



## Traducción de textos médicos: desarrollo de un portal de recursos lingüísticos sobre enfermedades raras

Translation of medical texts: Development of a language resource portal on rare diseases

**Elena Sánchez Trigo**

Universidade de Vigo  
Vigo, España  
etrigo@uvigo.gal

<https://orcid.org/0000-0002-1834-1722> 

**Tamara Varela Vila**

Universidade de Vigo  
Vigo, España  
tvarela@uvigo.gal

<https://orcid.org/0000-0002-8447-8102> 

**Resumen:** En este trabajo se presentan las características y potencialidades de uso de la última versión del portal de recursos lingüísticos multilingües sobre enfermedades raras que estamos desarrollando: RERCOR (3.1.2). Las enfermedades raras son un amplio grupo de dolencias que afectan a unos 30 millones de personas en la UE y a unos 47 millones en Iberoamérica. En estos momentos buscan mayor visibilidad y prioridad en las políticas de salud pública. El portal tiene como objetivo proporcionar herramientas multilingües que permitan conocer mejor los usos lingüísticos propios de este subdominio médico, especialmente su terminología y colocaciones. El gran número de enfermedades raras (más de 6000) y sus diversas consecuencias hacen que en este ámbito confluyan numerosas especialidades médicas y muchas otras áreas relacionadas con el ámbito de la salud en su sentido más amplio. Esta multidisciplinariedad confiere a RERCOR utilidad para la traducción, revisión y redacción de textos de todo el ámbito biomédico, así como para la investigación y la docencia. RERCOR está constituido por tres aplicaciones disponibles a través de una interfaz web de libre acceso: aplicación de consulta de corpus, aplicación de consulta de vocabularios y asistente para la redacción de textos. La primera facilita la búsqueda de concordancias e información morfosintáctica, colocacional y contextual en seis corpus —tres bilingües (francés y español) y dos trilingües (inglés, francés y español)—, comparables y paralelos, lematizados y etiquetados. La segunda permite consultar un glosario trilingüe y tres vocabularios bilingües que incluyen equivalencias en otras lenguas e información adicional como definición, sinónimos u otras denominaciones. La tercera aplicación consiste en un prototipo para facilitar la redacción de textos

sobre este subdominio médico. El diseño del portal permite una consulta completa, dinámica e intuitiva sin requerir conocimientos informáticos específicos.

**Palabras-clave:** terminología médica; traducción médica; enfermedades raras; corpus especializados; recursos multilingües.

**Abstract:** This paper presents the features and potential of use of the latest version of the portal of multilingual linguistic resources on rare diseases that we are developing: RERCOR (3.1.2). Rare diseases are a large group of diseases that affect about 30 million people in the EU and about 47 million in Latin America. They are currently seeking greater visibility and priority in public health policies. The portal aims to provide multilingual tools to better understand the linguistic uses of this medical subdomain, especially its terminology and collocations. The large number of rare diseases (more than 6000) and their diverse consequences mean that this field brings together numerous medical specialties and many other areas related to the field of health in its broadest sense. This multidisciplinary nature makes RERCOR useful for the translation, revision and editing of texts from the entire biomedical field, as well as for research and teaching purposes. RERCOR consists of three applications available through a freely accessible web interface: a corpus query application, a vocabularies query application and an assistant for writing texts. The first facilitates the search for concordances and morphosyntactic, collocational and contextual information in six corpora — three bilingual (French and Spanish) and two trilingual (English, French and Spanish) —, comparable and parallel, lemmatized and labeled. The second allows consultation of a trilingual glossary and three bilingual vocabularies that include equivalents in other languages and additional information such as definition, synonyms or other names. The third application consists of a prototype to facilitate the writing of texts on this medical subdomain. The design of the portal allows a complete, dynamic and intuitive consultation without requiring specific computer knowledge.

**Keywords:** medical terminology; medical translation; rare diseases; specialised corpora; multilingual resources.

## I. Introducción

La traducción de textos médicos es una de las ramas más antiguas de la actividad traductora (Montalt *et al.*, 2018; van Hoof, 2001; Fisbach, 1998) y está en permanente crecimiento. La rápida evolución del conocimiento médico derivado de la investigación (técnicas, especialidades, procesos, terapias...), así como la internacionalización que caracteriza a este sector (organizaciones de la salud, comercialización de medicamentos, contactos entre investigadores, movilidad de pacientes...) constituyen factores determinantes para su desarrollo actual.

Si bien se trata de un ámbito muy diverso, sus principales características desde el punto de vista de la traducción se pueden sintetizar en los siguientes puntos (Montalt Ressurrecció & González Davies, 2007; Muñoz Miquel, 2009, 2014):

- amplitud del dominio de conocimiento: están implicadas muchas especialidades interrelacionadas, pero también con claras diferencias entre ellas. Una consulta de las diferentes especialidades médicas reconocidas por el Ministerio de Sanidad español es



ilustrativa de esta situación;

- complejidad y variedad de los conocimientos temáticos: en correspondencia con las características de un ámbito que, además, está en constante evolución;
- complejidad terminológica: como consecuencia de la amplitud y variedad de áreas y de su rápida evolución;
- gran variedad de situaciones comunicativas: junto a comunicación entre médicos de igual o diferente especialidad, existen otras muchas situaciones derivadas de la comunicación también entre estos y otros profesionales de la salud (personal de enfermería, fisioterapeutas, nutricionistas...), entre médicos y pacientes o, por ejemplo, entre profesionales sanitarios y el público en general para difundir determinadas medidas. Estos son solamente algunos de los múltiples ejemplos;
- gran variedad de géneros textuales: como consecuencia de las situaciones comunicativas mencionadas existen textos de naturaleza muy diversa.

Teniendo en cuenta estas cuestiones, así como las competencias profesionales de los traductores descritas en la Norma UNE-EN ISO 17100 (AENOR, 2015) —ya que permiten identificar las áreas de interés para conformar el perfil de esta actividad— y nuestra experiencia como traductoras de textos biomédicos, decidimos desarrollar, desde el año 2018, el portal *RERCOR: Recursos sobre enfermedades raras* (Sánchez Trigo & Varela Vila, 2018-2024), al que en adelante nos referiremos como RERCOR. Se trata de un proyecto novedoso de acceso libre a través de una interfaz web con recursos lingüísticos en francés, inglés y español sobre las enfermedades raras (ER), que hemos ido enriqueciendo de manera paulatina desde sus inicios. Su objetivo es contribuir a mejorar el conocimiento sobre los usos lingüísticos propios de este subdominio médico, especialmente en relación con su terminología y colocaciones<sup>1</sup>. Ahora bien, la confluencia en el ámbito de las ER de muy diversas especialidades médicas y de diferentes áreas de la salud confiere a los recursos elaborados utilidad para la traducción, revisión y redacción de textos del conjunto del ámbito biomédico tanto para la actividad profesional, como para la docente e investigadora.

En este trabajo se presentan las características y potencialidades de uso de la última versión del portal RERCOR, en estos momentos la versión 3.1.2, que, frente a las anteriores, incluye más recursos y nuevas funcionalidades. En primer lugar, se expone cómo surge este proyecto, el interés del ámbito temático seleccionado y se resumen las bases teóricas, metodológicas y las etapas seguidas para la creación del portal. A continuación, se presentan las aplicaciones que lo integran. En tercer lugar, se describen las características de la interfaz web de libre acceso que se ha diseñado y se muestran algunas de las posibilidades de explotación con ejemplos de diferentes consultas que se pueden realizar. El portal permite una consulta fácil sin que sean necesarios conocimientos informáticos específicos.

---

<sup>1</sup> En este sentido, entendemos por ‘terminología’ el conjunto de unidades terminológicas, tanto monoléxicas (p. ej., ‘disnea’) como poliléxicas (p. ej., ‘enfermedad inflamatoria intestinal’) empleadas para hacer referencia a un concepto, mientras que utilizamos el término ‘colocación’, de una forma amplia, para referirnos a combinaciones de palabras o términos que en este ámbito de especialidad se usan con mayor frecuencia que otras combinaciones también posibles (p. ej., ‘interrumpir + tratamiento’, ‘referir + síntoma’ o ‘estrechamente + monitorizado’).

## 2. Origen, bases del proyecto y fases seguidas para la creación del portal

Debemos empezar destacando que RERCOR es un proyecto que surge directamente ligado a la traducción. Desde 1997 colaboramos con la Federación española contra las Enfermedades Neuromusculares (ASEM) traduciendo textos sobre ER de la reconocida Association française contre les myopathies (AFM-Téléthon). Esta potente asociación francesa publica documentos de gran rigor científico y permanentemente actualizados sobre estas enfermedades porque consideran que disponer de una información de calidad permite: “mejorar el conocimiento sobre las enfermedades raras, que durante mucho tiempo han sido ignoradas y aún siguen siendo poco conocidas, así como compartir y explicar los avances en la investigación y los progresos terapéuticos<sup>2</sup>” (AFM-Téléthon, 2002, traducción nuestra). Por su parte, ASEM consideró importante que estos textos estuviesen disponibles también español para continuar, así, la interesante línea iniciada por la AFM y ampliar el número de personas que podrían acceder a esta relevante documentación.

Las ER son un amplio y heterogéneo grupo de más 6000 dolencias que se caracterizan por “una gran diversidad de síntomas y signos que varían no solo de una enfermedad a otra, sino también de un paciente a otro con la misma enfermedad<sup>3</sup>” (EURORDIS, traducción nuestra). En la UE se consideran ER aquellas que afectan a menos de 1 de cada 2000 personas (Alliance Maladies Rares). Se trata, por lo tanto, de enfermedades que de manera individualizada se caracterizan por su baja prevalencia, es decir, de enfermedades minoritarias o poco frecuentes —otras de las denominaciones con las que se las conoce—. Sin embargo, consideradas en conjunto, las ER afectan a un porcentaje de la población estimado de entre el 3,5 % y el 5,9 % (Nguengang Wakap *et al.*, 2020). Es decir, se estima que habría unos 300 millones de personas con ER en total en el mundo: 30 millones en la UE, 25 millones en EE. UU. y 47 millones en Iberoamérica (SIO ALIBER, 2020). Esta nueva perspectiva hace que ya no pueda considerarse que estas enfermedades afectan a un grupo minoritario de personas. Se trata de un colectivo relevante al que debe prestarse atención.

Nuestro trabajo de traducción sobre textos de ER, además de implicarnos en un proyecto con marcado valor social, nos permitió conocer este amplio ámbito que presenta un gran interés por diversos motivos. Entre otros, reúne dos aspectos que consideramos fundamentales. Por un lado, su gran actualidad desde el punto de vista sociosanitario. Se trata de una característica que es el resultado de las constantes iniciativas para conseguir una mayor visibilidad y prioridad en las políticas de salud pública. Esto ha permitido que estas enfermedades sean cada vez más conocidas por toda la sociedad. Por otro lado, su naturaleza multidisciplinar, que es el resultado de la confluencia en el ámbito de las ER de numerosas especialidades médicas (neurología, rehabilitación, traumatología, cirugía, cardiología, neumología, anestesia, oftalmología...) junto a muchas otras relacionadas con el ámbito de la salud en su sentido más amplio (psicología, higiene, dietética, accesibilidad, calidad de vida de los enfermos...). Esta característica, derivada del amplio número de enfermedades que se incluyen en las ER, confiere a esta área, y en consecuencia también al portal RERCOR, representatividad del conjunto del ámbito médico.

---

<sup>2</sup> “*améliorer la connaissance sur les maladies rares, longtemps ignorées, encore méconnues, partager et expliquer les avancées de la recherche et les progrès thérapeutiques*” (AFM-Téléthon).

<sup>3</sup> “*a wide diversity of symptoms and signs that vary not only from disease to disease but also from patient to patient suffering from the same disease*” (EURORDIS).

Nuestra actividad traductora nos permitió constatar, asimismo, la necesidad de desarrollar recursos actualizados de calidad y de libre acceso que pudiesen ser utilizados en la práctica profesional de la traducción y actividades afines — así como también para la docencia (López-Rodríguez, 2016) o la investigación— en un ámbito complejo y en permanente actualización como este en el que nos situamos.

En relación con las bases teóricas, en la concepción del portal abordamos el estudio de la terminología desde un punto de vista integrador, no prescriptivo, que atiende al uso real de los términos y al contexto en que se utilizan. En consonancia con estos planteamientos, nuestro marco teórico toma como referencia destacada las propuestas de la teoría comunicativa de la terminología (Cabré, 1999, 2003, 2009) o las realizadas sobre términos y contexto (Pearson, 1988). Estos enfoques teóricos ponen de manifiesto que nos situamos en el marco de la función comunicativa de la terminología y han determinado también nuestra metodología. Así, desde el punto de vista metodológico, el portal se inscribe en el productivo campo de los CBTS (*Corpus Based Translation Studies*) porque permite el necesario análisis empírico y cuantitativo de los textos que nuestro enfoque investigador conlleva<sup>4</sup>. En estos momentos, los corpus constituyen una herramienta fundamental para estudiar muestras reales de la lengua y para crear recursos en diferentes idiomas. El tomar como base la lingüística de corpus confiere a los recursos que integran RERCOR una gran versatilidad como fuente de datos.

Los corpus que figuran en el portal han sido diseñados y compilados aplicando diferentes criterios propuestos por los principales expertos. En cuanto al diseño, se ha seguido la clasificación realizada por Laviosa (2002) para determinar qué tipo de corpus se adaptaba más a las necesidades del proyecto. En relación con los criterios de compilación, y centrándonos en la necesidad de crear corpus fiables, se han seguido las propuestas de EAGLES (1996) y se han tenido especialmente en cuenta los criterios de cantidad —también estudiados por otros autores como Berber Sardinha (2002) o Walsh (2013)— y calidad. Del mismo modo, se ha primado el criterio de autenticidad (Bowker & Pearson, 2002) garantizando que todos los textos fuesen creados en condiciones de comunicación natural. Además, siguiendo a Sinclair (2005), se ha aplicado una serie de criterios externos más concretos que determinan la composición del corpus, como los géneros y subgéneros incluidos, el grado de especialización de los textos o su fecha de publicación. Todos estos criterios son esenciales para garantizar la representatividad de los corpus compilados, aspecto primordial para autores como Biber (1993) o McEnery (2003).

Con estos antecedentes y enfoques teóricos y metodológicos concebimos el portal de recursos lingüísticos multilingües sobre el subdominio médico de las ER. El objetivo era crear una herramienta versátil, de acceso libre y ampliable (no estática).

Antes de este proyecto ya habíamos elaborado otros recursos como el Buscador sobre enfermedades neuromusculares (MYOCOR) y la Base de conocimiento terminológico sobre errores innatos del metabolismo (ONTERMET). El primero (Miquel Verges & Sánchez Trigo, 2010) es un recurso documental, creado para facilitar el acceso a la información a este grupo de ER constituido por unas 200 enfermedades. Se construyó en un momento en que el acceso a la información sobre enfermedades raras (ENM) era muy difícil por su dispersión y la práctica

---

<sup>4</sup> Existe un debate sobre si el trabajo con corpus es una metodología o constituye una línea de investigación en sí misma. Si bien no vamos a profundizar en esta cuestión por no tratarse del objetivo de este trabajo, remitimos a Parodi (2008) y Muñoz Cabrera (2017).

inexistencia de fuentes en español de interés para los pacientes y su entorno.

El segundo (Varela Vila & Sánchez Trigo, 2020) permite acceder de forma conceptualmente estructurada a la terminología de este también amplio grupo de ER en relación con la denominación de las enfermedades, síntomas y signos.

La creación de RERCOR conllevó dos fases principales. En un primer momento se llevó a cabo la compilación de un conjunto de corpus comparables y paralelos que se etiquetaron morfosintácticamente y lematizaron para poder realizar búsquedas más complejas en ellos. La lematización de los corpus se hizo con Freeling (TALP Research Center, 2020), un etiquetador de código abierto disponible, entre otras, para las lenguas de los corpus compilados (francés, inglés y español). En relación con la alineación, en el caso de los corpus paralelos, se utilizó Intertext (Vondříčka, 2016) y se realizó una alineación oracional en formato XML. Tras esta fase de compilación y tratamiento de los corpus, se pasó a la segunda fase de elaboración de las aplicaciones que integran el portal. A continuación, presentamos sus principales características.

### **3. Aplicaciones que integran RERCOR (versión 3.1.2)**

Como hemos indicado, la primera versión de RERCOR se publicó en 2018 (versión 1.0) y desde entonces hemos ido ampliando los recursos del portal e implementando nuevas funcionalidades. La versión actual RERCOR (3.1.2), que es la séptima, presenta tres aplicaciones: una aplicación principal para la consulta de corpus con un analizador de concordancias, una aplicación complementaria de consulta de vocabularios y un asistente para la redacción de textos. Presentamos a continuación sus características.

#### **3.1 Aplicación de consulta de corpus**

Se trata de la aplicación principal y permite realizar búsquedas en seis corpus integrados por 1745 textos escritos en tres lenguas (704 en francés, 344 en inglés y 697 en español), que suman 8 450 293 palabras.

Desde el punto de vista temático, dos de los corpus se centran en grupos concretos de enfermedades (metabólicas y neuromusculares). El resto lo hacen en aspectos importantes vinculados con las ER en su conjunto. Así, cuatro de los corpus abordan diversas áreas temáticas relevantes para pacientes con estas enfermedades: medicamentos huérfanos (desarrollados con apoyo institucional para tratar específicamente ER ya que no resultan interesantes a los laboratorios dada su baja rentabilidad), productos de apoyo (elementos y tecnologías que permiten compensar discapacidades derivadas de las enfermedades) y terapia ocupacional (facilita la autonomía de las personas con diversidad funcional en sus actividades de la vida diaria). El corpus restante se centra en un género textual importante en el ámbito: las guías de práctica clínica.

En la siguiente tabla (tabla 1) se recogen, junto al nombre que permite identificar cada uno de los corpus que integran la aplicación, sus datos pormenorizados en relación con la temática, número de palabras y elementos gramaticales, y lenguas.



Tabla 1: Temática, palabras ortográficas, elementos gramaticales y lenguas corpus

Corpus	Temática	Palabras ortográficas	Elementos gramaticales	Lenguas
EMCOR	corpus de enfermedades metabólicas	751 776	898 364	francés/español
ERCOR	corpus de guías de práctica clínica sobre enfermedades raras	212 865	258 110	francés/español
MYOCOR 2.0	corpus de enfermedades neuromusculares	911 269	1 109 770	francés/español
ORPHACOR	corpus de medicamentos huérfanos	4 391 283	5 241 619	francés/inglés/español
PRODAPCOR	corpus de productos de apoyo	1 046 859	1 254 763	francés/inglés/español
TERAPCOR	corpus de terapia ocupacional	1 136 241	1 365 531	francés/inglés/español

Fuente: Autoras (2024)

[Descripción] Corpus: EMCOR, Temática: corpus de enfermedades metabólicas, Palabras ortográficas: 751 776, Elementos gramaticales: 898 364, Lenguas: francés/español, Corpus: ERCOR, Temática: corpus de guías de práctica clínica sobre enfermedades raras, Palabras ortográficas: 212 865, Elementos gramaticales: 258 110, Lenguas: francés/español, Corpus: MYOCOR 2.0, Temática: corpus de enfermedades neuromusculares, Palabras ortográficas: 911 269, Elementos gramaticales: 1 109 770, Lenguas: francés/inglés/español, Corpus: ORPHACOR, Temática: corpus de medicamentos huérfanos, Palabras ortográficas: 4 391 283, Elementos gramaticales: 5 241 619, Lenguas: francés/inglés/español, Corpus: PRODAPCOR, Temática: corpus de productos de apoyo, Palabras ortográficas: 1 046 859, Elementos gramaticales: 1 254 763, Lenguas: francés/inglés/español, Corpus: TERAPCOR, Temática: corpus de terapia ocupacional, Palabras ortográficas: 1 136 241, Elementos gramaticales: 1 365 531, Lenguas: francés/español [Fin de la descripción].

Desde el punto de vista cronológico, las muestras textuales han sido publicadas entre 1993 y 2020, un período que ha permitido incluir textos representativos del ámbito. Los corpus son comparables (en cuyo caso se han aplicado criterios similares para seleccionar los textos que componen los distintos subcorpus) o paralelos (en cuyo caso los originales y sus traducciones están alineados). Tres de ellos contienen una parte paralela y una parte comparable, ya que ante las dificultades existentes para localizar textos de calidad sobre algunos aspectos de las ER, consideramos que incluir tanto textos comparables como paralelos permitía aumentar significativamente la riqueza del corpus. Como hemos indicado, los corpus están lematizados y etiquetados. En relación con los géneros textuales, todos son multigenéricos, con excepción de ERCOR, al que nos hemos referido en el párrafo anterior, que se centra en las guías de práctica clínica. El portal incluye en la actualidad 14 géneros textuales y 8 subgéneros, 5 de estos en relación con artículo científico (artículo original, caso clínico, comunicación breve, resumen y revisión) y 3 en guía de información sobre salud (guía de aspectos médicos/vida diaria, guía enfermedad, guía investigación). En la tabla 2 se presentan los datos de cada corpus en relación con estas cuestiones.

Tabla 2: Cronología, tipo y géneros textuales corpus

Corpus	Cronología	Tipo corpus	Géneros textuales
EMCOR	2001-2012	comparable	artículo científico (artículo original, caso clínico, resumen y revisión), guía de información sobre la salud (guía aspectos médicos/vida diaria y guía enfermedad)
ERCOR	2007-2013	comparable	guía de práctica clínica
MYOCOR 2.0	1993-2017	comparable (38 %) paralelo (62 %)	artículo científico (artículo original, caso clínico, comunicación breve y revisión), conferencia/comunicación en congreso, convención, guía de información sobre salud (guía aspectos médicos/vida diaria, guía enfermedad y guía investigación), guía de productos de apoyo, guía de práctica clínica, manual, monografía, noticia sobre aspectos médicos
ORPHACOR	2016-2021	paralelo	ficha técnica, prospecto



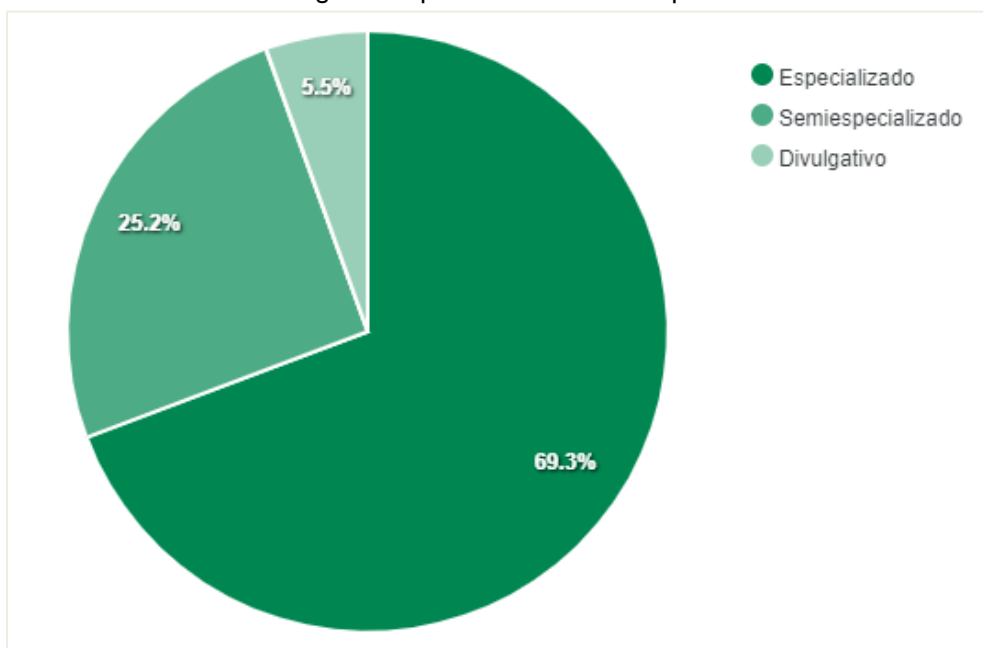
PRODAPCOR	2001-2019	comparable (40 %) paralelo (60 %)	glosario, guía de productos de apoyo, informe, manual, norma
TERAPCOR	1999-2020	comparable (60 %) paralelo (40 %)	artículo científico (artículo original) clasificación, documento de posición, guía de información sobre la salud (guía aspectos médicos/vida diaria y guía enfermedad)

Fuente: Autoras (2024)

[Descripción] Corpus: EMCOR, Cronología: 2001-2012, Tipo corpus: comparable, Géneros textuales: artículo científico (artículo original, caso clínico, resumen y revisión), guía de información sobre la salud (guía aspectos médicos/vida diaria y guía enfermedad), Corpus: ERCOR, Cronología: 2007-2013, Tipo corpus: comparable, Géneros textuales: guía de práctica clínica, Corpus: MYOCOR 2.0, Cronología: 1993-2017, Tipo corpus: comparable (38 %), paralelo (62 %), Géneros textuales: artículo científico (artículo original, caso clínico, comunicación breve y revisión), conferencia/comunicación en congreso, convención, guía de información sobre salud (guía aspectos médicos/vida diaria, guía enfermedad y guía investigación), guía de productos de apoyo, guía de práctica clínica, manual, monografía, noticia sobre aspectos médicos, Corpus: ORPHACOR, Cronología: 2016-2021, Tipo corpus: paralelo, Géneros textuales: ficha técnica, prospecto, Corpus: PRODAPCOR, Cronología: 2001-2019, Tipo corpus: comparable (40 %), paralelo (60 %), Géneros textuales: glosario, guía de productos de apoyo, informe, manual, norma, Corpus: TERAPCOR, Cronología: 1999-2020, Tipo corpus: comparable (60 %), paralelo (40 %), Géneros textuales: artículo científico (artículo original), clasificación, documento de posición, guía de información sobre la salud (guía aspectos médicos/vida diaria y guía enfermedad) [Fin de la descripción].

Como se puede observar, la mayor parte de los textos son especializados, si bien con grados de especialización diferentes, pero, por razones de representatividad (Vargas Sierra, 2006), se ha incluido asimismo un pequeño porcentaje de textos divulgativos con la finalidad de representar un lenguaje más amplio (Figura 1).

Figura 1: Especialización de los corpus



Fuente: Autoras (2024)

[Descripción de la imagen] Imagen de un gráfico circular que representa la distribución porcentual de los textos de los corpus en función de su grado de especialización. En la esquina superior derecha figuran en columna y precedidos de un punto verde con diferentes intensidades estos grados: especializados, semiespecializados, divulgativos. El gráfico circular está dividido en 3 porciones de color verde de diferente intensidad y tamaño en las que se indican porcentajes. En el sentido de las agujas del reloj, en la porción en verde más intenso y de mayor tamaño se indica 69,3 %, porcentaje de los textos especializados. A continuación, en verde un poco menos intenso y en la porción de tamaño medio se indica



25,2 %, porcentaje de los textos semiespecializados. En verde más claro y en la porción de tamaño menor se indica 5,5 % que corresponde a los textos divulgativos [Fin de la descripción].

### 3.2 Aplicación de consulta de vocabularios

Facilita la consulta de 3 vocabularios y un glosario que cuentan con un total de 1898 entradas en español, francés y/o inglés que pueden consultarse tanto de forma individual como conjunta. Estas entradas incluyen, junto a los términos equivalentes en otra lengua, la categoría gramatical, otras denominaciones o, por ejemplo, notas de uso. En la tabla 3 se resumen los datos de cada uno de los 4 recursos que en este momento integran esta aplicación.

Tabla 3: Datos vocabularios

Nombre	Lenguas	Entradas
Glosario de enfermedades neuromusculares	francés-español-inglés	114
Vocabulario de errores innatos del metabolismo	francés-español	1390
Vocabulario de productos de apoyo	francés-español	182
Vocabulario de productos de apoyo	inglés-español	212

Fuente: Autoras (2024)

[Descripción] Nombre: Glosario de enfermedades neuromusculares, Lenguas: francés-español-inglés, Entradas: 114, Nombre: Vocabulario de errores innatos del metabolismo, Lenguas: francés-español, Entradas: 1390, Nombre: Vocabulario de productos de apoyo, Lenguas: francés-español, Entradas: 182, Nombre: Vocabulario de productos de apoyo, Lenguas: francés-español, Entradas: 212 [Fin de la descripción].

Ofrecer la posibilidad de una consulta conjunta de todos conllevó un trabajo en dos fases. La primera, para identificar los campos comunes y elaborar un esquema de relaciones de equivalencia entre ellos. La segunda, para diseñar una base de datos, a partir de dicho esquema, en la que introducir las entradas y activar los campos pertinentes en cada caso.

### 3.3 Asistente para la redacción de textos

Se trata de un prototipo que facilita la redacción de textos. A medida que se escribe, propone terminología validada en francés, inglés o español. Está vinculado con las dos aplicaciones que a las que acabamos de hacer referencia, por lo que de forma rápida es posible consultar en los glosarios o en los corpus del portal la terminología propuesta con el fin de comprobar su adecuación al texto redactado.

Las tres aplicaciones del portal, que son multilingües, complementarias y están vinculadas, facilitan el acceso a recursos basados en textos reales de diferentes géneros y niveles de especialización del subdominio de las ER.

## 4. Interfaz web: funcionalidades para recuperar la información

Para acceder a las aplicaciones indicadas y recuperar la información se ha diseñado una interfaz de libre acceso al portal (<http://www.rercor.org>) que está disponible en cuatro lenguas (inglés, francés, español y gallego). En la página de inicio (ver figura 2) aparece, en la parte central,



una breve presentación tanto de los objetivos como de los contenidos del portal y, en la parte superior, figuran las pestañas que permiten el acceso a las diferentes aplicaciones del portal.

Figura 2: Página de inicio del portal



Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Veamos las opciones disponibles en cada una de estas aplicaciones.

#### 4.1 Funcionalidades de la aplicación de consulta de corpus

Se ofrecen dos tipos de búsqueda: simple y avanzada.

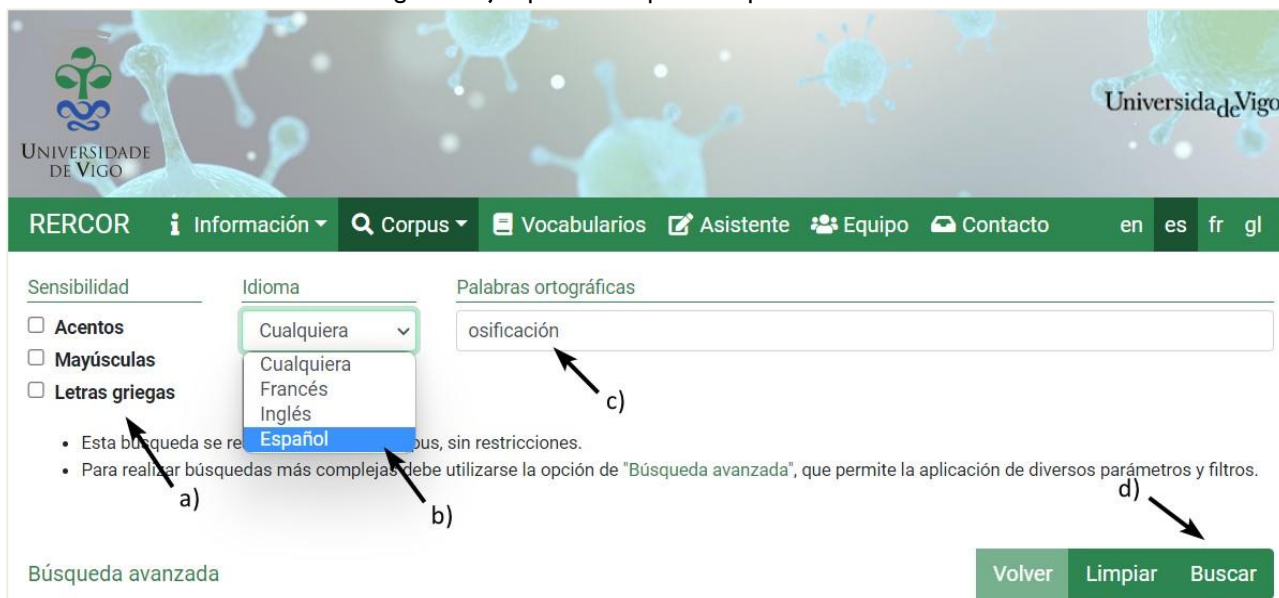
La primera está concebida para realizar consultas generales cuya finalidad es obtener resultados globales. Se adecua a ese tipo de consultas iniciales que puede querer realizar un usuario sobre un término o secuencia de hasta cinco palabras. Por ejemplo, para saber si una forma es de uso frecuente o no. La búsqueda simple, como se recoge en la figura 3, permite:

- a) seleccionar el idioma (uno o todos los disponibles: francés, inglés y español);
- b) decidir si tener en cuenta o no acentos, mayúsculas y letras griegas ( $\beta$ ,  $\gamma$ , etc. frecuentes en textos de este ámbito). Esto es útil, por ejemplo, para comprobar si una forma se utiliza

mayoritariamente con o sin mayúsculas ('Apgar'/'APGAR') o con o sin acento ('oftalmoplejía'/'oftalmoplejía').

Se introduce en la caja de búsqueda el término o secuencia de hasta cinco palabras (c) y se pincha en el botón 'Buscar' (d). A modo de ejemplo, hemos buscado 'osificación'):

Figura 3: Ejemplo de búsqueda simple: 'osificación'



Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Los resultados, como se presenta en la figura 4, se muestran en forma de concordancias. Es decir, cada coincidencia con la búsqueda realizada aparece en una línea distinta con la forma o formas consultadas en el centro destacadas en negrita y con el contexto anterior y posterior más cercano. Se incluye asimismo información sobre el corpus del que procede, idioma y año. La búsqueda se realiza en todos los corpus incluidos en el portal, pero puede haber resultados solo en algunos de ellos.

Figura 4: Ejemplo de resultados en búsqueda simple: 'osificación'

Resultados 1 a 24 de 24		
1	EMCOR / es / 2004	minución de las siluetas renales; centros de <b>osificación</b> correspondientes a la edad pero con franco retardo de madurez
2	MYOCOR / es / 2017	• Los brotes dolorosos seguidos de <b>osificación</b> de los músculos que se vuelven "duros como piedras" son resp
3	ORPHACOR / es / 2019	nómalo de costillas o vértebras lumbares y <b>osificación</b> deficiente en numerosos huesos.
4	ORPHACOR / es / 2021	ales vasos sanguíneos, una reducción de la <b>osificación</b> y una disminución del peso corporal del feto.
5	ORPHACOR / es / 2019	l peso fetal (acompañado de retrasos de la <b>osificación</b> del esqueleto) en ratas y conejos.
6	MYOCOR / es / 2003	ón, al inicio de los brotes podría prevenir la <b>osificación</b> .
7	MYOCOR / es / 2003	El «Simposio internacional sobre la <b>osificación</b> ectópica en especial en la FOP» ha reunido en Oxford, del 20 al
8	ORPHACOR / es / 2019	disminución del peso fetal con retraso de la <b>osificación</b> , fetos ligeramente edematosos y aumento de la incidencia de c
9	ORPHACOR / es / 2021	histopatológica entre esta observación y la <b>osificación</b> , la fibrosis y la inflamación.
10	ORPHACOR / es / 2018	aciones esqueléticas (sitios adicionales de <b>osificación</b> frontal/parietal).
11	ORPHACOR / es / 2018	ida la disminución de los pesos fetales y la <b>osificación</b> esquelética retardada.
12	ORPHACOR / es / 2019	cción del peso fetal promedio, retraso en la <b>osificación</b> e incidencias ligeramente mayores de anomalías esqueléticas r
13	ORPHACOR / es / 2018	ciones esqueléticas menores, retraso de la <b>osificación</b> y variantes de las esternebras) se corresponden a dosis equival
14	MYOCOR / es / 2003	alidad esencial retrasar los fenómenos de <b>osificación</b> .
15	ORPHACOR / es / 2019	rduladas fetales y una menor incidencia de <b>osificación</b> esquelética fetal.
16	ORPHACOR / es / 2018	días, una disminución del peso fetal y de la <b>osificación</b> del esqueleto.
17	ORPHACOR / es / 2019	rada; ausencia de osificación de la pelvis u <b>osificación</b> incompleta; aumento medio de las costillas torácicas supernun
18	ORPHACOR / es / 2019	morfología hepática alterada; ausencia de <b>osificación</b> de la pelvis u osificación incompleta; aumento medio de las cos
19	ORPHACOR / es / 2019	l disminución del peso fetal con retraso de la <b>osificación</b> , fetos ligeramente edematosos y aumento de la incidencia de c
20	MYOCOR / es / 2003	La <b>osificación</b> muscular derivada de un brote puede detectarse también medik
21	ORPHACOR / es / 2019	cción del peso corporal fetal y retraso de la <b>osificación</b> ) a la dosis más alta probada (200 mg/kg/día).
22	MYOCOR / es / 2003	á «transformación» del músculo en hueso ( <b>osificación</b> ).
23	MYOCOR / es / 2017	a un tratamiento susceptible de prevenir la <b>osificación</b> secundaria de los músculos.
24	EMCOR / es / 2001	La fisonomía de la <b>osificación</b> endocondral y el des arrollo de cambios osteoartríticos, particu

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

El contexto de cada concordancia se puede ampliar haciendo clic en el número que figura a su izquierda (indicado en la columna izquierda de la figura 4). De esta forma, se muestran la frase de la coincidencia, las dos frases anteriores y las dos posteriores. Además, si se trata de textos paralelos, se incluye, como se muestra en la figura 5, la frase de la coincidencia en las otras lenguas disponibles en función del corpus.

Figura 5: Ejemplo de ampliación de contexto

<b>Corpus:</b> ORPHACOR	<b>Idioma:</b> Español	<b>Género:</b> Ficha técnica	<b>Subgénero:</b> Cualquiera
<b>Año:</b> 2019	<b>Especialización:</b> Especializado	<b>ID de documento:</b> ORFI094ES	
<b>Título:</b> Torisel - Ficha técnica			
No hay datos adecuados relativos al uso de temsirolimus en mujeres embarazadas.			
Los estudios realizados en animales han mostrado toxicidad para la reproducción.			
<input checked="" type="checkbox"/> En los estudios de reproducción en animales, temsirolimus causó toxicidad embriofetal que se manifestó con mortalidad y disminución del peso fetal (acompañado de retrasos de la <b>osificación</b> del esqueleto) en ratas y conejos.			
<input type="checkbox"/> In reproduction studies in animals, temsirolimus caused embryo/foeto toxicity that was manifested as mortality and reduced foetal weights (with associated de lays in skeletal ossification) in rats and rabbits.			
<input checked="" type="checkbox"/> Les études de reproduction chez l'animal ont montré l'embryo/foetotoxicité du temsirolimus, qui se manifestait par une augmentation de la mortalité et une réduction du poids foetal (associé à un retard pour l'ossification du squelette) chez le rat et le lapin.			
Se observaron efectos teratogénicos (onfalocele) en conejos (ver sección 5.3).			
Se desconoce el riesgo potencial en el ser humano.			

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Junto a la posibilidad de búsqueda simple, a la que acabamos de referirnos, la segunda opción disponible en la aplicación de consulta de corpus es la búsqueda avanzada. Con ella se puede realizar una consulta más definida y completa, ya que ofrece un conjunto de funcionalidades muy amplio. Como se presenta en la figura 6, las diferentes secciones diseñadas en la interfaz cuentan con desplegables en los que se puede elegir:

- a) en qué corpus deseamos realizar las búsquedas (en uno, en varios o en todos los disponibles);
- b) qué tipo de búsqueda queremos hacer: palabras ortográficas o elementos gramaticales. En ambos casos la búsqueda permite incluir hasta 5 palabras o elementos. En la búsqueda por elementos gramaticales se pueden combinar en una misma consulta palabras ortográficas, lemas, etiquetas y elementos gramaticales, de modo que puedan localizarse fácilmente. Por ejemplo, permite localizar verbos como *s'accumuler* en francés (formado por dos elementos gramaticales), como 'administrársele' en español (con pronombres enclíticos), todas las ocurrencias del verbo 'diagnosticar' en cualquier tiempo, o términos que se correspondan con una determinada estructura morfosintáctica (p. ej., buscar sustantivo + prep. de + nombre propio para localizar epónimos). También pueden realizarse búsquedas por palabras gramaticales próximas y elementos gramaticales próximos, que permiten localizar contextos en que una palabra o elemento se encuentre cerca de otra (a una distancia de entre 1 y 10);

- c) marcar o desmarcar la sensibilidad a acentos, mayúsculas y letras griegas en la consulta.

Figura 6: Búsqueda avanzada: selección corpus, tipo búsqueda y sensibilidad

The screenshot shows the RERCOR search interface. The top navigation bar includes 'RERCOR', 'Información', 'Corpus', 'Vocabularios', 'Asistente', 'Equipo', and 'Contacto', along with language options 'en es fr gl'. The main search area is divided into 'Búsqueda' and 'Resultado' sections. In the 'Búsqueda' section, there are three sub-sections: 'Corpus' with a dropdown menu showing 'ORPHACOR' (labeled 'a'), 'Tipo' with a dropdown menu showing 'Palabras ortográficas' (labeled 'b'), and 'Sensibilidad' with three checkboxes: 'Acentos', 'Mayúsculas', and 'Letras griegas' (labeled 'c'). The 'Resultado' section includes 'Tipo de resultado' (set to 'Expresiones coinc'), 'Ordenación' (set to 'Total' and 'Coincidencia'), and 'Tamaño de página' (set to '50'). Below this is the 'Filtros' section with dropdowns for 'Género' (Prospecto), 'Subgénero' (Cualquiera), 'Especialización' (Cualquiera), 'Documentos' (Cualquiera), 'Idioma' (Francés), 'Desde' (2020), and 'Hasta' (2021). At the bottom, there is a search input field containing 'hypo\*' and buttons for 'Descargar', 'Volver', 'Limpiar', and 'Buscar'.

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Del mismo modo, (ver figura 7) puede indicarse cómo queremos que se muestren los resultados:

- en forma de frecuencia simple (número de ocurrencias, frecuencia normalizada y número de documentos en que figuran);
- frecuencia completa (datos estadísticos más completos, por género textual, grado de especialización e idioma, con la posibilidad de crear gráficas);
- concordancias, cuyas características ya se han explicado más arriba;
- expresiones coincidentes (lista de palabras o secuencias, resultado de la búsqueda realizada, ordenadas por frecuencia);
- colocaciones (sirve para analizar qué palabras o elementos se usan habitualmente con la palabra o secuencia que estamos buscando).

La aplicación permite seleccionar también:

- cómo desean ordenarse y agruparse los resultados;
- cuántos resultados desean visualizarse por página.

Figura 7: Búsqueda avanzada: tipo de resultados

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Por último, como se presenta en la figura 8, es posible filtrar los resultados utilizando distintos criterios:

- 1) género y subgénero textual (seleccionando alguno de los que hemos presentado en el apartado precedente);
- 2) grado de especialización (especializado, semiespecializado, divulgativo);
- 3) tipo de documento (paralelo o comparable);
- 4) idioma (español, francés, inglés);
- 5) fecha de publicación (entre 1993 y 2021).

Si no hace una selección específica de las diferentes posibilidades que se ofrecen en los despleables, se recuperan los datos disponibles en todas ellas.

Figura 8: Búsqueda avanzada: filtros

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

En la parte inferior de la pantalla debe introducirse la búsqueda. El formulario cambia en función del tipo de búsqueda mencionado anteriormente (palabras ortográficas, elementos gramaticales...). La opción de expresiones coincidentes permite obtener listados de todas las palabras o elementos gramaticales del corpus o subcorpus seleccionado ordenados por frecuencia, introduciendo, para ello, un asterisco en la caja de búsqueda de la sección 'palabras ortográficas'.

El ejemplo que sigue, en la figura 9, muestra precisamente cómo buscar palabras ortográficas que comiencen por el prefijo *hypo-* (introduciendo en la caja de búsqueda *hypo* seguido de asterisco, es decir, *hypo\**), en prospectos publicados entre el año 2020 y el 2021 y que los resultados se muestren en forma de expresiones coincidentes<sup>5</sup>:

<sup>5</sup> En relación con las búsquedas con comodines, como en este caso en que se ha utilizado un asterisco, ver el final de este apartado.



Figura 9: Búsqueda avanzada: consulta prefijo *hypo-* en francés

The image shows the RERCOR search interface. At the top, there is a navigation bar with the RERCOR logo and links for 'Información', 'Corpus', 'Vocabularios', 'Asistente', 'Equipo', and 'Contacto'. The language is set to 'fr' (French). The search interface is divided into several sections:

- Búsqueda:** Includes a search bar with the text 'Palabras ortográficas' and a dropdown menu for 'Corpus' set to 'ORPHACOR'.
- Resultado:** Includes dropdowns for 'Tipo de resultado' (set to 'Expresiones coinc...'), 'Ordenación' (set to 'Total' and 'Coincidencia'), and 'Tamaño de página' (set to '50').
- Tipo:** A dropdown menu set to 'Palabras ortográficas'.
- Agrupar:** A dropdown menu set to 'Palabra ortográfica'.
- Sensibilidad:** Checkboxes for 'Acentos', 'Mayúsculas', and 'Letras griegas'.
- Filtros:** Includes dropdowns for 'Género' (set to 'Prospecto'), 'Subgénero' (set to 'Cualquiera'), 'Especialización' (set to 'Cualquiera'), 'Documentos' (set to 'Cualquiera'), 'Idioma' (set to 'Francés'), 'Desde' (set to '2020'), and 'Hasta' (set to '2021').

At the bottom, there is a search bar with the text 'hypo\*' and a button labeled 'Buscar'. Other buttons include 'Descargar', 'Volver', and 'Limpiar'.

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

De este modo, como se muestra en la figura 10, se recuperan todas las formas que comienzan por *hypo-* en francés y se presentan ordenadas por frecuencia (se indica a la derecha el número de ocurrencias/número de documentos). Haciendo clic en el número de ocurrencias/documentos podemos acceder a las concordancias de la forma en cuestión.

Figura 10: Búsqueda avanzada: resultados prefijo *hypo-* en francés

		Total (88 272 / 30)
1	hypotension	7 / 6
2	hyponatrémie	4 / 2
3	hypoxie	3 / 2
4	hypoglycémie	2 / 2
5	hypoactive	1 / 1
6	hypoalbuminémie	1 / 1
7	hypocalcémie	1 / 1
8	hypokaliémie	1 / 1
9	hypomagnésémie	1 / 1
10	hypophosphatémie	1 / 1
11	hypothermie	1 / 1
12	hypothyroïdie	1 / 1
Resultados 1 a 12 de 12 (24)		

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Lo que mostramos es solo un pequeño ejemplo de las búsquedas que pueden realizarse con esta aplicación, que es muy potente y versátil. Por ejemplo, RERCOR es útil para recuperar todas las palabras o grupos de hasta 5 palabras ordenados por frecuencia o para comprobar de forma rápida una colocación (p. ej., para saber que se dice *traité par placebo* y no *traité avec placebo*).

La aplicación, que está creada para poder establecer subcorpus específicos, permite a los usuarios realizar una selección textual que se adapte a sus necesidades y a sus proyectos de trabajo. Se trata de una utilidad que sirve, por ejemplo, para comprobar si un determinado rasgo lingüístico es común a distintos géneros y grados de especialización, así como su comportamiento en diferentes lenguas.

Tanto en la búsqueda simple como en la avanzada, la aplicación de corpus permite emplear expresiones regulares (comodines) para sustituir caracteres y no condicionar los resultados. Se pueden emplear de forma individual o combinada para delimitar la búsqueda, lo que aporta versatilidad y permite obtener resultados diversos de manera rápida. En concreto:

- a) el signo de interrogación [?] sustituye a un carácter: por ejemplo, con el fin de identificar las distintas ocurrencias de una palabra que contenga un carácter variable (por ejemplo, 'glucogenosis' y 'glicogenosis'), se puede realizar una búsqueda reemplazando ese carácter por una interrogación;
- b) el asterisco [\*] sustituye a cero, uno o varios caracteres: se puede emplear para buscar todas las palabras acabadas por un sufijo, por ejemplo -oide, obteniendo resultados como *thyroïde*, 'linfoïde' o *myéloïde*. Además, en la opción elementos gramaticales, se puede especificar la categoría gramatical de la búsqueda, lo que permite buscar, por ejemplo,



todos los adverbios terminados en un sufijo determinado. También se pueden consultar todas las palabras que comienzan por un prefijo, como cardio-, introduciendo el asterisco después de este. De la misma forma, al introducir el asterisco en medio de una palabra, se pueden encontrar las ocurrencias que empiezan por un prefijo, como eco-, y terminan por un sufijo, por ejemplo -grafía, obteniendo tanto 'ecografía' como 'ecocardiografía'. Además, el asterisco se puede introducir más de una vez dentro de la misma palabra, por ejemplo, para localizar todas las ocurrencias que comienzan por el prefijo eco- y terminan por el sufijo -grafic, con género y número variable;

- c) el signo de exclamación [!] equivale a 'no': restringe los resultados, de modo que en ellos figurarán todas las ocurrencias menos aquellas que coincidan con las formas a las que hemos antepuesto el signo de exclamación;
- d) la barra vertical [|] permite ofrecer alternativas: por ejemplo, para buscar un nombre con dos grafías distintas (McArdle|MacArdle).

Estos comodines pueden también combinarse con el fin de realizar búsquedas más precisas.

Los resultados de todas las búsquedas pueden, como se recoge en la figura 11, descargarse y exportarse a un archivo en formato CSV. De este modo, se pueden analizar los datos con detenimiento y utilizarlos como se considere oportuno<sup>6</sup>.

Figura 11: Resultados en formato CSV

Column1	Column2	Column3	Column4	Column5	Column6	Col
Corpus	Idioma	Género	Subgénero	Especialización	Título	Contexto a la izquierda
MYOCOR 2.0	Francés	Monografía	Cualquiera	Semiespecializado	Dystrophie myotonique de Steinert	de la succion : l'incapacit
EMCOR	Francés	Artículo científico	Caso clínico	Especializado	Grossesse chez une patiente atteinte de cytopathie mit...	L'atteinte peut toucher d
MYOCOR 2.0	Francés	Guía de información sobre salud	Guía enfermedad	Especializado	Principales maladies neuromusculaires	• Fatigue musculaire et m
MYOCOR 2.0	Francés	Monografía	Cualquiera	Especializado	Maladies de Charcot-Marie-Tooth. Rééducation (princi...	Heureusement, la désorg
EMCOR	Francés	Guía de información sobre salud	Guía enfermedad	Semiespecializado	Céroïde Lipofuscinoses Neuronales	L'IRM montre rapidemen
EMCOR	Francés	Artículo científico	Revisión	Especializado	L'enfant souffrant d'une maladie métabolique connue ...	Vertèbres ovoïdes, parfoi
ORPHACOR	Francés	Ficha técnica	Cualquiera	Especializado	ADCETRIS - Résumé des caractéristiques du produit	Une
ORPHACOR	Francés	Prospecto	Cualquiera	Semiespecializado	Zolgensma - Notice	Cela provoque une faibles
MYOCOR 2.0	Francés	Monografía	Cualquiera	Semiespecializado	Critères diagnostiques des maladies neuromusculaires	Déficit et
EMCOR	Francés	Artículo científico	Artículo original	Especializado	Hypotrophie et manifestations digestives des désordres...	chorion sans atrophie vill
MYOCOR 2.0	Francés	Monografía	Cualquiera	Semiespecializado	Critères diagnostiques des maladies neuromusculaires	, des douleurs Musculaire
EMCOR	Francés	Artículo científico	Revisión	Especializado	Néphropatie de l'adulte par mutation A3243G de l'ADN...	L'imagerie peut déceler d
EMCOR	Francés	Artículo científico	Revisión	Especializado	Manifestations ORL de la maladie de Fabry	révélaît la présence de gh
EMCOR	Francés	Artículo científico	Artículo original	Especializado	Hypotrophie et manifestations digestives des désordres...	Les 4 patients atteints de
EMCOR	Francés	Artículo científico	Revisión	Especializado	Adrénoleucodystrophie liée à l'X	L'apparition d'une
ORPHACOR	Francés	Ficha técnica	Cualquiera	Especializado	BESPONSA - Résumé des caractéristiques du produit	répétées menées chez les
MYOCOR 2.0	Francés	Guía de información sobre salud	Guía enfermedad	Especializado	Principales maladies neuromusculaires	• Examens complémentai
MYOCOR 2.0	Francés	Guía de información sobre salud	Guía enfermedad	Especializado	Principales maladies neuromusculaires	• Faiblesse et
MYOCOR 2.0	Francés	Monografía	Cualquiera	Semiespecializado	Dystrophie myotonique de Steinert	L'histologie des muscles r

Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

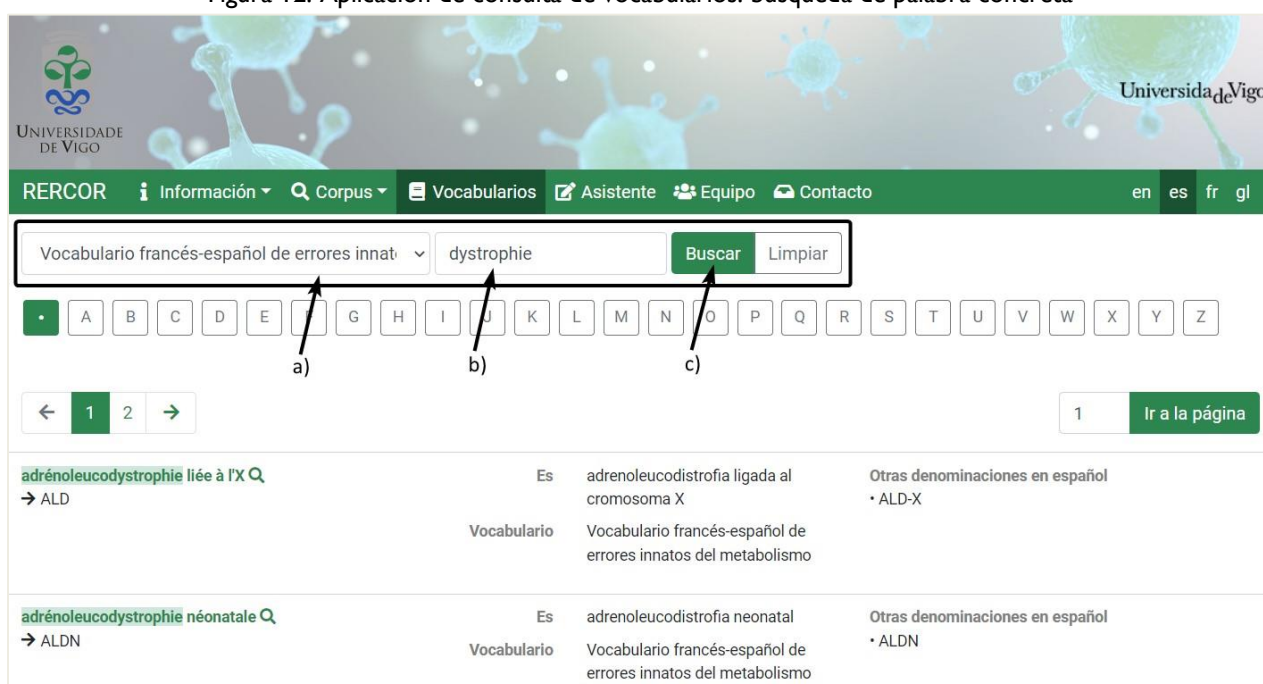
<sup>6</sup> Puede obtenerse más información sobre cómo abrir el documento descargado en el sitio web del portal RERCOR. Concretamente, en la sección Preguntas frecuentes de la pestaña superior Información.

## 4.2 Funcionalidades de la aplicación consulta de vocabularios

Esta aplicación permite dos formas de búsqueda: mediante la consulta de una palabra concreta y mediante el listado alfabético.

En el primer caso, de acuerdo con la figura 12, debe seleccionarse en el desplegable el vocabulario que se desea consultar —que puede ser uno o todos (a)—, introducir la palabra (b) y pinchar en buscar (c). De este modo se mostrarán todas las entradas en las que haya un término que coincida con la búsqueda realizada independientemente del campo en el que figure la correspondencia y de la lengua. Las correspondencias se destacan en verde.

Figura 12: Aplicación de consulta de vocabularios: búsqueda de palabra concreta



Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Es posible, asimismo, consultar las entradas a partir del listado alfabético, pinchando en la letra correspondiente. Al hacer clic en el punto que figura al inicio de dicho listado, se vuelven a mostrar todos los términos del vocabulario o vocabularios seleccionados por orden alfabético.

En ambos sistemas de búsqueda la lupa que figura junto a algunas de las entradas permite mostrar las ocurrencias de ese término en la aplicación de corpus.

## 4.3 Funcionalidades del asistente de redacción

Se trata de un prototipo que sugiere, según el idioma que se seleccione y la secuencia que se escriba en el cuadro de texto, los términos que estén recogidos en alguno de los vocabularios y que coincidan con el texto redactado. Para utilizarlo, como se indica en la figura 13, debe seleccionarse la lengua (a) y comenzar a escribir el texto que se desee (b). La aplicación propone en el cuadro de la derecha (c) los términos que figuran en los vocabularios.

Figura 13: Asistente de escritura: proceso de escritura



Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Si se pincha el término deseado, este se incluye en el texto sin necesidad de teclear. Además, como se muestra en la figura 14, todos los términos del cuadro de texto que consten en una o más entradas de los vocabularios aparecerán destacados en negrita y color verde. A medida que se escribe, van apareciendo nuevas sugerencias.

Figura 14: Asistente de escritura: fragmento de texto resultante



Fuente: Autoras (2024)

[Descripción de la imagen] Conforme a lo mencionado en el texto [Fin de la descripción].

Asimismo, al clicar el icono que figura después de cada término sugerido se puede consultar esa entrada en el vocabulario en el que está incluido. Además, como ya indicamos, al clicar el icono de la lupa que se encuentra a la derecha de algunas entradas de los vocabularios, es posible acceder directamente a las correspondencias de dicho término en el corpus. Por tanto, las tres aplicaciones están vinculadas, lo que permite sacar el máximo rendimiento de todas las funcionalidades del portal.

## 5. Consideraciones finales

Como hemos mostrado, las aplicaciones y funcionalidades de RERCOR (3.1.2) son numerosas. Cuando estamos traduciendo, podemos usarlas, por ejemplo, para:

- comprobar si es más habitual ‘ortesis’ u ‘órtesis’, ‘hemiplejia’ o ‘hemiplejía’;
- verificar si ‘ELISA’ (nombre de una técnica) es una sigla o un epónimo y, por lo tanto, si debe escribirse en mayúsculas o solo con mayúscula inicial;
- analizar si para traducir el inglés *active ingredient* es más frecuente decir ‘sustancia activa’, ‘principio activo’ o ‘ingrediente activo’;
- decidir si en textos especializados se puede usar ‘tomar quimioterapia’ o si es preferible usar ‘recibir quimioterapia’;
- saber si la colocación *cliniquement pertinente* en francés, suele traducirse al español con más frecuencia por clínicamente ‘pertinente’, ‘relevante’, ‘significativo’ o ‘importante’;
- confirmar si debemos decir ‘ensayo clínico de fase I’ o ‘ensayo clínico en fase I’;
- buscar qué preposición o preposiciones se utilizan en francés con el verbo *traiter* y sus analizar sus contextos;
- estudiar qué sustantivos suelen seguir en inglés a *therapeutic*;
- analizar las ocurrencias de *cases of* y *have been* en contextos en que se encuentren a una distancia menor o igual a 10 palabras;
- obtener todas las colocaciones formadas por cualquier palabra más *hématologique* con sus variantes de género y número;
- extraer listados de bigramas, trigramas, etc. introduciendo dos, tres o el número correspondiente de asteriscos;
- consultar ‘Addison’ en los vocabularios, encontrar *maladie d’Addison* y analizar sus ocurrencias en los corpus.

Estas son solo algunas de las utilidades del portal, que permite acceder a corpus de textos comparables y paralelos lematizados y etiquetados, establecer corpus *ad hoc* para adaptarse a los intereses de cada usuario, obtener terminología real en contexto, analizar los resultados de múltiples formas (documentar usos, conocer frecuencias, identificar diferencias, registrar combinaciones habituales...) y exportar los resultados. Por otra parte, RERCOR 3.1.2 es útil para alimentar herramientas habituales de traducción como, por ejemplo, MemoQ, que mediante LiveDocs permite añadir corpus para ser utilizados en los proyectos de traducción, así como crear memorias y bases de datos terminológicas.

En el momento en que iniciamos este proyecto, nuestro objetivo fue crear un portal versátil y fácilmente ampliable. Esta característica ha permitido que hayamos ido enriqueciéndolo. Así, si se



compara RECOR 3.1.2, con la primera versión que construimos (Sánchez Trigo & Varela Vila, 2019), se puede apreciar que, como apuntamos más arriba, la versión actual presenta diferencias relevantes. Por un lado, se ha desarrollado una nueva aplicación para la redacción de textos, que inicialmente no existía y sobre la que pretendemos seguir trabajando para mejorarla. Por otro lado, se han ampliado mucho tanto el tamaño como las funcionalidades disponibles en de las aplicaciones de consulta de corpus y de vocabularios. En la primera se han incluido más corpus (tras la compilación y tratamiento de 3 nuevos corpus) y más lenguas (además del francés y el español, ahora también el inglés). Estos cambios han conllevado un importante incremento del número de palabras (en 6 289 255 más), de textos (se han añadido 1033 textos) y de géneros (ahora 16, frente a los 8 géneros iniciales), lo que confiere mayor representatividad a los corpus. En relación con la aplicación de consulta de vocabularios, además de reformular y reorganizar la estructura de las entradas, se ha creado un nuevo recurso sobre productos de apoyo con 182 nuevas entradas en francés y español. Las entradas de los vocabularios recogen un número importante de sinónimos, equivalentes y notas de uso, aspecto que hace que la consulta de los vocabularios aporte una información relevante. Por otra parte, la vinculación entre las tres aplicaciones disponibles que permite realizar consultas combinadas confiere versatilidad a las búsquedas y a los resultados. De este modo se contribuye a un mejor conocimiento y uso de la terminología propia tanto del campo específico de las ER como de la medicina en general, imprescindible para redactar textos.

Como ayuda para facilitar las consultas, en la pestaña Información se puede acceder a datos sobre los distintos recursos, tutoriales y preguntas frecuentes sobre el portal.

Confiamos en poder seguir poniendo a disposición de los usuarios nuevos corpus y vocabularios, y de este modo seguir ayudando a traductores y profesionales de actividades afines, así como a docentes e investigadores, del ámbito biomédico.

## Referencias

- AENOR (2015). *Servicios de traducción. Requisitos para los servicios de traducción*. (Norma UNE-EN ISO 17100).
- AFM-Téléthon. (2022, 9 de junio). *Nos missions. AFM-Téléthon : innover pour guérir*. <https://www.afm-telethon.fr/fr/l-association/qui-sommes-nous/nos-missions>
- Alliance Maladies Rares. (s. f.). *Nos combats*. <https://alliance-maladies-rares.org/nos-combats/>
- Berber Sardinha, T. (2002). Tamanho de corpus. *The ESpecialist*, 23(2), 103–122.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243–257. <https://doi.org/10.1093/lc/8.4.243>
- Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. Routledge.
- Cabré, M. T. (1999). *La terminología: representación y comunicación*. Instituto Universitario de Lingüística Aplicada.
- Cabré, M. T. (2003). Theories of Terminology: Their Description, Prescription and Explanation. *Terminology*, 9(2), 163–199. <https://doi.org/10.1075/term.9.2.03cab>
- Cabré, M. T. (2009). La Teoría Comunicativa de la Terminología, una aproximación lingüística a los términos. *Revue Française de Linguistique Appliquée*, 14(2), 9–15. <https://doi.org/10.3917/rfla.142.0009>



- EAGLES. (1996). *Preliminary Recommendations on Corpus Typology*, EAG-TCWG-CTYP/P. <https://www.ilc.cnr.it/EAGLES96/corpusyp/corpusyp.html>
- EURORDIS (Rare Diseases Europe). (s. f.). *What is a rare disease?* <https://www.eurordis.org/information-support/what-is-a-rare-disease>
- Fisbach, H. (Ed.). (1998). *Translation and Medicine*. John Benjamins.
- Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, Findings, Applications*. Rodopi.
- López-Rodríguez, C. I. (2016). Using Corpora in Scientific and Technical Translation Training: Resources to Identify Conventionality and Promote Creativity. *Cadernos de Tradução*, 36(1), 88–120. <https://doi.org/10.5007/2175-7968.2016v36nesp1p88>
- McEnery, T. (2003). Corpus Linguistics. In M. Ruslan (Ed.), *Oxford Handbook of Computational Linguistics* (pp. 448–463). Oxford University Press.
- MemoQ (s.f.). <https://www.memoq.com/>
- Miquel Verges, J., & 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. <https://doi.org/10.7202/044246ar>
- Montalt Ressorecció, V., & González Davies, M. (2007). *Medical Translation Step by Step*. St. Jerome.
- Montalt, V., Zethsen, K., & Karwacka, W. (2018). Medical translation in the 21st century - Challenges and Trends. *MonTI*, 10, 27–42. <http://dx.doi.org/10.6035/MonTI.2018.10.1>
- Muñoz Cabrera, R. (2017). *Glosario militar de la Segunda Guerra Mundial. Trabajo de extracción terminológica y de lingüística de corpus mediante el análisis de textos de época*. [Tesis doctoral]. Universidad de Sevilla. <https://bit.ly/3qkWyWN>
- Muñoz Miquel, A. (2009). El perfil del traductor médico: diseño de un estudio de corte socioprofesional. *Panace@*, 10(30), 157–167.
- Muñoz Miquel, A. (2014). El perfil y las competencias del traductor médico desde el punto de vista de los profesionales: una aproximación cualitativa. *TRANS: revista de traductología*, 18, 163–181. <https://doi.org/10.24310/TRANS.2014.v0i18.3251>
- Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, Ch., Lanneau, V., Murphy, D., Le Cam, Y., & Rath, A. (2020). Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database. *European Journal of Human Genetics*, 28, 165–173. <https://doi.org/10.1038/s41431-019-0508-0>
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 93–119. <http://dx.doi.org/10.4067/S0718-48832008000100006>
- Pearson, J. (1988). *Terms in Context*. John Benjamins.
- Sánchez Trigo, E., & Varela Vila, T. (2018-2024). *RERCOR: Recursos sobre enfermedades raras (versión 3.1.2)*. Universidade de Vigo. <http://www.rercor.org>
- Sánchez Trigo, E., & Varela Vila, T. (2019). RERCOR: portal de recursos lingüísticos multilingües sobre enfermedades raras. *Estudios de traducción*, 9, 131–150. <https://doi.org/10.5209/estr.65706>
- Sinclair, J. (2005). Corpus and Text - Basic Principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1–16). Oxbow Books.
- SIO ALIBER. (2020). *Iberoamérica unida por las Enfermedades Raras*. Alianza Iberoamericana de Enfermedades Raras. <https://bit.ly/3VFFFvs>
- TALP Research Center. (2020). *Freeling 4.1 User Manual*. Universitat Politècnica de Catalunya. <https://freeling-user-manual.readthedocs.io/en/v4.1/>
- Van Hoof, H. (2001). Notes pour une histoire de la traduction pharmaceutique. *Meta*, 46(1), 154–175. <https://doi.org/10.7202/004551ar>
- Varela Vila, T., & Sánchez Trigo, E. (2020). Diseño e implementación de una base de conocimiento terminológico sobre enfermedades raras. *Onomázein*, 49, 1–20. <https://doi.org/10.7764/onomazein.49.01>



- Vargas Sierra, C. (2006). Diseño de un corpus especializado con fines terminográficos: el Corpus de la Piedra Natural. *Debate Terminológico*, 2(7).
- Vondříčka, P. (2016). *InterText Editor v1.5 Comprehensive Guide*. Institute of the Czech National Corpus Charles University, Faculty of Arts. [https://wanthalf.saga.cz/repository/intertext/InterText\\_Editor-User\\_Guide.pdf](https://wanthalf.saga.cz/repository/intertext/InterText_Editor-User_Guide.pdf)
- 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). Springer.

## Notas

### Contribución de autoría

**Concepción y elaboración del manuscrito:** E. Sánchez Trigo, T. Varela Vila

**Recolección de datos:** E. Sánchez Trigo, T. Varela Vila

**Análisis de datos:** E. Sánchez Trigo, T. Varela Vila

**Discusión de los resultados:** E. Sánchez Trigo, T. Varela Vila

**Revisión y aprobación:** E. Sánchez Trigo, T. Varela Vila

### Datos de la investigación

No se aplica.

### Financiación

No se aplica.

### Derechos de uso de la imagen

No se aplica.

### Aprobación de comité de ética en investigación

No se aplica.

### Conflicto de intereses

No se aplica.

### Licencia de uso

Los autores ceden a *Cadernos de Tradução* los derechos exclusivos de primera publicación, con el trabajo simultáneamente licenciado bajo la [Licencia Creative Commons](#) Atribución 4.0 Internacional (CC BY). Esta licencia permite a terceros remezclar, adaptar y crear a partir del trabajo publicado, otorgando el crédito adecuado de autoría y publicación inicial en esta revista. Los autores están autorizados a celebrar contratos adicionales por separado para distribuir de manera no exclusiva la versión del trabajo publicado en esta revista (por ejemplo, publicarlo en un repositorio institucional, en un sitio web personal, en redes sociales académicas, realizar una traducción o republicar el trabajo como un capítulo de libro), siempre y cuando se reconozca la autoría y la publicación inicial en esta revista.

### Publisher

*Cadernos de Tradução* es una publicación del Programa de Posgrado en Estudios de Traducción de la Universidad Federal de Santa Catarina. La revista *Cadernos de Tradução* está alojada en el [Portal de Periódicos UFSC](#). Las ideas expresadas en este artículo son responsabilidad de sus autores y no representan necesariamente la opinión del equipo editorial o de la universidad.

### Editores del número especial

Andréia Guerini – Fernando Ferreira Alves – Orlando Grossegeisse

### Editor de sección

Willian Moura



## Corrección de normas

Alice S. Rezende – Ingrid Bignardi – João G. P. Silveira – Kamila Oliveira

## Histórico

Recibido el: 03-06-2023

Aprobado el: 17-02-2024

Revisado el: 02-03-2024

Publicación: 04-2024

