplicación o solo alguno(s); b) buscar palabras ortográficas o elementos gramaticales, ya que no siempre coinciden. Así, palabras ortográficas como *s'ossifier* (fr) o 'acentuarse'(es) están compuestas por dos elementos gramaticales (pronombre personal y verbo), cada uno con una etiqueta morfosintáctica y lema; c) atender o no a la presencia de mayúsculas y /o acentos, lo que permite comprobar, por ejemplo, el número de ocurrencias de 'síndrome de GRACILE' frente a 'síndrome de gracile' o de 'ortesis' frente a 'órtesis' y tener datos sobre las formas de uso mayoritario en los corpus; d) obtener todas las muestras textuales o centrarse en alguno de los géneros y/o subgéneros, de los grados tres de especialización, de los años del total de 15 a los que pertenecen los textos y/o de los dos idiomas.

Ce signe a été	classiquement décrit dans diverses atteintes de les motoneurons de la corne antérieure de la moelle à le niveau thoracique , [...]
L'ostéoporose est une complication	classiquement décrite dans l'homocystinurie classique liée à le déficit en cystathionine β-synthase (OMIM : 236200) .
Cette délétion ,	déjà décrite chez un enfant présentant un tableau similaire , entraîne l'excision de l'exon 6 de le gène [...]
De les polymorphismes ,	déjà décrits à le niveau de l'exon 1 (p.ASA , p.A20A et p.H33Q) , de l'exon [...]
IA de type II ont été rajoutés à les quatre	déjà décrits dans la revue Cochrane de 2009 .
ine de ces thérapeutiques et nous allons	également décrite de nouvelles perspectives thérapeutiques en_cours de développement dans le domaine de l'enzymothérapie subs
	Initialement décrite par Kaufman , l'atteinte ovarienne de la galactosémie se traduit sur le plan clinique par un hypogonadisme [...]
Les crises sont polymorphes , assez	mal décrites dans la littérature , avec soit de les chutes ou de les absences atoniques , soit de les [...]
OCT , le_plus fréquent , est l'anomalie la	mieux décrite à ce jour
Les mêmes mutations sont ainsi	parfois décrites chez de les malades publiés comme de les variants dits cardiaques et chez de les malades publiés comme [...]
Il s'agit d'une complication rare ,	peu décrite chez l'enfant et de physiopathologie encore mal connue .
Les cas sont plus rares et	peu décrits à ce jour (dans notre série , 5 enfants étaient âgés de_plus de 6 ans dont 2 [...]
Les 2 interventions les	plus décrites chez les patients atteints de DMM sont l'intervention pour la cataracte et la cholécystectomie .
• Rare (1 cas sur 1 million d'individus)	principalement décrite à le Moyen-Orient et en Asie (Japon , Chine)

Figura 4 RERCOR: aplicación corpus. Concordancias patrón ADV + *décrire* + PREP

Estos parámetros, de los que solo hemos destacado parte de los posibles, son de utilidad para precisar las búsquedas y obtener resultados concretos. Permiten, por ejemplo, buscar: *dystrophie* como palabra ortográfica, en el género artículo científico, subgénero caso clínico, desde 2005 a 2008, teniendo en cuenta mayúsculas y acentos. Es posible también, consultar simultáneamente más de

un elemento gramatical para analizar el comportamiento de patrones sintácticos (Figura 4). Se trata de un aspecto fundamental para la recuperación de la información fraseológica, tan necesaria para traducir.

Los resultados por defecto son líneas de concordancia con la ocurrencia destacada en negrita. Se incluyen datos del corpus del que procede, lengua y fecha. Estos pueden ampliarse con información de naturaleza morfosintáctica, colocacional y contextual. La aplicación proporciona, además, datos numéricos de frecuencias totales (*Frecuencia simple*) y/o por género, nivel de especialización y lengua (*Frecuencia completa*).

La segunda aplicación del portal (glosarios), permite la consulta, simultánea o individualizada, de tres recursos terminológicos multilingües: *Vocabulario francés-español de errores innatos del metabolismo*, *Vocabulario inglés-español de productos de apoyo* y *Glosario francés-español-inglés de enfermedades neuromusculares*. Junto a los términos equivalentes en otra lengua, proporcionan información de naturaleza descriptiva (categoría gramatical, sinónimos o, notas de uso, entre otras). Las búsquedas se pueden realizar directamente en el listado alfabético o introduciendo un término en la caja de búsqueda.

Todos <input type="text" value="myopathie"/> <input type="button" value="Buscar"/> <input type="button" value="Limpiar"/>		
<input type="button" value="A"/> <input type="button" value="B"/> <input type="button" value="C"/> <input type="button" value="D"/> <input type="button" value="E"/> <input type="button" value="F"/> <input type="button" value="G"/> <input type="button" value="H"/> <input type="button" value="I"/> <input type="button" value="J"/> <input type="button" value="K"/> <input type="button" value="L"/> <input type="button" value="M"/> <input type="button" value="N"/> <input type="button" value="O"/> <input type="button" value="P"/> <input type="button" value="Q"/> <input type="button" value="R"/> <input type="button" value="S"/> <input type="button" value="T"/> <input type="button" value="U"/> <input type="button" value="V"/> <input type="button" value="W"/> <input type="button" value="X"/> <input type="button" value="Y"/> <input type="button" value="Z"/>		
<input type="button" value="Anterior"/> <input checked="" type="button" value="1"/> <input type="button" value="2"/> <input type="button" value="3"/> <input type="button" value="4"/> <input type="button" value="Siguiente"/>		
cardiomyopathie → atteinte du myocarde → myocardiopathie	Es Glosario	miocardiopatía Vocabulario francés-español de errores innatos del metabolismo Otras denominaciones en español • cardiomiopatía
cardiomyopathie dilatée → cardiomyopathie dilatative	Es Glosario	miocardiopatía dilatada Vocabulario francés-español de errores innatos del metabolismo Otras denominaciones en español • cardiomiopatía dilatada
cardiomyopathie familiale	Es Glosario	miocardiopatía familiar Vocabulario francés-español de errores innatos del metabolismo

Figura 5 RERCOR: aplicación glosarios. Resultados búsqueda *myopathie*

La aplicación cuenta con cerca dos mil entradas. Aporta información terminológica y permite completar la información ofrecida por el análisis de concordancias en la consulta de corpus.

3 RESULTADOS Y DISCUSIÓN

Los recursos creados ponen de manifiesto que los resultados obtenidos con la metodología basada en corpus son versátiles y muy productivos. Con independencia de sus diferentes finalidades, estos recursos muestran también la evolución de nuestra línea de investigación: a) los corpus en los que se basan se han ido ampliado (*RERCOR* incluye más de 2 millones de palabras) (En línea 3), lo que aumenta la representatividad y enriquece los resultados que se ofrecen;

b) las interfaces tienen más funcionalidades, lo que permite parametrizar las búsquedas en función de los intereses concretos de los usuarios; c) se han incluido nuevas opciones con especial atención al contexto (concordancias), lo que es de especial utilidad para traducir; e) se reúnen diferentes recursos y aplicaciones en un mismo portal facilitar la consulta.

Todo esto ha permitido que las herramientas que ofrezcan resultados más precisos. En la actualidad, seguimos trabajando en una nueva versión del portal RERCOR (En línea 3) con la finalidad de ampliar y mejorar sus aplicaciones.

CONCLUSIONES

Las diferentes herramientas multilingües de libre acceso para traductores, intérpretes o redactores científicos presentadas, son innovadoras por: a) centrarse en un ámbito biomédico poco estudiado (las ER) y multidisciplinar (confluyen numerosas áreas), lo que aporta novedad y riqueza a los resultados; b) ofrecer resultados basados en la realidad textual (derivados de trabajar con corpus representativos); c) incluir, junto al inglés, francés y español, lenguas con poco recursos especializados en biomedicina; d) contribuir a la difusión de la información sobre ER lo que la vincula con una demanda social de actualidad.

REFERENCIAS BIBLIOGRÁFICAS

- Bowker, L. y Pearson, J. (2002) *Working with specialized language: a practical guide to using corpora*. Londres: Routledge.
- EAGLES (1996) *Preliminary Recommendations on Corpus Typology*, EAG-TCWG-CTYP/P. Disponible en <http://www.ilc.cnr.it/EAGLES/corpusyp/corpusyp.html> [Consultado el 16 de mayo de 2019].
- Koester, A. (2010) Building small specialised corpora. In A. O’Keeffe, and M. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics* (pp. 66-79). Londres: Routledge.
- Laviosa, S. (2002) *Corpus-based translation studies: theory findings applications*. Amsterdam: Rodopi.
- McEnery, T. (2003) Corpus linguistics. In R. Mitkov (ed.) *Oxford Handbook of Computational Linguistics* (pp. 448-463). Oxford: Oxford University Press.
- Miquel Verges, J. y Sánchez Trigo, E. (2010) 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): 374-386. Disponible en <https://bit.ly/2I8i5te> [Consultado el 16 de mayo de 2019].
- Sánchez Trigo, E. y Varela Vila, T. (2019) RERCOR: portal de recursos lingüísticos multilingües sobre enfermedades raras. *Estudios de Traducción*, 9: 131-150. Disponible en <https://revistas.ucm.es/index.php/ESTR/article/view/65706> [Consultado el 16 de mayo de 2019].
- Sardinha, T. (2002) Tamanho de corpus. *The ESPecialist*, 23 (2): 103-122. Disponible en <https://revistas.pucsp.br/esp/article/view/9381> [Consultado el 16 de mayo de 2019].
- Sinclair, J. (2004) Corpus and Text: Basic Principles. In M. Wynne, (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books. Disponible en http://icar.cnrs.fr/ecole_thematique/contact/documents/Baude/wynne.pdf [Consultado el 16 de mayo de 2019].
- Tognini-Bonelli, E. (2010) Theoretical overview of the evolution of corpus linguistics. In O’Keeffe, A. y McCarthy, M. (eds.) *The Routledge Handbook of Corpus Linguistics* (pp. 14-27). Londres: Routledge.

- Varela Vila, T. y Sánchez Trigo, E. (aceptado, en prensa) Diseño e implementación de una base de conocimiento terminológico sobre enfermedades raras. *Onomázein*, 49.
- Walsh, S. (2013) Corpus Linguistics and Conversation Analysis at the Interface: Theoretical Perspectives, Practical Outcomes. In J. Romero-Trillo (ed.) *Yearbook of Corpus Linguistics and Pragmatics 2013. New Domains and Methodologies* (pp. 37-51). Dordrecht (Países Bajos): Springer.

RECURSOS DIGITALES

- 1) [En línea 1] MYOCOR. Disponible en <http://www.asemgalicia.com/servicios/> [Consultado el 16 de mayo de 2019].
- 2) [En línea 2] ONTERMET. Disponible en <http://ontermet.com/es/> [Consultado el 16 de mayo de 2019].
- 3) [En línea 3] RERCOR. Disponible en <https://www.rercor.org/> [Consultado el 16 de mayo de 2019].

CORPUS-BASED RESEARCH IN TRANSLATION: CREATING OF MULTILINGUAL RESOURCES ON RARE DISEASES

Abstract. This paper presents a set of multilingual, open-access Web tools created by our research group which focus on the medical field of rare diseases. The characteristics of three tools are described: the search engine *MYOCOR*, the terminological knowledge base *ONTERMET*, and most notably, the recently developed portal of language resources on rare diseases, *RERCOR*. These resources aim to provide query tools for translators, interpreters and editors (trainees and consolidated professionals alike), enabling them to further their knowledge of the textual characteristics and terminology applied in the field of biomedicine. The innovative approach of our contributions derives from three elements: our chosen methodology (corpus-based approach); the subject area, which is a priority for EU health policy and a domain seeking healthcare improvements and increased visibility; and the choice of mainly French and Spanish in a field where English would seem to be the only language. These resources can be considered significant for two further reasons. First, a link has been created between our research and our translation activities. Secondly, because they respond to current social demands, patient associations call for the dissemination of information on rare diseases and our resources play a part in this by providing tools to improve the quality of translations.

Key words: multilingual resources, specialised corpora, rare diseases, medical terminology, medical translation

Elena Sánchez Trigo (Catedrática de traducción e interpretación) trabaja en la Universidad de Vigo (España). Entre sus intereses académicos están la traducción de textos médicos, la terminología y las aplicaciones de los corpus lingüísticos a la traducción. Correo electrónico: etrigo@uvigo.es