



Universitat
Pompeu Fabra
Barcelona

Treball de fi de màster de Recerca

Una aproximación al uso de la neología especializada en el ámbito de las enfermedades raras

Jorge Mario Porras Garzón

Màster: Lingüística Teòrica i Aplicada

Edició: 2015-2016

Directors: Dra. Rosa Estopà Bagot

Any de defensa: 2016

Col·lecció: Treballs de fi de màster

Programa oficial de postgrau

"Comunicació lingüística i mediació multilingüe"

Departament de Traducció i Ciències del Llenguatge

RESUMEN

Este trabajo pretende hacer una aproximación al uso que se le da a la neología especializada en el ámbito de las enfermedades raras, en los textos dirigidos a especialistas y en los textos dirigidos a pacientes y sus familiares. La finalidad de este trabajo era construir dos corpus diferentes para hacer una detección de las unidades terminológicas nuevas y luego un análisis que permitió observar cuáles eran los recursos de formación de neologismos especializados más utilizados. Además se pudo observar que en ámbito hay una gran dependencia denominativa del inglés. Los resultados también nos mostraron que en los textos dirigidos a pacientes y familiares, las unidades terminológicas nuevas no se introducen como nuevas, sino que se utilizan como si ya fueran conocidas por el usuario de estos textos. Esta es una interferencia lingüística adicional que impide una asimilación completa y satisfactoria de la información que se pretende dar en estos textos.

KEYWORDS

Neology, specialized neology, terminology, rare disease, health literacy, medical terminology, Neurofibromatosis, Turner Syndrome, WAGR Syndrome, specialized communication.

ÍNDICE

1. INTRODUCCIÓN.....	4
2. OBJETIVOS E HIPÓTESIS	7
3. LA NEOLOGÍA EN LA LENGUA GENERAL Y LA NEOLOGÍA ESPECIALIZADA	9
4. METODOLOGÍA Y RESULTADOS	13
4.1. Elaboración de los corpus de trabajo	14
4.1.1. Selección de los textos para construir los corpus	15
4.1.1.1. Corpus pacientes y familiares (CPAC).....	16
4.1.1.2. Corpus especialistas (CESP)	18
4.2. Procesamiento de los corpus con WikiYATE	22
4.2.1. Conversión de los archivos PDF a texto plano	23
4.2.2. Procesamiento lingüístico de los documentos.....	24
4.2.3. Extracción de los patrones terminológicos.....	25
4.2.4. Análisis con la Wikipedia	26
4.3. Comparación de los resultados de WikiYATE.....	28
4.3.1. Revisión con WinMerge y extracción de unidades nuevas.....	30
4.4. Validación de los candidatos a neologismos especializados	32
4.4.1. Aspectos tenidos en cuenta para la validación	33
4.4.2. Depuración de los candidatos.....	34
4.5. Análisis de las unidades nuevas validadas.....	36

5. TIPOLOGÍA DE LOS NEOLOGISMOS ESPECIALIZADOS	37
5.1. Neologismos especializados CESP.....	38
5.2. Neologismos especializados CPAC.....	39
6. DIFERENCIAS ENTRE EL CESP Y EL CPAC.....	40
6.1. Densidad neologismos especializados	40
6.2. Tipos más comunes de neologismos.....	42
7. EL INGLÉS COMO MEDIO DENOMINATIVO.....	43
8. COTEXTOS DE LOS NEOLOGISMOS.....	45
8.1. Cotextos en el CESP	45
8.2. Cotextos en el CPAC	46
9. CONCLUSIONES Y TRABAJO FUTURO.....	49
10. REFERENCIAS	52
ANEXOS	56

1. INTRODUCCIÓN

Una *enfermedad rara*, también llamada *enfermedad huérfana* o *minoritaria*, es cualquier enfermedad que afecta a un pequeño porcentaje de la población. La mayoría de estas enfermedades son genéticas y están presentes durante toda la vida de una persona, incluso si los síntomas no aparecen inmediatamente. Así pues, las enfermedades raras son poco frecuentes y en su mayoría crónicas y complejas. En Europa se considera que una enfermedad o desorden es raro cuando afecta a 1 de cada 2.000 ciudadanos (EURORDIS, 2009). Estas enfermedades en conjunto afectan a un grupo importante de personas, aproximadamente entre el 6% y el 8% de la población mundial (Aguirre, Angulo, Guerrero, Motero, & Prieto, 2012).

No obstante, las enfermedades raras no solo afectan a las personas diagnosticadas, sino también a sus familias, amigos, cuidadores y a toda la sociedad. Una gran parte de estas enfermedades son genéticas y se presentan en la edad pediátrica, esto es algo que hace que la comunicación se dé principalmente entre profesionales sanitarios y familias o cuidadores. Aunque no sólo se presentan en esta edad, estas enfermedades raras también surgen más allá de la infancia y la adolescencia. Además, los pacientes afectados por estas enfermedades requieren una atención especial en la que un conjunto de personas son responsables del tratamiento global, el pediatra o el médico de familia en atención primaria, el especialista o los especialistas que entienden de los problemas clínicos específicos del enfermo, la atención de enfermería y fisioterápica, y los servicios sociales y el apoyo psicológico (Palau, 2010). Como para los pacientes afectados los problemas asociados a la enfermedad aparecen, generalmente, a una temprana edad, es necesario que la acción

sanitaria, social, escolar y laboral de todas estas personas involucradas se realice por un largo período de tiempo y de manera planificada.

Por otra parte, en una sociedad en la que los pacientes y sus familias han comenzado a participar activamente en los procesos médicos que los afectan, la asimilación de la información escrita recibida de parte de los profesionales es uno de los elementos clave para que la comprensión de temas relacionados con la salud sea una realidad. Es un hecho que las interferencias lingüísticas y cognitivas de los textos generados por profesionales y asociaciones destinados a pacientes y familiares impiden una asimilación completa y satisfactoria de esta información. En el ámbito de las enfermedades raras estas interferencias a menudo están presentes en la comunicación escrita que se da entre profesionales y pacientes y familiares. En los textos que se producen en la comunicación escrita, muchas veces, la densidad terminológica, las unidades especializadas nuevas, la enorme cantidad de siglas, la variación terminológica, todas características comunes de los textos médicos, frenan la comunicación entre profesionales y pacientes.

Además, esta comunicación suele tener consecuencias que repercuten en la salud del paciente. A veces son consecuencias positivas (el paciente se puede sentir mucho mejor después de hablar con el especialista); y otras, en cambio, son efectos contraproducentes (el paciente puede no entender el mensaje, puede interpretarlo de manera equivocada, lo que activa su imaginario con consecuencias indeseables, o incluso puede entender lo contrario de lo que se deseaba comunicar) (Estopà, 2012).

De igual manera, dentro de esta comunicación especializada entre profesionales y pacientes y familiares las unidades terminológicas o términos son un factor clave, y además, ayudan a

la difusión del conocimiento, como ya afirma Cabré (1993:103), *los términos reflejan la estructuración conceptual de una disciplina, y por ello son la base de la comunicación especializada.*

Al hablar de *estructuración conceptual* es necesario hablar también de necesidades denominativas, es decir, la aparición de un nuevo concepto conlleva normalmente la aparición de una nueva denominación. En terminología pueden presentarse dos tipos de situaciones que requieren una determinada intervención denominativa: la primera, cuando en un concepto confluyen dos o más denominaciones y ese fenómeno entorpece la comunicación; la segunda, cuando una lengua de especialidad no posee la denominación necesaria para expresar un concepto. En el primer caso, para garantizar una comunicación precisa puede ser necesario simplificar las distintas formas designativas y reducirla a una sola; en el segundo caso, debe crearse una nueva forma de denominación para expresar el nuevo concepto (Cabré, 1993), es decir un neologismo especializado.

Con respecto a la neología, y la neología especializada no es la excepción, Rey (1976) afirma que es un proceso dinámico que produce unidades léxicas nuevas y unidades terminológicas nuevas, bien por la necesidad de transmitir las novedades de un mundo en evolución constante, bien para designar conceptos que ya existían, por razones difíciles de clasificar y que pueden estar ligadas a fenómenos totalmente subjetivos y colectivos como la necesidad de expresarse de una manera nueva o el esnobismo de renovar el inventario léxico de una lengua. No se puede negar que en la actualidad hay una gran inversión económica en investigación científica dentro campos como la medicina, lo que quiere decir que con el avance investigativo la necesidad denominativa de nuevos conceptos continuará estando presente.

Teniendo en cuenta todos los aspectos previamente mencionados, este trabajo tendrá como punto de partida el análisis de las unidades terminológicas nuevas utilizadas en español en el ámbito de las enfermedades raras, específicamente en textos dirigidos a profesionales sanitarios y en textos dirigidos a pacientes y familiares. En estos contextos comunicativos entre familias de pacientes y profesionales sanitarios siempre son las familias las que se encuentran en desventaja.

Este trabajo se desarrollará en el ámbito de las enfermedades raras y es por esto que se contará con el apoyo del proyecto RecerCaixa JUNTOS, un proyecto de investigación, dirigido por la Dra. Rosa Estopà (UPF) y el Dr. Manuel Armayones (UOC), cuyo objetivo principal es analizar la información escrita que reciben las familias que tienen un niño afectado por una enfermedad rara, con el fin de estructurar protocolos para adecuar la información especializada a las necesidades y competencias comunicativas de estas familias (Estopà & Armayones, 2015).

2. OBJETIVOS E HIPÓTESIS

El objetivo principal de este trabajo es realizar una aproximación al uso que se le da a la neología especializada del ámbito de las enfermedades raras, en textos dirigidos a profesionales sanitarios y en textos dirigidos a pacientes y familiares. Este objetivo se concreta en los objetivos secundarios siguientes:

- a. Elaborar dos corpus textuales a manera de muestra: un corpus con documentos dirigidos a especialistas y otro con documentos escritos dirigidos a pacientes y familiares, ambos sobre el mismo tema (tres enfermedades raras diferentes).

- b. Detectar y clasificar tipológicamente los neologismos especializados utilizados en dos corpus compuestos por textos del ámbito de las enfermedades raras dirigidos a usuarios distintos, uno a profesionales sanitarios y el otro a pacientes y familiares.
- c. Hacer un análisis contrastivo de los neologismos especializados encontrados en ambos corpus.
- d. Explorar hasta qué punto hay dependencia denominativa en español del inglés en el ámbito de las enfermedades raras.
- e. Identificar los contextos en los que ocurren las unidades terminológicas nuevas y hasta qué punto se las introduce como nuevas o no.

La finalidad primera de este trabajo es identificar y analizar la neología especializada usada en dos corpus textuales de niveles de especialidad distintos del ámbito de las enfermedades raras, destinados a usuarios distintos. Es decir, se pretende hacer un análisis de todos los neologismos especializados, organizados por tipología, extraídos de ambos corpus textuales. De igual manera, se desea contrastar estos neologismos encontrados en ambos corpus, con el fin de observar si hay algún tipo de variación en cuanto a los términos utilizados y analizar la manera como se introducen en el texto, es decir el cotexto. Establecer también cuál es el grado de densidad de neologismos especializados; y cuáles son las tipologías de neologismos más frecuentes en ambos corpus. Con los resultados que se obtengan de este trabajo se espera continuar con más investigaciones acerca de cómo todos estos factores pueden estar afectando la comprensión de este mismo tipo de textos.

Para realizar el trabajo partimos de las hipótesis siguientes:

- a. Los textos dirigidos a profesionales sanitarios contienen un número mayor de neologismos especializados que los textos destinados a pacientes y familiares.
- b. El ámbito de la medicina tiene una alta dependencia denominativa del inglés, lo cual se podría evidenciar con una presencia alta de neologismos especializados del tipo *préstamo* o *calco*.
- c. En los textos dirigidos a pacientes y familiares el cotexto que acompaña a las unidades terminológicas nuevas las presenta como neologismos, mientras que en los textos dirigidos a profesionales no se intenta dar una introducción o un contexto que indique que las unidades terminológicas nuevas, efectivamente lo son.

3. LA NEOLOGÍA EN LA LENGUA GENERAL Y LA NEOLOGÍA ESPECIALIZADA

La capacidad de crear palabras nuevas, o creatividad léxica, es un hecho inherente a la especie humana, ya sea para adaptarse a las necesidades denominativas de la realidad o simplemente referirse a las cosas, los pensamientos y sentimientos o a los seres humanos de una manera diferente a la habitual, simplemente por el hecho de llamar la atención o autoafirmarse. La competencia léxica, dentro de la competencia lingüística en general, presupone una serie de conocimientos inherentes al hablante por el simple hecho de ser un ser hablante. Son conocimientos que están ligados a la especie humana (Cabré, 2015). En términos generales, la neología es la materia que se ocupa de todos estos aspectos relativos a los fenómenos nuevos que aparecen en las lenguas. Tradicionalmente la neología ha estudiado los fenómenos lingüísticos que surgen en un momento determinado de la evolución de una lengua.

Se habla específicamente de neología léxica para hacer referencia a la aparición de palabras nuevas dentro de una lengua. Es decir, la neología comprende toda unidad que pueda constituir una entrada de diccionario, ya sea una unidad simple o una unidad sintagmática. Estopà (2010: 17) afirma que *L'entrada de neologismes en les llengües vives és un fet natural: tota llengua activa actualitza permanentment el seu lèxic. Els neologismes són la conseqüència lingüística dels canvis constants d'una societat.*

Pero para delimitar el concepto de *neologismo*, es necesario destacar la obra del francés Alain Rey (1976), para quien la neología es un proceso dinámico que produce unidades léxicas nuevas y también unidades terminológicas nuevas, bien sea por la necesidad de transmitir las novedades de un mundo en evolución constante, bien para designar conceptos que ya existían, por razones difíciles de clasificar y que pueden estar ligadas a fenómenos totalmente subjetivos y colectivos como la necesidad de expresarse de una manera nueva o el esnobismo de renovar el inventario léxico de una lengua. Desde esta definición podemos concluir que existen unidades terminológicas nuevas en contextos comunicativos especializados, los neologismos especializados, los cuales surgen en su gran mayoría por necesidades denominativas y suelen tener una estabilidad más duradera que los neologismos léxicos de la lengua general.

Sin embargo, las unidades nuevas absolutamente para todo el mundo, en todos los contextos y situaciones, son muy raras. La neología tiene un carácter temporal (lo que en un momento determinado es un neologismo deja de serlo) y uno subjetivo (lo que es un neologismo para alguien, en algún lugar o en un determinado contexto de uso puede no serlo para otro hablante, ámbito geográfico o contexto de comunicación) inherentes (Cabré, 2015).

A pesar de que debido a esto es difícil abordar un objeto relativo con una metodología científica, Cabré (2008: 9-36, 2015: 125-136), siguiendo algunos de los principios propios de la teoría comunicativa de la terminología (TCT), desarrolla una teoría de los neologismos, en la cual se plantean una serie de criterios que permitirían abordar los neologismos sin perder de vista su complejidad. Para este trabajo vamos a tener en cuenta estos mismos principios:

- El primero es el *principio de poliedricidad*, según el cual todo objeto de conocimiento es multidisciplinar, en cuanto a que se puede abordar desde materias y campos del conocimiento diferentes, y multidimensional, en cuanto hay que tiene diferentes aspectos que no se pueden describir simultáneamente.
- El *principio de descripción situada*, el cual quiere dar razón del hecho de que una descripción de los neologismos requiere que el analista se sitúe en una perspectiva determinada dentro de la cual sea capaz de dar cuenta de una faceta determinada de este objeto, sin contradecirse con la descripción que se lleva a cabo desde las otras perspectivas.
- Esta condición de articulación entre las diversas descripciones es el *principio de no contradicción*.
- El *principio de descripción orientada*, que se basa en el criterio de multifuncionalidad de los objetos de conocimiento, pretende dar razón del hecho de que cada perspectiva de descripción persigue unos objetivos diferentes y que se debe explicitar en función de qué metas se lleva a cabo.
- El *Principio de explicitación*, que obliga a precisar qué aspectos de los neologismos se describen.

- *El principio de focalización* sobre el objeto se propone asegurar que en una descripción se explicita qué parte del objeto se considera la más relevante.
- *El principio del punto de vista* o *principio de la perspectiva* quiere asegurar que, dada la naturaleza enunciativa de los neologismos, se explicita en cada descripción desde qué vertiente del proceso enunciativo actuamos: desde la producción de neologismos, desde la recepción o desde de la transmisión.
- Finalmente, para articular esta larga lista de principios, es necesario un principio integrador, *el principio de interrelación entre situación, dimensión, focalización, orientación y perspectiva*, que estructure y articule adecuadamente y jerárquicamente este conjunto de principios.

Tener en cuenta todos estos principios nos permitirá hacer un análisis más completo del comportamiento que tiene la neología especializada en el ámbito de las enfermedades raras.

En cuanto a las características de los *neologismos especializados*, estas no se diferencian en ningún aspecto de las de los términos no neológicos. Su creación está justificada por la necesidad de denominar un concepto; suelen ser más descriptivos que las palabras; son más estables en cuanto a perdurabilidad, y se valen de series fijas de derivativos dentro de cada campo de especialidad. Desde el punto de vista de la forma, los neologismos terminológicos utilizan los mismos recursos que los términos no neológicos (Cabré, 1993).

Finalmente, se deben tener en cuenta diferentes criterios para establecer si un término es nuevo o no. Rey (1976) identifica tres criterios que permiten establecer conjuntos de palabras nuevas: temporal, psicolingüístico y lexicográfico. En nuestro caso, los parámetros que más se van a tener en cuenta para establecer si una unidad es un neologismo

especializado o no, son el lexicográfico y el temporal. En cuanto a la clasificación de los estos neologismos, es necesario tener en cuenta que existen diferentes maneras de producir nuevas unidades terminológicas y también existen diferentes tipos de unidades que se pueden crear en una lengua determinada.

De acuerdo con Estopà (2011), en las lenguas romances existen básicamente dos maneras de producir neologismos: con los recursos propios del idioma (recursos fonológicos, morfológicos, semánticos y sintácticos) o con préstamos de otro idioma. Dentro de los recursos propios del idioma, uno de los recursos más comúnmente utilizados para la creación de nuevas unidades terminológicas es la sintagmación (Simón, Fontáns, Duarte, & Barité, 2013). Además en el ámbito de la medicina el porcentaje de unidades terminológicas poliléxicas (UTP) es muy alto, según Estopà (1999) el 52% de las unidades terminológicas son poliléxicas.

4. METODOLOGÍA Y RESULTADOS

La metodología a seguir para este trabajo se divide en las etapas siguientes:

- elaboración de los dos corpus textuales en español con textos publicados entre 2010 y 2015;
- implementación de la herramienta WikiYATE¹ en ambos corpus dos veces, una primera vez teniendo como punto de partida la versión 2008 de la Wikipedia y una segunda vez a partir de la versión de finales de 2014;

¹ Extractor automático de terminología. Herramienta de extracción automática de candidatos a término utilizando la Wikipedia. Fue creado por Jorge Vivaldi del grupo IULATERM.

- comparación entre las unidades potencialmente terminológicas resultantes los dos análisis hechos a cada uno de los corpus para encontrar las unidades nuevas (punto de referencia temporal);
- validación de las unidades resultantes candidatas a neologismos especializados;
- análisis de los neologismos especializados utilizados en el ámbito de las enfermedades raras (específicamente en las tres seleccionadas);
- y resultados preliminares.

4.1. Elaboración de los corpus de trabajo

La condición de relatividad de los neologismos es uno de los obstáculos que dificultan la identificación de los mismos. Los neologismos son relativos, por lo que su condición de “novedad” no es absoluta, sino que es necesario tener siempre un punto de referencia para poder clasificarlos como tal. Este punto de referencia puede ser temporal, social, geográfico, funcional o temático (Cabré, 2015). Para este trabajo tendremos en cuenta el punto de referencia temporal y por lo tanto utilizaremos dos corpus con textos publicados entre el 2009 y el 2015, los cuales nos van a ayudar a detectar más fácilmente los neologismos de una manera semiautomática con la ayuda de un corpus de un corpus de contraste, que en este caso será una versión de la Wikipedia almacenada en el servidor local del grupo IULATERM², puesto que aún es necesario que una persona sea la encargada de validar los candidatos a neologismos especializados que resulten.

² <https://www.iula.upf.edu/iulaterm/tpreses.htm>

4.1.1. Selección de los textos para construir los corpus

Como el objetivo de este trabajo es hacer una primera exploración de la presencia de neología especializada en el ámbito de las enfermedades raras, se elaboró un corpus con tres textos dirigidos a pacientes y familiares (CPAC) y otro corpus con seis textos dirigidos a especialistas (CESP). El número de textos para el segundo corpus es el doble, puesto que el tipo de documentos disponibles para especialistas (artículos de revisión, artículos de genética práctica, guías de práctica, protocolos, etc.) tienen una tendencia a ser menos extensos y más concretos que los documentos disponibles para las familias y pacientes (guías de enfermedades para familias, entre otros). Entonces, en el momento de la selección, se tuvo en cuenta el número total de palabras de todos los textos para que el tamaño de los dos corpus fuera aproximadamente el mismo, ambos corpus cuentan con un total aproximado de 30.000 palabras.

Para que el análisis arrojara unos resultados que pudieran reflejar de mejor manera lo que realmente ocurre en cuanto a neología especializada en este tipo de textos, se hizo una selección de documentación sobre el mismo tema para ambos corpus, es decir, se buscaron textos sobre tres enfermedades raras, tanto para el CESP como para el CPAC, las enfermedades escogidas al azar fueron las siguientes: el *síndrome de Turner*, el *síndrome de WAGR* y la *neurofibromatosis*. La búsqueda se hizo en una ventana temporal de 6 años, de 2010 a 2015.

Lo que primero se hizo fue un sondeo en de internet con la ayuda de diferentes motores de búsqueda y bases de datos (*Google Scholar*, *Orphanet*³, *FEDER*⁴, *EURORDIS*⁵, *Guies*

³ Portal de información de enfermedades raras y medicamentos huérfanos *Orphanet*. <http://goo.gl/voGNY7>

BibTIC UPF⁶) con el objetivo de establecer cuáles eran las mejores fuentes documentales para conseguir los textos que tuvieran las mejores características para construir los dos corpus. El principal criterio de búsqueda, además del temporal, era que los textos estuvieran disponibles en línea en español. También se tuvo en cuenta que la documentación de una enfermedad específica estuviera disponible en español para ambos corpus.

4.1.1.1. *Corpus pacientes y familiares (CPAC)*

Después de una búsqueda exhaustiva, los tres textos para construir el CPAC se seleccionaron de tres fuentes diferentes. La búsqueda se dificultó un poco puesto que la información disponible en español, en línea, no es mucha y la mayor parte es anterior al 2010, es decir que es información que se descarta no corresponde a la ventana temporal establecida para este trabajo. Sin embargo, fue posible encontrar algunos documentos que servían para este trabajo y se seleccionaron los siguientes: *Preguntas y respuestas sobre el Síndrome de Turner* (Turner_PAC__1); *Neurofibromatosis 1, NF-1* (Neuro_PAC_1); y el tercero se llama: *Aniridia, tumor de Wilms y síndrome WAGR* (WAGR_PAC_1). A cada uno de los documentos se les asignó un nombre más corto, y estándar al corpus, que permite hacer un trabajo de análisis más organizado.

Se seleccionaron estos textos porque son textos escritos específicamente para ayudar a los pacientes afectados por estas enfermedades raras, y sus familias. La descripción de cada uno de ellos se puede ver a continuación:

⁴ Federación Española de Enfermedades Raras (FEDER). <http://goo.gl/iGsPg>

⁵ EURORDIS es una alianza no-gubernamental dirigida por organizaciones de pacientes y personas individuales activas en el campo de las enfermedades raras. <http://goo.gl/8HMsw>

⁶ Ciències Experimentals i de la Salut: Bases de dades. <http://goo.gl/Rk2nZ3>

1. Turner_PAC_1: esta es la segunda edición del libro editado en 1997. El libro es una guía para pacientes afectadas por el Síndrome de Turner y sus familias. En el libro se incluye información acerca de la enfermedad, la genética de la misma, el apoyo social y emocional a las pacientes, el paso de la infancia a la edad adulta, entre otras cosas. El coordinador del libro es Juan Pedro López Sigüero, del Servicio de Endocrinología Pediátrica Hospital Materno-Infantil de Málaga. El libro fue publicado en 2011 por *Novo Nordisk Pharma* (López Sigüero, 2011).

2. Neuro_PAC_1: esta es una guía informativa sobre la neurofibromatosis 1 dirigida a pacientes y familias, disponible en el sitio web de Infogen desde 2014. La guía incluye información acerca de lo que es la neurofibromatosis, cómo se clasifica, sus síntomas, su diagnóstico y su tratamiento. Infogen es una asociación civil no lucrativa, fundada en 1997, en México. Está integrada por un grupo de voluntarios y profesionales de la salud y surge como respuesta a la necesidad de información que existe en torno a los defectos que se presentan en el nacimiento (Infogen, 2014).

3. WAGR_PAC_1: este documento es una hoja informativa para pacientes afectados por el síndrome de WAGR, hace parte de una guía para la práctica clínica que también incluye el protocolo diagnóstico, el protocolo de seguimiento y la información mínima necesaria acerca de un paciente afectado. Los autores de este texto son Isabel Lorda-Sanchez, Eddy René González Flores y Carmen Ayuso, del Grupo de Trabajo sobre Cáncer en Síndromes Genéticos Polimalformativos (GT-CSGP). Fue publicado por el Servicio de Genética, Capiro, Fundación Jiménez Díaz, en Madrid; y el centro de Investigación Biomédica en Red de Enfermedades Raras, CIBERER (Lorda-Sanchez, González Flores, & Ayuso, 2015).

El resumen de las características de los textos que componen el corpus CPAC se presenta en la tabla siguiente:

Nombre	Año	Número de palabras
Turner_PAC_1	2011	23.383
Neuro_PAC_1	2014	3.857
WAGR_PAC_1	2015	1.675
Total		28.915

Tabla 1. Descripción del CPAC

4.1.1.2. *Corpus especialistas (CESP)*

Después de seleccionar los tres textos para el CPAC, se empezó a hacer la búsqueda de textos en español dirigidos a especialistas sobre las mismas enfermedades que se seleccionaron para el CPAC. En total se escogieron dos textos de cada una de las enfermedades seleccionadas para, de alguna manera, equiparar el tamaño de ambos corpus. Entre los textos seleccionados hay artículos de revisión, los cuales son artículos extensos que resumen el conocimiento médico actual de la enfermedad; artículos de genética práctica, es decir, artículos de revisión focalizados en el conocimiento genético y en la batería de pruebas necesarias para llegar al diagnóstico del paciente; y guías para la práctica clínica, que son resúmenes de recomendaciones para el manejo de los afectados, elaborados por organismos oficiales.

El primer texto seleccionado acerca del síndrome de Turner para el CESP es *Actualización en el tratamiento y seguimiento del síndrome de Turner (ST)* (Turner_ESP_1); el segundo

se titula *Reemplazo estrogénico en pacientes con síndrome de Turner* (Turner_ESP_2). Para la neurofibromatosis se seleccionaron: *Características fenotípicas y genéticas en la neurofibromatosis tipo 1 en edad pediátrica* (Neuro_ESP_1) y *Neurofibromatosis tipo 1 (NF1) y su diagnóstico molecular como estrategia del diagnóstico diferencial y a edades tempranas* (Neuro_ESP_2). Finalmente, para el Síndrome WAGR se seleccionaron los dos textos siguientes: *Tumor de Wilms unilateral asociado a aniridia: a propósito de un caso* (WAGR_ESP_1) y *Aniridia, Tumor de Wilms y Síndrome WAGR* (WAGR_ESP_2). La descripción de cada uno de estos textos sigue a continuación:

1. Turner_ESP_1: este es un artículo de revisión dirigido a especialistas publicado en 2010 en la Revista Española de Endocrinología Pediátrica. Los autores son José I. Labarta, Antonio de Arriba, Esteban Mayayo, Ángel Ferrández-Longás, de la Unidad de Endocrinología del Hospital Infantil Universitario Miguel Servet, Zaragoza (Labarta, de Arriba, Mayayo, & Ferrández-Longás, 2010).

2. Turner_ESP_2: este es un artículo de revisión dirigido a especialistas publicado en 2015 en Perinatología y Reproducción Humana. Los autores son Pérez Bueno, F. Aguilera Pérez, J.R. Garza Ríos, P. del Instituto Nacional de Perinatología, México D.F., México. Este artículo revisa la literatura médica sobre terapia hormonal de reemplazo en pacientes con síndrome de Turner (Pérez Bueno, Aguilera Pérez, & Garza Ríos, 2015).

2. Neuro_ESP_1: este es un artículo de genética práctica publicado en los Anales de Pediatría, la Publicación Oficial de la Asociación Española de Pediatría. Es un artículo producido por diferentes personas: A. Duat Rodríguez, de la Sección de Neurología del Hospital Infantil Universitario Niño Jesús, Madrid, España; G.Á. Martos Moreno, del

Servicio de Pediatría y Endocrinología del Hospital Infantil Universitario Niño Jesús, Instituto de Investigación Sanitaria la Princesa, España; Y. Martín Santo-Domingo, del Servicio de Genética Molecular, Hospital Universitario Ramón y Cajal, Madrid, España; A. Hernández Martín, de la Sección de Dermatología, Hospital Infantil Universitario Niño Jesús, Madrid, España; J.M. Espejo-Saavedra Roca, del Servicio de Psicología Clínica, Hospital Universitario 12 de Octubre, Madrid, España; M.L. Ruiz-Falcó Rojas también del Servicio de Pediatría y Endocrinología del Hospital Infantil Universitario Niño Jesús; y por J. Argente, del Departamento de Pediatría, Universidad autónoma de Madrid (Duat Rodríguez et al., 2015).

3. Neuro_ESP_2: este es un artículo de revisión publicado en la Revista médica de Chile, número 143 de 2015, por Martha Gómez, del Centro Especializado de Genética (CEGEN), Facultad de Ciencias Naturales y Exactas, Universidad Autónoma de Chiriquí (UNACHI), ciudad de David, provincia de Chiriquí, República de Panamá; y por Oriana Batista, también de CEGEN y del Centro Gendiagnostik, S.A., ciudad de David, provincia de Chiriquí, República de Panamá. Es un artículo financiado por el Centro Gendiagnostik, el Sistema Nacional de Investigación (SNI), la Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT/ APY-I10-002 A) y la Vicerrectoría de Investigación y Posgrado de la UNACHI (VIP 184-CN-01-A092-022011) (Gómez & Batista, 2015).

4. WAGR_ESP_1: este documento es un caso clínico relacionado con el estudio de un paciente afectado con tumor de Wilms. El objetivo radica en mostrar las características clínicas que presenta un paciente con tumor de Wilms unilateral asociado a aniridia, presente en el Síndrome WAGR. Los autores son Mayra Victoria-Rocha Choque¹, Auxiliar del Departamento de Educación Médica y Planificación (DEMyP); Alejandro Méndez

Pardo, Estudiante de Medicina, Facultad de Medicina de la Universidad Mayor de San Simón. Cochabamba, Bolivia; Claudia Terrazas Saavedra; y Antonio José Pardo Novak, M.D. Médico Ginecólogo-Obstetra. Jefe de Guardia, Hospital Materno Infantil German Urquidi. El artículo fue publicado en 2014 en la Gaceta Médica Boliviana (Rocha Choque, Méndez Pardo, Terrazas Saavedra, & Pardo Novak, 2014).

5. WAGR_ESP_2: esta es una guía para la práctica clínica que incluye el protocolo diagnóstico, el protocolo de seguimiento y la información mínima necesaria acerca de un paciente con Síndrome WAGR. Los autores de este texto son Isabel Lorda-Sanchez, Eddy René González Flores y Carmen Ayuso, del Grupo de Trabajo sobre Cáncer en Síndromes Genéticos Polimalformativos (GT-CSGP). Fue publicado por el Servicio de Genética, Capiro, Fundación Jiménez Díaz, en Madrid; y el centro de Investigación Biomédica en Red de Enfermedades Raras, CIBERER (Lorda-Sanchez et al., 2015).

El resumen de las características de los textos que componen el corpus CESP se presenta en la tabla siguiente:

Nombre	Año	Número de palabras
Turner_ESP_1	2010	4.685
Turner_ESP_2	2011	5.816
Neuro_ESP_1	2015	6.088
Neuro_ESP_2	2015	6.219
WAGR_ESP_1	2014	2.546
WAGR_ESP_2	2015	5.300
Total		30.654

Tabla 2. Descripción del CESP

De acuerdo con la información de las tablas 1 y 2, el tamaño de los corpus es aproximadamente el mismo lo cual nos facilitó la comparación: El CPAC tiene 28.915 palabras, mientras que el CESP tiene 30.654.

4.2. Procesamiento de los corpus con WikiYATE

Los textos que componen ambos corpus se procesaron con el extractor de terminología WikiYATE, el cual es un extractor de candidatos a término a partir de la *Wikipedia*. Ha sido desarrollado por Jorge Vivaldi en el ámbito del grupo de investigación IULATERM de la Universitat Pompeu Fabra en Barcelona. Los primeros experimentos de Jorge Vivaldi con la Wikipedia como extractor terminológico empezaron en 2010 y continúan hasta el día de hoy (Vivaldi, Cabrera-Diego, Sierra, & Pozzi, 2012; Vivaldi & Rodríguez, 2010). Las características más relevantes de esta herramienta son: a) uso de combinación de estrategias y b) uso intensivo de información semántica. El proceso de la herramienta se resume en la figura 1.

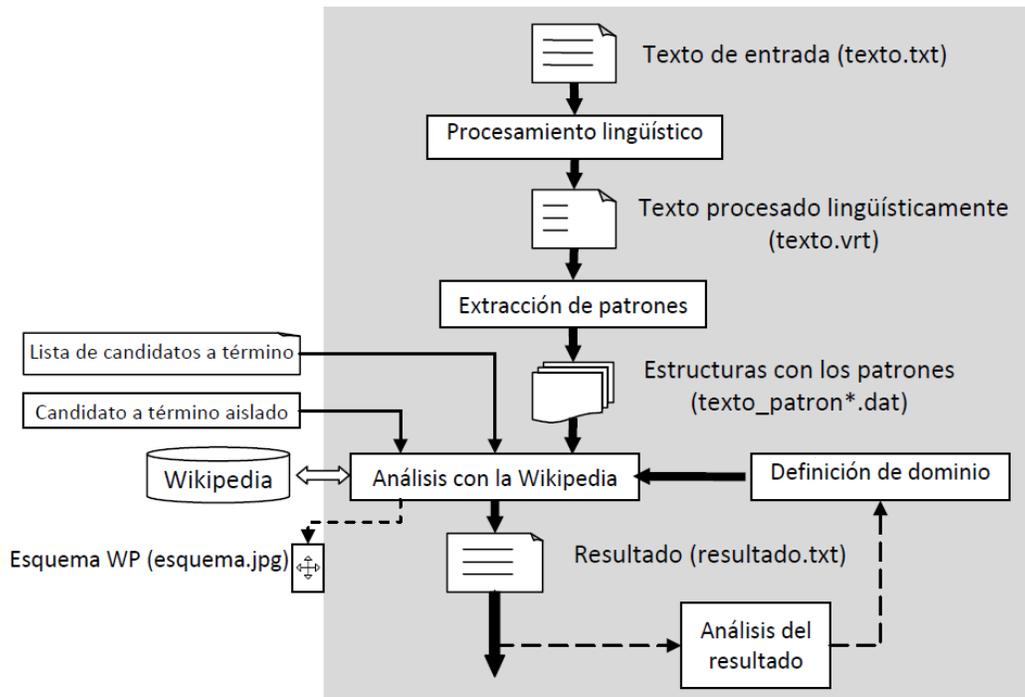


Figura 1. Esquema del procesamiento

Para hacer el procesamiento de cada uno de los textos con esta herramienta, fue necesario seguir los pasos siguientes:

4.2.1. Conversión de los archivos PDF a texto plano

Después de tener los textos seleccionados, fue necesario convertirlos de su formato original (todos ellos en PDF) a texto plano con codificación iso-8859-i o UTF-8 (cualquiera de los dos formatos de codificación que permite que todos los caracteres de los textos en español sean reconocidos, sin embargo al ejecutar algunos comando en el proceso de WikiYATE es necesario seleccionar uno de los dos, dependiendo de cuál sea el proceso que se vaya a

llevar a cabo), ya que era necesario tenerlos así para poder hacer el análisis de los textos. Esta tarea se realizó con la ayuda un programa de reconocimiento óptico de caracteres⁷.

Los ficheros de texto obtenidos después de la conversión contenían algunos errores (palabras enganchadas, errores de reconocimiento, formato dañado por tablas, etc.), por lo que fue necesario hacer una revisión de los mismos con el procesador de texto MS WORD. Para la detección rápida de estos problemas, se utilizó la herramienta de corrección de ortografía y gramática de MS Word. Para el resto de correcciones y organización de formato, se utilizó el editor de texto EditPlus, puesto que es una herramienta que permite utilizar expresiones regulares para llevar a cabo sustituciones con el fin de organizar el formato de los textos. Esta revisión general permitió organizar el resultado obtenido después de la conversión de los textos y de esta manera pudimos asegurarnos de que toda la información inicial estuviera disponible de la mejor manera posible para el procesamiento con WikiYATE.

4.2.2. Procesamiento lingüístico de los documentos

Este procesamiento permite detectar problemas que hayan quedado después del proceso anterior. De igual manera, este procesamiento es necesario para poder obtener los patrones terminológicos de un texto. Para utilizar las herramientas de WikiYATE disponibles en este procesamiento, es necesario tener acceso a un directorio local del grupo IULATERM en el que están todos los archivos que se necesitan para poder ejecutar una orden en el intérprete de comandos, de Windows en este caso, (símbolo de sistema o cmd) que nos ayuda a procesarlos lingüísticamente. El comando que se ejecuta es el siguiente:

⁷ ABBYY FineReader 12

```
perl P:\IULA\CORPUS\UTILS\Preproceso\hector.pl -lang es -inputmimetype iso-8859-1 -
outputmimetype iso-8859-1 -inputtext nombrefichero.txt -outputtext nombrefichero.vrt -
annotationformat IULACT
```

Los parámetros se describen en la tabla siguiente:

Parámetro	Descripción
-lang	Idioma del texto a analizar (es nuestro caso español <i>es</i>)
-inputmimetype	Parámetro necesario para tener la codificación del fichero de entrada definida (en nuestro caso iso-8859-1)
-outputmimetype	Parámetro necesario para tener la codificación del fichero de salida definida (en nuestro caso iso-8859-1)
-inputtext	Texto a procesar. Nombre del fichero con el texto plano a analizar
-outputtext	Nombre del fichero resultado (será texto verticalizado con extensión “vrt”)
- annotationFormat	Formato de la anotación (en nuestro caso utilizamos el formato del corpus técnico del IULA, IULACT)

Tabla 3. Descripción parámetros procesamiento lingüístico

4.2.3. Extracción de los patrones terminológicos

La extracción de los patrones terminológicos se hace con el programa “filtroCatMCR.pl”, disponible en el directorio interno del IULA, P:\IULA\CORPUS\UTILS\YatePlus. Esta

acción consiste en leer el resultado del procesamiento lingüístico previo y extraer las secuencias potencialmente terminológicas por medio del comando siguiente:

```
perl P:\IULA\CORPUS\UTILS\YatePlus\filtroCatMCR.pl -lang es -i nombrefichero.vrt -o nombrefichero -DocNoCtBaseDir "directorio donde está el fichero"
```

Los parámetros se describen en la tabla 4 que se presenta a continuación:

Parámetro	Descripción
-lang	Idioma del texto a analizar (es nuestro caso español <i>es</i>)
-i	Nombre del fichero con extensión “vrt” a analizar
-o	Directorio en el cual se van a almacenar todos los ficheros resultantes del proceso
- DocNoCtBaseDir	Directorio en el que se encuentran los ficheros que no hacen parte del corpus técnico del IULA

Tabla 4. Descripción parámetros extracción patrones terminológicos

4.2.4. Análisis con la Wikipedia

Luego de extraer los patrones potencialmente terminológicos, otro programa, ubicado también en el directorio interno del IULA, se encarga de buscar en una copia (en un servidor local) de la Wikipedia los candidatos potencialmente terminológicos. Este programa se llama *exploraWPenAPyPP_5.pl* y hace una exploración del árbol de categorías de la Wikipedia. Se encuentra en *P:\IULA\CORPUS\UTILS\Wikipedia*. El comando que se ejecuta para llevar a cabo la tarea es el siguiente:

perl P:\IULA\CORPUS\UTILS\wikipedia\exploraWPenAPyPP_5.pl -lang es -doc

nombrefichero -indir directoriodelfichero -lmincat # -wp2008⁸

Parámetro	Descripción
-lang	Idioma del texto a analizar (es nuestro caso español <i>es</i>)
-doc	Este es el nombre que se le asignó a todos los ficheros .dat creados en el proceso anterior
-indir	Este es el directorio en el que están los ficheros creados en el paso anterior
-lmincat #	Esta es la longitud mínima de un candidato a término para ser analizado
-wp2008	Este parámetro se introduce para hacer el análisis con la versión de la Wikipedia de 2008 (por defecto se hace con la de 2014)

Tabla 5. Descripción parámetros análisis con la Wikipedia

Este último paso fue necesario realizarlo dos veces, una primera vez con la versión de la Wikipedia que se hace por defecto, es decir, la de 2014, y una segunda vez con la versión de 2008. La versión de octubre de 2014 cuenta con más de 2.534.720 páginas disponibles y 226.841 categorías diferentes, mientras que la versión de 2008 cuenta con 1.000.282 artículos y 80.482 categorías distintas.

⁸ Esta opción se utilizó cuando se hizo el análisis con la versión de 2008 de la Wikipedia.

4.3. Comparación de los resultados de WikiYATE

El resultado que hay al final del proceso con WikiYATE es un archivo de texto plano que se puede abrir utilizando una herramienta de hojas de cálculo como Microsoft Excel. En esta herramienta se pueden visualizar los resultados en una tabla que permite manejarlos y analizarlos más rápidamente. Como cada uno de los corpus fue procesado con WikiYATE dos veces, se obtuvieron dos listas de cada uno de ellos. Al abrir estas listas en Microsoft Excel cada una de las unidades potencialmente terminológicas tenía asociada una serie de informaciones de las cuales sólo tuvimos en cuenta las que se describen a continuación⁹:

Parámetro	Significado
CAT	Candidato a término (en nuestro caso a neologismo especializado)
Lema	Lema
Patrón	Patrón morfosintáctico (es el patrón utilizado en el Corpus Técnico del IULA)
Fichero ¹⁰	Parámetro interno de control con el nombre del fichero
CDwp_nc	Coefficiente de dominio en base al número de caminos ¹¹
CDwp_nc_w	Coefficiente de dominio (ponderado) al número de caminos

⁹ Sólo tuvimos en cuenta estos parámetros porque eran los relevantes para este trabajo, sin embargo la tabla completa se puede ver en los anexos.

¹⁰ Este es un parámetro que se agregó en este trabajo para tener mayor control sobre las unidades terminológicas.

¹¹ El número de caminos (relación) que hay entre un candidato a término y un dominio específico (por ej. medicina, derecho, economía, etc.). Este es el modo de exploración del árbol de categorías de la Wikipedia.

DminTopDom	Distancia mínima a un top del dominio
Detección ¹²	-page: existe como página de la Wikipedia -category: existe como categoría de la Wikipedia -components: obtenido mediante análisis de sus componentes -nil: no encontrado en esta versión de la Wikipedia

Tabla 6. Descripción parámetros resultantes

Por otro lado, lo que primero se hizo fue hacer una cuenta de las unidades que había en ambas listas. En la tabla que se muestra a continuación están los resultados de este proceso:

Corpus	Unidades potencialmente terminológicas	
	wp2008	wp2014
CESP	4.167	3.822
CPAC	3.499	3.438

Tabla 7. Resultados de análisis con WikiYATE

En la tabla 7 se muestran las unidades potencialmente terminológicas que resultaron después de ambos análisis. Se puede observar que a pesar de que la versión de 2014 tiene muchos más artículos, la cantidad de unidades potencialmente terminológicas es menor, esto se debe a que al tener más datos disponibles, WikiYATE puede hacer un mejor análisis con ayuda de los árboles de la Wikipedia en ambos corpus y por lo tanto da unos resultados más confiables. En decir, descarta algunas unidades que antes consideraba terminológicas y después no, ya que con la versión más actualizada tiene la información necesaria para decir que no lo son.

¹² Este parámetro fue modificado para este trabajo. Sólo tuvimos en cuenta los mecanismos de detección por página, categoría y componentes.

4.3.1. Revisión con WinMerge¹³ y extracción de unidades nuevas

Con la ayuda de Microsoft Excel se organizaron los resultados de cada lista en orden alfabético y se guardaron los ficheros de nuevo en formato de texto plano. Después con la ayuda de WinMerge se hizo una comparación de ambas listas para saber cuáles eran las unidades potencialmente terminológicas nuevas. WinMerge es una herramienta de diferenciación y combinación de código abierto para Windows, puede comparar tanto carpetas como archivos, presentando las diferencias en un formato de texto visual fácil de entender y controlar.

A pesar de que esta herramienta permite alinear los archivos y resalta automáticamente las diferencias entre los mismos, fue necesario hacer la revisión manualmente para saber cuáles eran las diferencias entre ambas listas y así determinar los candidatos a unidades nuevas. Estas unidades resultantes fueron nuestros candidatos a neologismos especializados (Hay que tener siempre en cuenta que WikiYATE es una herramienta de extracción de candidatos a término, pero para este trabajo la aprovechamos para extraer los candidatos a neologismo especializado).

Tomando como punto de partida los resultados de la lista resultante de la versión 2014 de la Wikipedia, se tuvieron en cuenta dos criterios para decidir que una unidad era nueva, a saber: que en el parámetro de detección de cada candidato en la lista resultante de la wp de 2008 fuera –nil y en la lista de 2014 cualquiera de las otras opciones, es decir –components, –category, page, etc.; o que los criterios de detección en ambas listas fueran –nil. No obstante, algunas unidades cumplían el segundo criterio pero no se etiquetaron como

¹³ <http://winmerge.org/about/?lang=es>

nuevas ya que eran unidades que presentaban algún tipo de error (/exón, 50y, 8315del, etc.), por lo que eran descartadas en el proceso de revisión. Después de haber hecho la revisión manual de todas las unidades, teniendo en cuenta los criterios previamente mencionados, se obtuvieron los resultados siguientes:

Unidades potencialmente terminológicas	CESP	CPAC
Descartadas	3.663	3.323
Nuevas	159	116
Total	3.822	3.439

Tabla 8. Comparación de resultados de WikiYATE

En total se etiquetaron 274 unidades como nuevas, 158 en el CESP y 116 en el CPAC. Inicialmente se descartaron los candidatos que no cumplieran con alguno de los dos criterios anteriores, es decir unidades que ya aparecían en la Wikipedia: *aniridia*, *anticonceptivos hormonales*, *biometría flemática*, *cáncer de mama*, etc. También se descartaron unidades que en principio cumplían con los criterios para ser etiquetados como nuevas, pero que después de una revisión más minuciosa de la lista se encontró que estaban repetidas y que en definitiva no se deberían tener en cuenta para el análisis. Esto pasó porque WikiYATE llevaba a cabo el análisis sobre una unidad pero si encontraba otra unidad idéntica, la descartaba. Esto causaba que hubiera muchos falsos candidatos, ya que muchas unidades potencialmente terminológicas se repetían a lo largo de los textos, algunos ejemplos son los siguientes:

##CAT	Patrón	CDwp_nc	CDwp_nc_w	DminTop Dom	Detección wp2008	Detección wp2014
ángulo camerular	NJ	0	-1	-1	components	components
ángulo camerular	NJ	-1	-1	-1	nil	nil
células nerviosas estriatales	NJJ	0.15	-1	-1	components	components

células nerviosas estriatales	NJJ	-1	-1	-1	nil	nil
cardiomiopatía hipertrófica obstructiva	NJJ	0.15	-1	-1	components	components
cardiomiopatía hipertrófica obstructiva	NJJ	-1	-1	-1	nil	nil
componente blastematoso	NJ	0	-1	-1	components	components
componente blastematoso	NJ	-1	-1	-1	nil	nil

Tabla 9. Unidades potencialmente terminológicas descartadas

En la tabla 9 se puede observar cómo dos unidades idénticas son valoradas de diferentes maneras. La primera aparición de las unidades obtiene una valoración que hace que la descartemos como nueva, pero al momento de ver la segunda aparición de la unidad se etiqueta como unidad nueva. Esto fue algo que dificultó el proceso de revisión, puesto que inicialmente se tenían unos resultados preliminares que, tras una revisión más minuciosa de las listas, cambiaron.

4.4. Validación de los candidatos a neologismos especializados

Para validar estos candidatos a neologismos especializados era necesario tener en cuenta que la manera óptima de interpretar la información disponible en los ficheros era ordenar con la ayuda una herramienta de hojas de cálculo los tres criterios siguientes de esta manera: 1) CDwp_nc (mayor a menor), 2) CDwp_nc_w (mayor a menor) y 3) DminTopDom (menor a mayor). Este orden personalizado era posible, puesto que estos tres parámetros tienen como resultados valores numéricos.

4.4.1. Aspectos tenidos en cuenta para la validación

De acuerdo con la guía de uso de la herramienta WikiYATE, disponible en el sitio web del grupo IULATERM¹⁴, la visualización resultante del ordenamiento anterior nos debería facilitar la valoración de las unidades resultantes potencialmente terminológicas, y neológicas para este trabajo. Lo que se debía hacer era dividir la lista resultante en cuatro zonas como se indica a continuación:

- Zona 1 (CDwp_nc entre 0,99-0,01). Se trataba de candidatos que tenían una valoración entre 0 y 1 en la columna CDwp_nc. En esta zona no hubo candidatos a neologismos especializados en ninguno de los dos corpus.
- Zona 2 (CDwp_nc igual a 1). Candidatos con la valoración máxima. Era importante buscar potenciales errores. Una indicación de un candidato potencialmente erróneo es cuando DminTopDom tenía un valor más alto de lo habitual (valores mayores que 5-6 eran sospechosos porque no son del dominio o son generales).
- Zona 3 (CDwp_nc igual a 0). Se trataba de candidatos que el sistema no valoraba como pertenecientes al dominio. Si considerábamos que debería estar incluido entre los términos del dominio era necesario examinarlos individualmente con la herramienta `exploraWPenAPyPP_5.pl`.
- Zona 4 (CDwp_nc igual a -1). Se trataba de candidatos que no se podían detectar con la Wikipedia ya sea porque no están incluidos o bien había algún error interno de la base de datos (en este caso en la columna detección mostraba el mensaje NoCategs)

¹⁴ Disponible en: http://www.iula.upf.edu/rec/aple/2/resul_esp.htm

4.4.2. Depuración de los candidatos

Teniendo en cuenta los parámetros anteriores se hizo una depuración de los candidatos a neologismo terminológico que hay en el ámbito de las enfermedades raras, específicamente en el ámbito de las tres enfermedades raras que seleccionamos para este trabajo, a saber: síndrome de Turner, síndrome de WAGR y neurofibromatosis.

En algunos casos fue fácil decidir si una unidad nueva era terminológica (*neurofibroma*, *progestina*, *micrognatia*, etc.), pero en otras ocasiones fue necesario recurrir a los contextos de las unidades para verificar su carácter terminológico, o bien, para descartarlas. La mayoría de las unidades que fueron necesarias validar con ayuda de los contextos fueron las que tenían el patrón morfosintáctico NJ¹⁵. Esto se presentaba porque muchas de las unidades que WikiYATE consideraba terminológicas eran adjetivos (J), por lo que era necesario validar si hacían parte de una unidad terminológica poliléxica.

Para extraer los contextos de estas unidades se utilizó la herramienta Terminus 2.0¹⁶, la cual es una estación de trabajo para la terminología que permite explorar corpus mediante frecuencias, concordancias, n-gramas y cálculo de asociación entre formas, y también permite la extracción de términos a partir de corpus textuales especializados. Todo este análisis se hizo de manera manual, puesto que era necesario para obtener unos mejores resultados y el tamaño del corpus lo permitía. Esto también permitió que se detectaran posibles errores que hubo durante el procesamiento de los dos corpus, como conversión incorrecta de algunas letras, en el cambio de formato de los textos, o eliminación de espacios entre palabras.

¹⁵Estas etiquetas son las que utiliza por defecto WikiYATE, son las mismas que están en el etiquetario del IULA para el español: <https://www.iula.upf.edu/corpus/etqfrmes.htm>

¹⁶ <http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl>

Los resultados que se obtuvieron después del análisis se presentan en la tabla siguiente:

CDwp_nc	CESP		CPAC	
	no validados	validados	no validados	validados
Zona 2 (1)	9	19	12	9
Zona 3 (0)	4	6	16	
Zona 4 (-1)	53	68	35	44
Total	66	93	63	53

Tabla 10. Neologismos especializados validados

Se descartaron el 46,9% de los candidatos a neologismos especializados en ambos corpus, algunas unidades como *documento, dedicatoria, edad, edición, efectiva, eficaz, eliminación, etc.* es decir, unidades que no se consideraban terminológicas.

De los 159 candidatos a neologismos especializados en el CESP se validaron 93 unidades. Mientras que en el CPAC, de los 116 candidatos iniciales, se validaron 53. Es decir, se validaron 146 neologismos especializados en ambos corpus. Sin embargo, 8 de los neologismos validados se repitieron en ambos corpus, por lo que en definitiva son 138 los neologismos especializados diferentes.

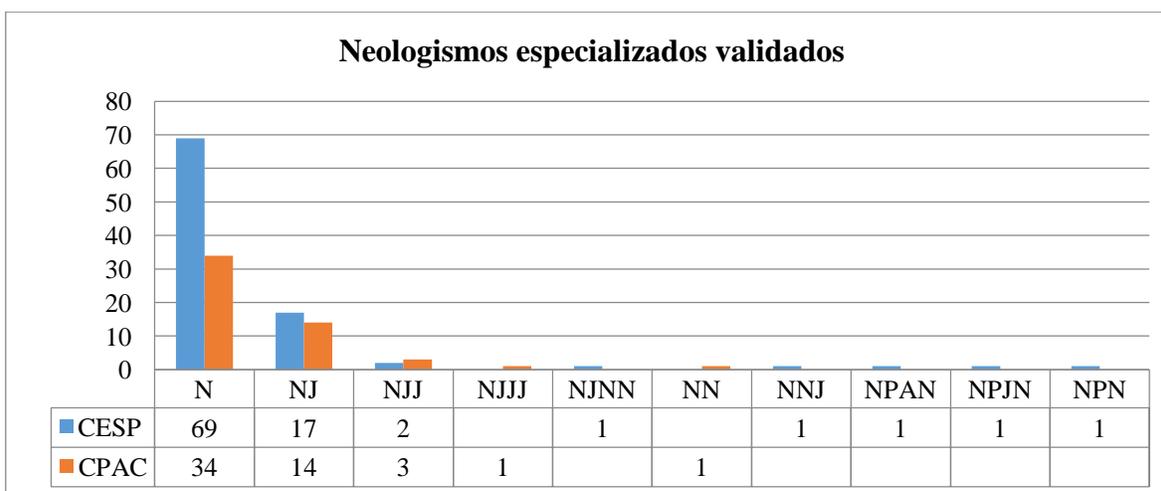


Gráfico 1. Patrones morfosintácticos de los neologismos validados. N: nombre; J: adjetivo; P: preposición; A: determinante.

En la tabla se ve como las unidades terminológicas monoléxicas representan el 74,2% de los neologismos especializados en el CESP y el 64,2% en el CPAC. Las unidades terminológicas poliléxicas de patrón NJ también estuvieron presentes con un 18,3% en el CESP y un 26,4% en el CPAC. Los demás patrones morfosintácticos poliléxicos¹⁷ son mucho menos frecuentes, sin embargo se presentan unos cuantos casos.

A continuación se muestran algunos ejemplos de los patrones más recurrentes¹⁸:

N: *ADNc; adrenarquia; anaplasia; antipirina; arg; audiología; BDNF; blastema; progestina; pubarca; queratopatía; quimiosensibilidad; quinasa; rasopatía; retromicrognatia.*

NJ: *componente blastemal; componente estromal; cribado mutacional; desarrollo puberal; heterogeneidad alélica; mccl asintomático; mesotelio peritoneal; nevus anémico; nevus melanocítico; oclusión dentaria; tejido ovárico; tumor anaplásico.*

4.5. Análisis de las unidades nuevas validadas

Después de obtener la lista de neologismos especializados, se comenzó con todo el análisis lingüístico del trabajo. En primer lugar y con la ayuda de una tabla en Microsoft Excel, para facilitar la manipulación de los datos, se hizo una clasificación tipológica de los neologismos de ambos corpus para definir cuáles son los tipos de neologismos más

¹⁷ Del patrón poliléxico NJNN se presenta un solo caso, *ecotropic viral integration site*, esta unidad no se considera parte de los neologismos creados por sintagmación, sino que se clasifica dentro de los préstamos.

¹⁸ Todos los neologismos especializados detectados se presentan en una tabla completa como anexo.

utilizados en estos dos corpus. De igual manera se hizo el análisis para definir si los textos dirigidos a especialistas o los dirigidos a pacientes tenían mayor presencia de neologismos especializados. De igual manera se hizo una revisión y descripción de los contextos de los neologismos especializados en ambos corpus.

5. TIPOLOGÍA DE LOS NEOLOGISMOS ESPECIALIZADOS

Una vez detectados los neologismos especializados utilizados, se procedió a clasificarlos por mecanismo de formación, es decir tipológicamente. Para hacer esta clasificación se tuvieron en cuenta los criterios que tiene el Observatorio de Neología (OBNEO) (Cabré, 2006; Freixa, Bernal, & Cabré (eds.), 2015). Esta clasificación se hizo de manera manual y en los casos en los que había alguna duda se hizo una consulta en línea en el Diccionario médico-biológico, histórico y etimológico de la Universidad de Salamanca¹⁹ y el Diccionario médico de Larousse²⁰ para ayudar establecer el criterio que correspondiera.

En general los mecanismos de formación más utilizados en la neología especializada del ámbito de las enfermedades raras²¹ son los formales: sufijación, prefijación, composición culta, conversión sintáctica, sintagmación, acronimia, abreviación y variación. También se presentaron casos de neologismos por préstamos y algunos latinismos.

En total se detectaron 146 neologismos especializados en ambos corpus (138 diferentes sin tener en cuenta los 8 en común), 93 en el CESP y 53 en el CPAC. En los gráficos siguientes se pueden observar con detalle los resultados según el recurso de formación que se haya utilizado:

¹⁹ <http://dicciomed.eusal.es/comosecita.php>

²⁰ <http://www.diccionarios.com.sare.upf.edu/diccionariomedico>

²¹ Hay que tener en cuenta que esta es una primera aproximación a la neología especializada de este ámbito en específico.

5.1. Neologismos especializados CESP

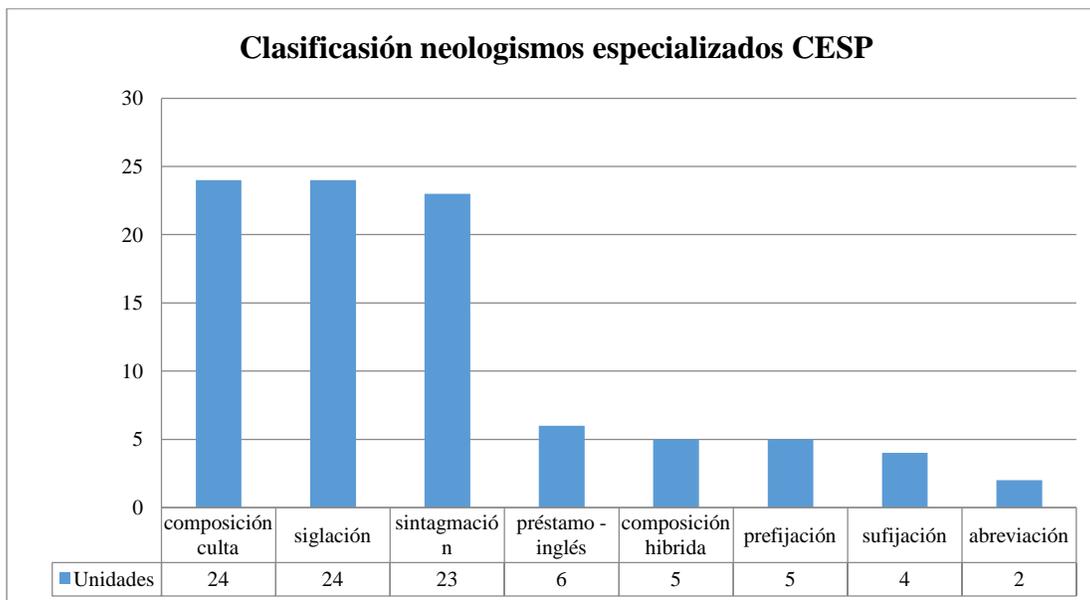


Gráfico 2. Neologismos especializados CESP

En el gráfico 2 se puede ver como los mecanismos de formación de unidades terminológicas más utilizados en los textos dirigidos a especialistas son la composición culta (*adrenarquia, anaplasia, audiología, blastema, coagulopatía, disgerminoma, haploinsuficiencia, linfopenia*) y la siglación, ambos con 24 unidades de 93, es decir el 25,8% de los neologismos especializados del CESP.

En cuanto a la siglación, hay algunas que son iniciales de un sintagma en inglés (*BDNF, CSRD, DHPL, DHPLC, FS, MLPA, mTOR, OMGP, rhGH, SHBG, SNPs*), otras que son iniciales de sintagma en español (*ADNc, DMO, GVO, NF, NFNS, NFP, PKB, ST*) y otras que son siglas alfanuméricas (*EVI2B, NF1, WT1, PAX6*). A continuación está la sintagmación, con un 24,7% (*cariotipo de alta resolución, coartación de la aorta, componente blastemal, componente estromal, cribado mutacional, desarrollo psicosocial, diagnóstico genético prenatal, gen supresor tumoral, heterogeneidad alélica*).

En cuarto lugar, con un 7,5%, está el préstamo del inglés (*ecotropic viral integration site, frameshift, heteroduplex, megabase, missense, nonsense*). Empatado en quinto lugar, con un 5,4%, están la composición híbrida (*cifoescoliosis, estrogenoterapia, morbimortalidad, quimiosensibilidad, troncoencéfalo*) y la prefijación (*antipirina, dehidroepiandrosterona, didrogesterona, hipertelorismo, hipertonía*).

Los recursos menos utilizados en el CESP son la sufijación, con un 4,3% (*estrogenización, malignización, obstipación, tumoración*) y la abreviación, con un 2,2% (*arg, indels*).

5.2. Neologismos especializados CPAC

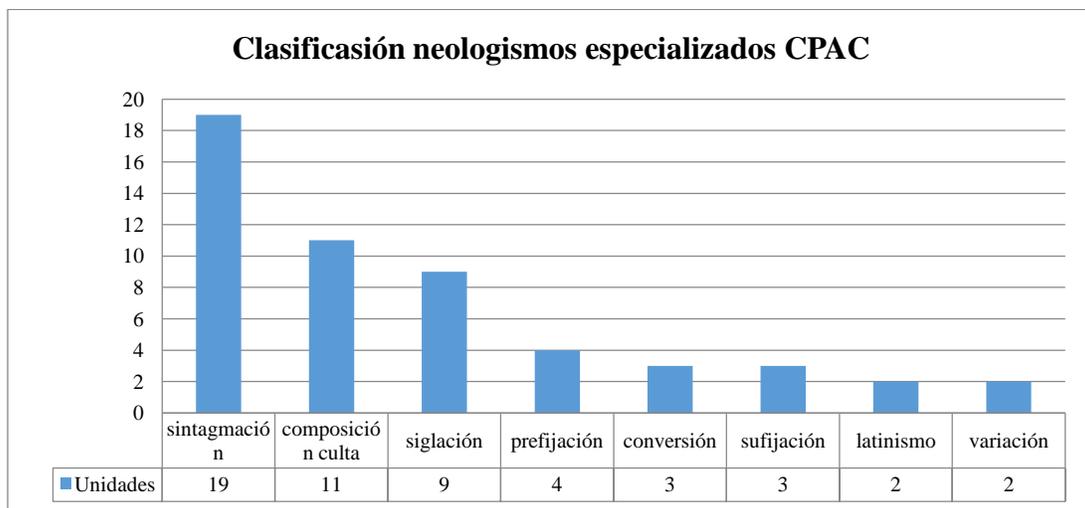


Gráfico 3. Neologismos especializados CPAC

En el gráfico 3 se puede ver de nuevo como en los textos dirigidos a pacientes y familiares, el recurso de formación de unidades terminológicas más utilizado es la sintagmación, 19 unidades de 53, es decir el 35,8% (*anticuerpo antiendomisio, auscultación cardiopulmonar, botón mamario, células germinativas, cicatrices hipertróficas, corazón izquierdo hipoplásico*). Lo siguen las unidades formadas por composición culta, con un 20,8% (*dermatomo, fibromatosis, gonadoblastoma, hamartoma, higroma, nefrouropatía*) y

en tercer lugar la siglación, con un 17%. Algunas de las siglas son iniciales de sintagmas en español (*FC, NF, ST, THS*), otras son iniciales de sintagmas en inglés (*MLPA, SHOX*) y también hay algunas que son siglas alfanuméricas (*NF1, NF2, T4L*).

En cuarto lugar está la prefijación con un 7,5% (*antehélix, hipertelorismo, hipertransaminasemia, neodermis*). Después, empatados con un 5,7%, están la conversión sintáctica (*internista, monosómico, trisómico*²²) y la sufijación (*estrogenización, extirpación, feminización*). En último lugar están empatados, con un 3,8%, los préstamos del latín o latinismos (*mamila* y *prolapso*) y la variación formal ortográfica (*gamagrama* variación de *gammaagrama*; *Z-plastia* variación de *plastia en Z*).

6. DIFERENCIAS ENTRE EL CESP Y EL CPAC

Al momento de hacer un análisis contrastivo de los neologismos especializados encontrados en ambos corpus, lo primero que es necesario resaltar es el grado de densidad neológica especializada de cada uno de ellos, de esta manera podemos mostrar cuáles son los textos que contienen el mayor número de neologismos especializados, si los dirigidos a profesionales o si los dirigidos a pacientes y familiares.

6.1. Densidad neologismos especializados

Para hacer el cálculo de la densidad neológica especializada, en cada uno de los corpus, se tuvo en cuenta la cantidad de unidades potencialmente terminológicas que tenían ambos corpus después de llevar a cabo el análisis con WikiYATE.

²² Los tres casos de conversión sintáctica se dan de adjetivo (J) a sustantivo (N).

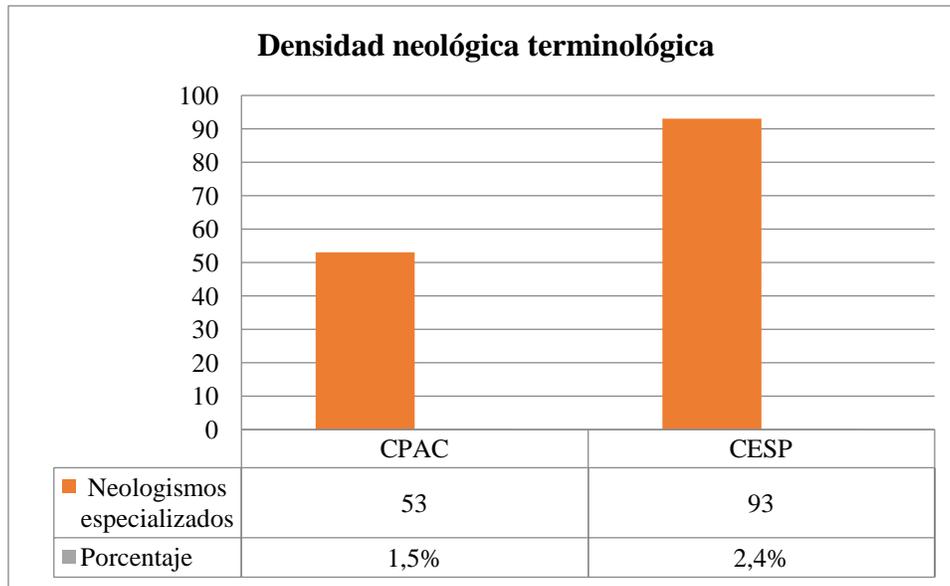


Gráfico 4. Porcentaje de densidad terminológica especializada

Después del análisis con WikiYATE el CPAC generó una lista con 3.822 unidades potencialmente terminológicas, de las cuales se validaron 53 como neologismos especializados, es decir una densidad neológica terminológica del 1,5%. Mientras que en el CESP, de 3.439 unidades se validaron 93, es decir, una densidad del 2,4%. Este dato nos muestra que la densidad neológica especializada de los textos especializados es significativamente mayor que la de los textos dirigidos a pacientes, lo cual representa un grado de dificultad mayor para la comprensión de los textos.

6.2. Tipos más comunes de neologismos

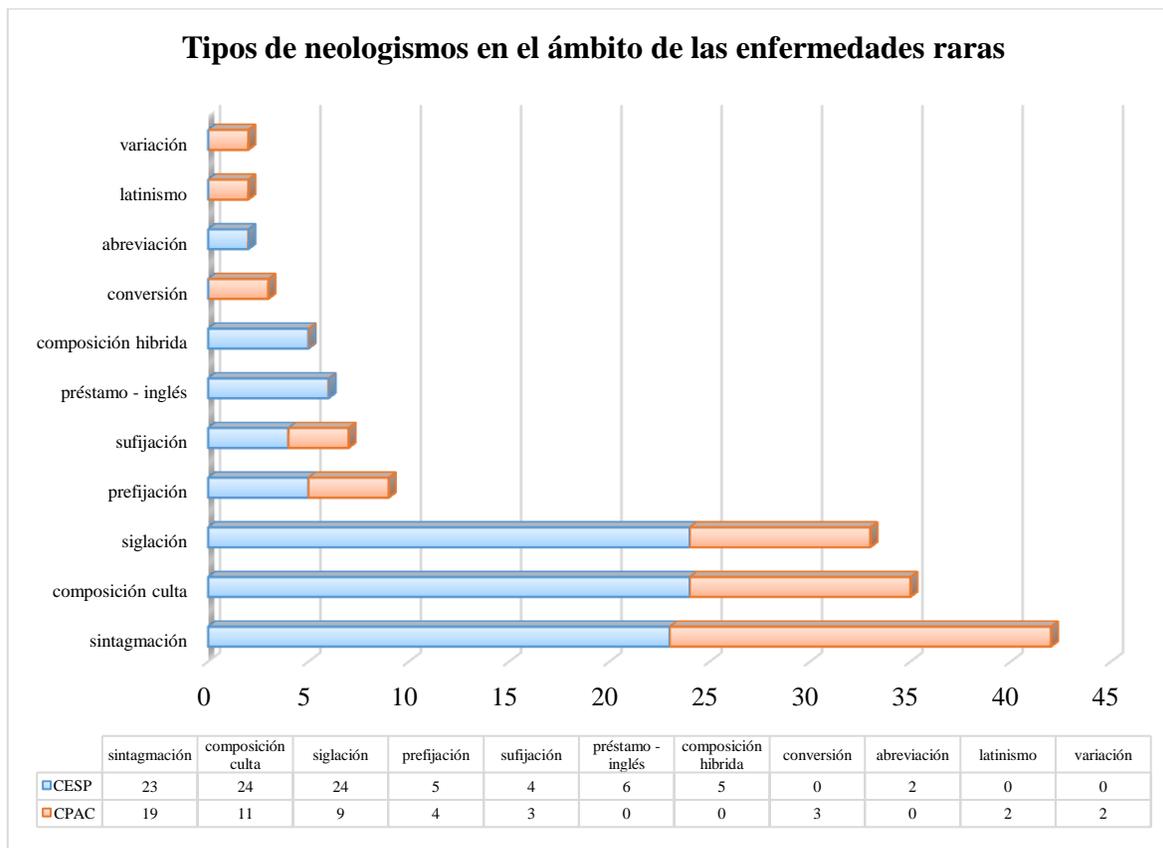


Gráfico 5. Comparación tipología de neologismos en ambos corpus

En general, se puede observar que los mecanismos de creación de neología especializada en ambos corpus son muy similares. La sintagmación está en primer lugar y representa el 30,4% del total de las unidades nuevas en ambos corpus y es un mecanismo que se utiliza casi igual cantidad de veces. Sin embargo, vale la pena destacar que los otros dos mecanismos más productivos, la composición culta y la siglación, se utilizan más del doble de veces en el corpus de especialistas que en el corpus de pacientes

Hay otros recursos de formación que se utilizan algunas veces en el CESP pero que no utilizan el CPAC, estos son el préstamo en inglés, la composición híbrida y la abreviación.

Por el contrario, otros mecanismos como la conversión sintáctica, los latinismos, y la variación formal ortográfica se utilizan en el CPAC pero no en el CESP.

7. EL INGLÉS COMO MEDIO DENOMINATIVO

Las unidades terminológicas utilizadas en situaciones comunicativas especializadas propias del ámbito médico, ámbito en el que enmarcamos las enfermedades raras, están en constante evolución, ya que son inherentes al desarrollo científico. La comunidad médico-científica estadounidense tiene un papel muy importante en este aspecto, puesto que muchos avances y descubrimientos en materia biomédica tienen lugar en el seno de ese grupo de especialistas.

Sánchez Ibáñez (2013) afirma que es por esta razón que en Estados Unidos es donde se decide qué subdominios de la medicina ocupan lugares centrales en las investigaciones y cuáles se quedan relegados a la periferia del avance disciplinar. Asimismo, la manera de transmitir la información que estos especialistas escojan, no sólo por el código empleado (el inglés) sino por las normas de comunicación especializada en general y de producción textual en particular, se convierten en norma para el resto de comunidades científicas, que ven cómo la adecuación de su manera de transmitir el conocimiento se ve condicionada, si quieren que éste sea validado por la comunidad internacional.

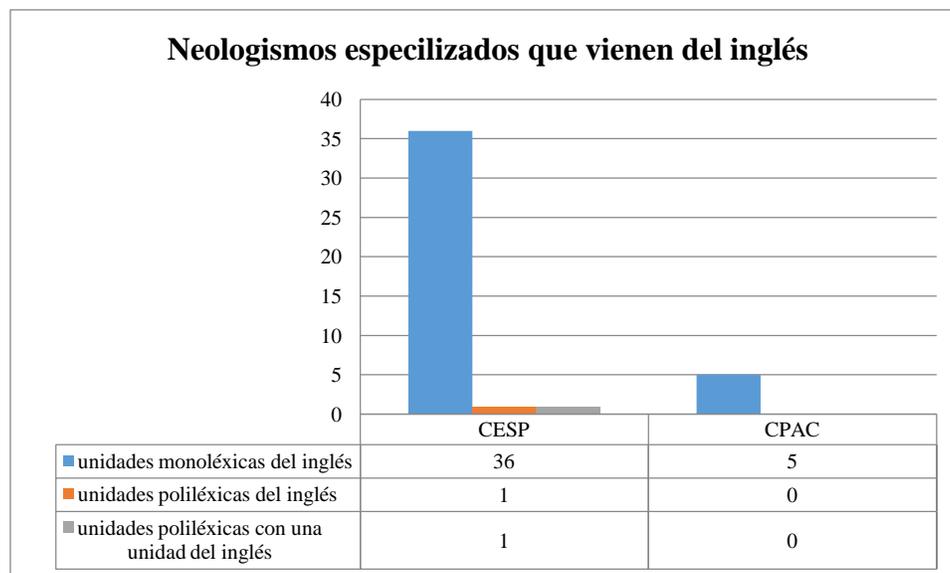


Gráfico 6. Neologismos con unidades del inglés

En el gráfico anterior se pueden observar la cantidad de unidades terminológicas nuevas que vienen o que tienen alguna influencia del inglés. Ya sean unidades monoléxicas o poliléxicas que son préstamos o que tienen un componente, en el caso de las poliléxicas, que viene del inglés. Como se puede observar en el gráfico, esta dependencia denominativa del inglés predomina en el corpus de textos para especialistas. 35,3% en el CESP y 2,6% en el CPAC.

De los 138 neologismos especializados que se detectaron, 43 son o tienen al menos una unidad que viene del inglés, es decir el 31,2%.

Dentro de las unidades monoléxicas hay siglas que son iniciales de sintagmas en inglés (*BDNF, CSRD, DHPL, DHPLC, FS, MLPA, mTOR, OMGP, rhGH, SHBG, SHOX, SNPs, TGGE*) o siglas alfanuméricas del inglés (*PAX6, WTI*).

También hay unidades que en español se forman a partir de la composición culta, pero que inicialmente fueron formadas con este mismo recurso en inglés, por lo que en el momento

de pasar al español se hizo uso de los mismos formantes cultos para tener un equivalente en esta lengua. Ejemplos de de este caso: *anaplasia*, *blastema*, *hamartoma*, *higroma*, *Schwannomatosis*, *xantogranuloma*; y por último, tenemos los préstamos del inglés (*frameshift*, *heteroduplex*, *megabase*, *missense*, *nonsense*).

En cuanto a las unidades poliléxicas, sólo se presentaron dos casos en total en ambos corpus. Uno de los casos es una unidad sintagmática préstamo del inglés, *ecotropic viral integration site*, y la otra es una unidad sintagmática en español que tiene como una de sus unidades una sigla que está formada por las iniciales de un sintagma en inglés, *mccl asintomático*.

8. COTEXTOS DE LOS NEOLOGISMOS

Tras un análisis de los cotextos de los neologismos especializados en ambos corpus, no se encontró ninguna información que introdujera a estas unidades como neologismos. En la mayoría de los casos simplemente se hacía uso de la nueva unidad para hacer referencia al concepto y en otros casos, menos frecuente, se daba una explicación breve de la unidad terminológica.

8.1. Cotextos en el CESP

De los 93 neologismos especializados encontrados en el corpus para especialistas, ninguno de ellos se introduce en el texto como un neologismo del ámbito. Sin embargo, se presentaron dos casos de unidades en las que se da una explicación del concepto al que hace referencia la unidad terminológica. En los demás casos²³, las unidades simplemente se

²³ La lista completa de los cotextos de los neologismos especializados están en un documento anexo.

utilizan en el texto sin dar ningún tipo de indicación de que son terminológicas, mucho menos neológicas.

Neologismo	Cotexto
<i>missense</i>	... <i>i</i> : inserción; <i>md</i> : deleción multiexón; <i>ms</i> : missense (cambio de aminoácido) ; <i>ns</i> : nonsense (mutación...
<i>nonsense</i>	... <i>ms</i> : missense (cambio de aminoácido) ; <i>ns</i> : nonsense (mutación puntual que produce un codón de parada) ...

Tabla 11. Cotextos neologismos CESP

En la tabla se puede ver como se da una explicación del concepto. También se puede observar que estos son dos casos de préstamos del inglés.

8.2. Cotextos en el CPAC

De los 53 neologismos especializados encontrados en el corpus para especialistas, ninguno de ellos se introduce en el texto como un neologismo del ámbito. Sin embargo, se presentaron 18 casos (casi el 34%) de unidades en las que se da una explicación del concepto al que hace referencia la unidad terminológica. Esta explicación consiste en dar alguna de las características del concepto, es decir, si es una enfermedad, se da una explicación de sus síntomas; o si es un dispositivo médico, se da una descripción de lo que hace, entre otras cosas. En la tabla siguiente se muestran estas 18 unidades con sus respectivos cotextos:

Neologismo	Contexto
<i>antehélix</i>	... <i>de la cabeza, con la falta de la curvatura o pliegue normal del interior (antehélix) haciendo que esta se...</i>

<i>cicatriz hipertrófica</i>	<i>... la cirugía la más frecuente es el desarrollo de cicatrices engrosadas y de color rojo (hipertróficas o queloides), en cuyo caso estaría indicado el tratamiento local...</i>
<i>dermatomo</i>	<i>... esta piel se obtiene por medio de unos aparatos que son los dermatomos que pueden ser manuales o eléctricos y que obtienen piel de un espesor de décimas...</i>
<i>diagnóstico presintomático</i>	<i>Si se encuentra la mutación, este estudio ofrece diagnóstico presintomático (niños de corta edad, genéticamente afectados, que aún no manifiesten la enfermedad), diagnóstico prenatal...</i>
<i>disgenesia gonadal</i>	<i>... mayores, clítoris y orificio vaginal. En el síndrome de Turner existe una disgenesia gonadal .esto significa que el desarrollo de las gónadas (ovarios) se detiene en el período embrionario, a las pocas semanas de la gestación.</i>
<i>displasia ungueal</i>	<i>Son frecuentes las displasias ungueales (alteraciones en el desarrollo de las uñas) y los hemangiomas superficiales...</i>
<i>FC</i>	<i>... cuando hablamos del fenotipo conductual (FC) nos referimos al patrón característico de las disfunciones motoras, cognitivas, lingüísticas y sociales que se asocian de forma compatible con un trastorno biológico, sin desestimar...</i>
<i>gonadoblastoma</i>	<i>... ya que existe una cierta predisposición a desarrollar un tumor en los restos ováricos denominado gonadoblastoma .en estos casos generalmente...</i>
<i>hamartoma</i>	<i>... estas manchas pueden ser el único síntoma. Nódulos de Lisch (llamados hamartomas): son áreas pigmentadas en el iris del ojo...</i>
<i>hipertelorismo</i>	<i>... más frecuentes están el estrabismo, ptosis palpebral (párpado caído),</i>

	<i>hipertelorismo (ojos separados) y epicantus (pliegue cutáneo en ángulo interno del ojo).</i>
<i>monosómico</i>	<i>... y si hay un cromosoma de menos, el individuo es monosómico (ej. síndrome de Turner). Si existe mosaicismo, el individuo presenta diferentes líneas celulares...</i>
<i>neurofibromina</i>	<i>... en el gen nf1 cuya función es la producción de una proteína llamada neurofibromina. esta proteína se produce en muchas células, incluyendo las células de los nervios y...</i>
<i>NF2</i>	<i>Neurofibromatosis tipo 2 (NF-2): La neurofibromatosis tipo 2 (NF2) también conocida como NF Acústica Bilateral, es mucho más rara...</i>
<i>pseudoartrosis</i>	<i>... lesiones óseas definidas como adelgazamiento de los huesos largos, falsas articulaciones (pseudoartrosis), sobre todo de la tibia; la escoliosis (encurvamiento de la columna vertebral)...</i>
<i>retrognatia</i>	<i>... (brazos rotados ligeramente hacia fuera a partir del codo), dedo anular corto y retrognatia (retrusión del maxilar inferior).</i>
<i>retromicrognatia</i>	<i>Es frecuente la presencia de un maxilar estrecho, retromicrognatia (mandíbula pequeña y hacia atrás) y paladar ojival.</i>
<i>SHOX</i>	<i>Podría estar relacionada con la pérdida de un gen homeobox denominado con la sigla en inglés SHOX (gen que contiene el homeobox de talla baja en el cromosoma X).</i>
<i>trisómico</i>	<i>... implicar severas consecuencias. Cuando existe un cromosoma de más, el individuo se denomina trisómico (ej. síndrome de down) y si hay un cromosoma de menos...</i>

Tabla 12. Cotextos neologismos CPAC

En la tabla se puede observar cómo se presentan estos neologismos especializados en los textos dirigidos a pacientes y familiares. En ninguno de los casos se da algún tipo de información que indique que las unidades son neológicas. Al igual que en el CESP, las demás unidades²⁴ simplemente se utilizan en el texto sin dar ningún tipo de indicación de que son terminológicas, mucho menos de que son neológicas.

9. CONCLUSIONES Y TRABAJO FUTURO

Hay que tener en cuenta que la finalidad primera de este trabajo era identificar y analizar la neología especializada usada en dos corpus textuales de niveles de especialidad distintos del ámbito de las enfermedades raras, destinados a usuarios distintos. A partir de esto, y después de haber construido ambos corpus de análisis, los resultados obtenidos nos muestran lo siguiente:

- Los mecanismos de creación de neologismos especializados más productivos en el ámbito de las enfermedades raras son la sintagmación, la siglación y la composición culta.
- En los textos dirigidos a especialistas la siglación, las sintagmación, los préstamos del inglés y los préstamos adaptados al español se utilizan mucho más que en los textos dirigidos a pacientes y familiares
- La densidad terminológica especializada es mayor en los textos dirigidos a especialistas que en los dirigidos a pacientes. Esto confirma nuestra primera hipótesis de que los textos dirigidos a profesionales sanitarios contienen un mayor

²⁴ La lista completa de los cotextos de los neologismos especializados están en un documento anexo.

número de neologismos especializados que los textos destinados a pacientes y familiares.

- En los textos dirigidos a especialistas hay una dependencia denominativa del inglés mucho mayor que en los textos dirigidos a pacientes. Esto se puede ver en el tipo de unidades terminológicas nuevas que hay en el corpus para especialistas, es decir, una alta presencia de préstamos, siglas de sintagmas en inglés y unidades poliléxicas con componentes que vienen del inglés. Esto nos confirma nuestra segunda hipótesis de que el ámbito de la medicina, del cual hacen parte las enfermedades raras, tiene una gran dependencia denominativa del inglés.
- Algunos de los mecanismos de formación de neologismos más utilizados en los textos dirigidos a especialistas son la siglación y la composición culta. Esto es algo que nos lleva a pensar que el nivel de opacidad de este tipo de textos es mucho mayor. Otro dato que apoya esta conclusión es que en estos textos también se utiliza el préstamo del inglés, la composición híbrida y la abreviación, mientras que en los textos dirigidos a pacientes y familiares no se utilizan ninguno de estos mecanismos para la creación neológica especializada.
- En cuanto a la tercera hipótesis, acerca de la manera cómo se presentan los neologismos en ambos corpus, no fue posible confirmarla puesto que en ninguno de los casos se da algún tipo de información que indique que se trata de un neologismo especializado. Sin embargo, en muchos casos, principalmente en los textos dirigidos a pacientes y familiares, se hace una definición o una descripción del concepto al que los neologismos hacen referencia, esto nos hace pensar acerca de hasta qué

punto los autores de estos textos hacen esta explicación teniendo en cuenta que son unidades nuevas.

Este trabajo representó una primera aproximación al uso que se le da a la neología especializada en el ámbito de las enfermedades raras. Aun así, estas primeras conclusiones nos confirman el interés que hay por el comportamiento de los neologismos especializados en este ámbito.

Teniendo en cuenta que no se pudo confirmar la tercera hipótesis, consideramos que es necesario seguir con un análisis más a profundidad de la neología especializada del ámbito de las enfermedades raras, en español, en todos los tipos de textos que reciben los pacientes afectados por una enfermedad rara y sus familiares. Primero, con el objetivo de establecer si los autores de estos textos son conscientes de que introducen una unidad nueva y la manera como lo hacen y adicionalmente para conocer cuál es la percepción de neologicidad que tienen los pacientes y sus familias en estos textos. Con los resultados que arroje este trabajo futuro se podría hacer una propuesta para facilitar la comprensión de los textos dirigidos a los pacientes y también la asimilación de la información contenida en estos.

Por otro lado, en cuanto a la extracción de las unidades potencialmente terminológicas, se podría continuar con el trabajo que hace el investigador WikiYATE y contribuir a la actualización de la nueva versión de la Wikipedia, versión 2016. Este trabajo contribuirá a mejorar los resultados que da el análisis con esta herramienta de extracción terminológica.

10. REFERENCIAS

- Aguirre, P., Angulo, M. del C., Guerrero, E., Motero, I., & Prieto, I. (2012). *Manual de atención al alumnado con necesidades específicas de apoyo educativo por padecer. Enfermedades Raras Y Crónicas. IEIM instituto errores inatos del metabolismo* (Vol. Luis). Andalucía (Sp): Junta de Andalucía, Consejería de Educación.
- Cabré, M. T. (1993). *La terminología. Teoría, metodología, aplicaciones*. (S. A. Editorial Empúries, Ed.) (Primera). Barcelona: Hurope, S.A.
- Cabré, M. T. (2006). La clasificación de los neologismos: una tarea compleja. *Alfa, São Paulo*, 50(2), 229-250.
- Cabré, M. T. (2008). El principio de poliedricidad: La articulación de lo discursivo, lo cognitivo y lo lingüístico en terminología. *Iberica*, 16, 9-36.
- Cabré, M. T. (2015). La neologia: un nou camp a la cerca de la seva consolidació científica. *Caplletra. Revista Internacional de Filologia*, 59, 125-136. <http://doi.org/0214-8188>
- Duat Rodríguez, A., Martos Moreno, G. Á., Martín Santo-Domingo, Y., Hernández Martín, A., Espejo-Saavedra Roca, J. ., Ruiz-Falcó Rojas, M. L. ., & Argente, J. . (2015). Características fenotípicas y genéticas en la neurofibromatosis tipo 1 en edad pediátrica. *Anales de Pediatría*, 83(3), 173-182. <http://doi.org/10.1016/j.anpedi.2014.10.010>
- Estopà, R. (1999). *Les unitats terminològiques polilexemàtiques en els lèxics especialitzats (dret i medicina)*. Universitat Pompeu Fabra.

- Estopà, R. (2010). La neologia especialitzada, repte constant per al mediador lingüístic. En E. Coromina & J. M. Mestres (Eds.), *Aspectes de terminologia, neologia i traducció* (pp. 15-39). Barcelona: Institut d'Estudis Catalans.
<http://doi.org/10.2436/15.2503.02.9>
- Estopà, R. (2011). Neologisms at the boundaries of prefixation, composition and syntagmatic composition in catalan: controversial and open questions. *Organon*, 25(50), 275-291. Recuperado a partir de <http://goo.gl/16F8ii>
- Estopà, R. (2012). Medicina i llenguatge: les paraules de la salut. *Llengua, Societat i Comunicació*, 10, 1-5.
- Estopà, R., & Armayones, M. (2015). RecerCaixa JUNTS. Superant barreres socioeducatives i afavorint l'alfabetització en salut. Barcelona (Sp).
- EURORDIS. (2009). ¿Qué es una enfermedad rara? Recuperado 26 de abril de 2016, a partir de <http://goo.gl/CVW2do>
- Freixa, J., Bernal, E., & Cabré, M. T. (Eds.). (2015). *La neología lèxica catalana*. Barcelona (Es): Institut d'Estudis Catalans.
- Gómez, M., & Batista, O. (2015). Neurofibromatosis tipo 1 (NF1) y su diagnóstico molecular como estrategia del diagnóstico diferencial y a edades tempranas. *Revista médica de Chile*, (143), 1320-1330.
- Infogen. (2014). Neurofibromatosis 1, NF-1. Recuperado 23 de mayo de 2016, a partir de <http://goo.gl/Ab2LzB>

- Labarta, J. I., de Arriba, A., Mayayo, E., & Ferrández-Longás, A. (2010). Actualización en el tratamiento y seguimiento del Síndrome de Turner (ST). *Revista Española de Endocrinología Pediátrica*, 1(1), 74-80.
<http://doi.org/10.3266/Pulso.ed.RevEspEP2010.vol1.SuplCongSEEP>
- López Sigüero, J. P. (ed. . (2011). *Preguntas y respuestas sobre el Síndrome de Turner* (2da ed.). Madrid (Es): Novo Nordisk Pharma S.A.
- Lorda-Sanchez, I., González Flores, E. R., & Ayuso, C. (2015). Aniridia, tumor de Wilms y síndrome WAGR. Recuperado 6 de mayo de 2016, a partir de <http://goo.gl/IY1oam>
- Palau, F. (2010). Enfermedades raras, un paradigma emergente en la medicina del siglo XXI. *Medicina clínica*, 134(4), 161-8. <http://doi.org/10.1016/j.medcli.2009.06.038>
- Pérez Bueno, F., Aguilera Pérez, J. R., & Garza Ríos, P. (2015). Reemplazo estrogénico en pacientes con síndrome de Turner. *Perinatología y Reproducción Humana*, 29(3), 122-129. <http://doi.org/10.1016/j.rprh.2015.08.002>
- Rey, A. (1976). Néologisme: un pseudo-concept? *Cahiers de Lexicologie*, 28, 3-17.
- Rocha Choque, M. V., Méndez Pardo, A., Terrazas Saavedra, C., & Pardo Novak, J. A. (2014). Tumor de Wilms unilateral asociado a aniridia: a propósito de un caso. *Gaceta Médica Boliviana*, 37(2), 90-93.
- Sánchez Ibáñez, M. (2013). *Neología y traducción especializada: clave para calibrar la dependencia terminológica español-inglés en el ámbito de la enfermedad del alzheimer*. Universidad de Salamanca. Recuperado a partir de <http://gredos.usal.es/jspui/handle/10366/123060>

Simón, L., Fontáns, E., Duarte, A., & Barité, M. (2013). Estructuras sintagmáticas en tesauros: aproximación cualitativa-cuantitativa. En *III Jornadas de Intercambio y Reflexión acerca de la Investigación en Bibliotecología* (pp. 1-15).

Vivaldi, J., Cabrera-Diego, L. A., Sierra, G., & Pozzi, M. (2012). Using Wikipedia to Validate the Terminology found in a Corpus of Basic Textbooks. En T. (et al. . Calzolari, Nicoletta; Choukri, Khalid; Declerck (Ed.), *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)* (pp. 3820-3827). París: European Language Resources Association (ELRA). Recuperado a partir de http://www.upf.edu/intranet/iula/repositori/pub_complet/vivaldi_2012_2089_cact_aut.pdf
http://www.lrec-conf.org/proceedings/lrec2012/pdf/515_Paper.pdf

Vivaldi, J., & Rodríguez, H. (2010). Using Wikipedia for term extraction in the biomedical domain: first experiences. *Procesamiento del Lenguaje Natural*, 45, 251-254.

ANEXOS

Todos los anexos van en un CD-ROM adjunto a este trabajo:

- Archivo de Excel con todos los neologismos especializados validados y las listas iniciales resultantes del análisis con WikiYATE.
- Tabla completa para la interpretación de los resultados de WikiYATE.