

La compilación de DiCoEnviro en español

María Teresa Ortego Antón

Departamento de Lengua Española – Universidad de Valladolid

Los componentes del OLST, conscientes de que los diccionarios especializados de medio ambiente no satisfacen las necesidades de los usuarios, se propusieron crear un diccionario electrónico en inglés, francés y español, DiCoEnviro, Dictionnaire fondamental de l'environnement, que sigue los principios de la lexicología combinatoria y explicatoria.

En el presente artículo explicamos la metodología seguida para la elaboración de las entradas en español a partir del análisis y extracción de datos de un corpus sobre medio ambiente compilado previamente: la selección de términos, la elección de los contextos, la descripción de la estructura actancial y de las relaciones léxicas y la vinculación de equivalentes.

Palabras clave: corpus, DiCoEnviro, lexicología combinatoria y explicativa, medio ambiente, término.

Members from OLST, being aware that specialized dictionaries about environment do not satisfy users' needs, decided to develop an e-dictionary in English, French and Spanish entitled DiCoEnviro, Dictionnaire fondamental de l'environnement, which follows the principles of Explanatory Combinatorial Lexicology.

In this paper, we explain the methodology to create the Spanish entries by analyzing and extracting data from an environmental corpus compiled previously: the headword selection, the choice of contexts, the description of actantial structure and lexical relations and the linking of equivalents.

Key words: corpus, DiCoEnviro, explanatory and combinatory lexicology, environment, term.

1. INTRODUCCIÓN

El dominio del medio ambiente cada vez cobra más importancia como consecuencia de la diferenciación por actividades socioeconómicas y los expertos necesitan traductores que permitan que el conocimiento traspase las barreras lingüísticas. Los traductores, a su vez, necesitan herramientas fiables como los diccionarios especializados. Sin embargo, la literatura lexicográfica española muestra la insatisfacción de los traductores con este tipo de herramientas (Gutiérrez Rodilla, 1998; García Palacios, 2002; Pastor y Alcina, 2010), entre otros) pero recientemente el panorama ha cambiado gracias a los avances de la informática y a la aplicación de la investigación en lingüística de corpus a la compilación de diccionarios.

Desde el OLST (Université de Montréal – Canadá), el equipo ÉCLECTIK, liderado por la profesora L’Homme (2007), consciente de la necesidad de un diccionario especializado sobre medio ambiente, siguiendo el ejemplo de DiCoInfo, se propuso crear un diccionario electrónico titulado DiCoEnviro, *Dictionnaire fondamental de l’environnement*, que sigue los principios de la lexicología combinatoria y explicativa (Mel’čuk et al. 1984-1999, 2007).

En este artículo pretendemos describir la metodología seguida para elaborar las fichas terminológicas en español de DicoEnviro a partir del análisis y extracción de datos de un corpus sobre medio ambiente, que comenzó en septiembre de 2010. Por el momento, el número de fichas en español asciende a 146, aunque aproximadamente otras 50 están en construcción y no son visibles en línea.

2. CARACTERÍSTICAS DEL CORPUS

El corpus se compone de 85 archivos cuya extensión supera 1,5 millones de palabras (1.571.274 palabras) sobre el dominio del medio ambiente y el subdominio del cambio climático. Dado que el dominio del medio ambiente es muy vasto y heterogéneo, se optó por centrarse únicamente en el subdominio del cambio climático.

La composición del mencionado corpus corrió a cargo de la investigadora Sarah Iveth Carreño Cruz en el marco de su memoria fin de máster.

Durante la selección de los textos que forman el corpus, se intentó incluir textos representativos de las variedades de español. De ahí que la procedencia geográfica de los textos sea variada: México, Argentina, Chile, Ecuador, España, El Salvador, Perú y Venezuela.

Las fuentes de las que proceden los textos que forman el corpus son organismos internacionales, que publicaron informes en línea.

En lo que concierne a la antigüedad de los textos, la mayoría fueron publicados entre 2000 y 2008, aunque una pequeña proporción fue publicada entre 1994 y 1999.

Para denominar los textos que forman el corpus, se ha incluido el nombre del organismo o su acrónimo y las iniciales del país de referencia.

La información que contiene el corpus es de gran relevancia porque permite extraer los términos, los contextos, la información gramatical, los actantes y las relaciones léxicas, semánticas, morfológicas y sintagmáticas.

3. ELECCIÓN DE LOS TÉRMINOS

La primera decisión es seleccionar los términos que formarán parte de la macroestructura de DiCoEnviro. A partir del corpus, utilizamos un extractor automático de términos: TermoStat Web 3.0 (Drouin, 2003). Este extractor genera una lista de candidatos a término basándose en criterios de especificidad. Cada candidato a término recibe una puntuación según la frecuencia en el corpus analizado y la frecuencia en otro corpus pretratado denominado corpus de referencia. Para español, los cinco primeros resultados son:

Tabla 1. Listado de los diez primeros candidatos a término extraídos con TermoStat 3.0

CANDIDATO A TÉRMINO	FRECUENCIA	ESPECIFICIDAD	VARIANTES ORTOGRÁFICAS	MATRIZ
climático	3503	266.2	<i>climático</i> <i>climáticos</i> <i>climáticas</i>	Adjetivo
emisión	4657	264.47	<i>emisión</i> <i>emisiones</i>	Sustantivo
cambio	4535	180.64	<i>cambio</i> <i>cambios</i>	Sustantivo
carbono	1465	174.56	<i>carbono</i>	Sustantivo
gas	2086	171.08	<i>gas</i> <i>gases</i>	Sustantivo

Una vez que disponemos de la lista de candidatos a término, verificamos si cumplen cuatro parámetros (L'Homme, 2008: 90-91):

- Si denotan una entidad ligada a un dominio, por ejemplo *carbono* es un tipo de gas.
- Si presentan actantes de naturaleza especializada, por ejemplo *absorber*: un ecosistema absorbe radiación de la atmósfera.
- Si los vínculos morfológicos van acompañados de vínculos semánticos, por ejemplo dos unidades léxicas que pertenecen a una misma familia: *clima* y *climático*.
- Si existen otros vínculos paradigmáticos, por ejemplo dos términos que son sinónimos o antónimos, por ejemplo *absorción* y *emisión*.

Una vez seleccionados los términos, se pasa a la siguiente etapa, la redacción de la ficha terminológica.

4. LA REDACCIÓN DE LA FICHA TERMINOLÓGICA

En esta fase nos ayudamos de TextSTAT⁵², un analizador de concordancias automático gratuito que busca en el corpus todas las ocurrencias de un término, con el contexto anterior y posterior. En nuestro caso, elegimos ver en pantalla 75 palabras antes y después del término escogido.

Una vez que tenemos todos los contextos en los que aparece el término, comenzamos con lo que propiamente se puede denominar la redacción de la ficha terminológica. Utilizamos una plantilla en formato XML y el programa Oxygen. La plantilla de ficha en blanco es común para todas las lenguas y similares a las utilizadas por DiCoInfo (L'Homme, 2008, 2009). Antes de comenzar con el análisis de datos, completamos la información de gestión: el autor (TOA) y la fecha (aaaa-mm-dd).

Continuamos escogiendo la forma de lematización del término, la más frecuente de los contextos y siguiendo estas normas:

- Los sustantivos se lematizan en singular.
- Los adjetivos se lematizan en masculino singular.
- Los verbos se lematizan en infinitivo.

Después, describimos las diferentes partes que componen la redacción de la ficha terminológica.

4.1. Significado y contexto

De la observación detallada de todas las ocurrencias del término en el corpus con el analizador de concordancias, distinguimos las diferentes acepciones (1, 2, etc.) y los significados próximos (1a, 1b, 1c). Ejemplo:

- “contaminante”: contaminante₁ (sustantivo) y contaminante₂ (adjetivo);
- “río”: río_{1a} (tipo de ecosistema) y río_{1b} (corriente de agua dulce).

Para cada significado diferente, intentamos extraer 20 contextos que sean representativos e ilustren los actantes y la fraseología con la que se utiliza la unidad léxica y la fuente de la que proceden. Se ordenan alfabéticamente según la denominación del texto del corpus del que proceden. En la versión en línea únicamente se pueden obtener los tres primeros contextos, por lo tanto estos contextos deben ilustrar con precisión los actantes y las posibles relaciones. Ejemplo:

ecosistema , n.m.

Contextos

Otras causas directas de la pérdida del capital natural son la contaminación de los ecosistemas, así como la introducción de especies invasoras y la variabilidad climática. (Fuente: 3A_COMUN_MX)

52 <http://neon.niederlandistik.fu-berlin.de/textstat/>

Con un gran nivel de certeza se puede asegurar que el cambio climático hará que parte de los ecosistemas acuáticos continentales españoles pasen de ser permanentes a estacionales (Fuente: 4A_COMUN_UNFCC_ES)

Las temperaturas y humedad inusualmente altas parecen estar afectando a los ecosistemas boscosos. (Fuente: CC_AMERICA_LATINA)

4.2. La estructura actancial

Una vez que hemos delimitado las acepciones y hemos seleccionado los contextos, nos centramos en la estructura actancial, que L’Homme (2008: 94) describe como “*la structure actancielle décrit les participants essentiels pour décrire le sens d’un terme. Il ne s’agit pas d’une définition à proprement parler, mais cette structure fournit déjà des éléments importants sur le sens des termes*”⁵³.

La estructura actancial describe los participantes involucrados en el significado del término: los agentes, los pacientes, el destino, la superficie, la fuerza natural, etc.

A partir de nuestro conocimiento terminológico y con la ayuda de los contextos exponemos la estructura actancial. Por ejemplo, para absorber:

absorber₁, v. tr.

Estructura actancial: absorber: Destino {ecosistema 1, gas 1} ~ Paciente {radiación 1} de Fuente {atmósfera 1}

Una vez determinada la estructura y las realizaciones que servirán de ejemplo, rastreamos el corpus en busca de posibles unidades léxicas que actúen como Destino, Paciente y Fuente de absorber y las incluimos en orden alfabético. Si alguna de ellas fuera objeto de una ficha del diccionario, la vinculamos, de ahí que el navegador la muestre en azul y permita la referencia cruzada con otras fichas de DiCoEnviro en español.

absorber₁, v. tr.

Estructura actancial: absorber: Destino {ecosistema 1, gas 1} ~ Paciente {radiación 1} de Fuente {atmósfera 1}

Relaciones lingüísticas de los actantes

destino
aerosol ₁ , árbol, biosfera ₁ , bosque ₁ , ecosistema ₁ , gas ₁ , gas de efecto invernadero, mar, océano ₁ , planta, suelo, superficie, sustancia, Tierra, vapor, vegetal
paciente
calor ₁ , dióxido de carbono, energía ₁ , gas ₁ , radiación ₁
fuelle
atmósfera ₁ , sol, Tierra

⁵³ “La estructura actancial describe los participantes esenciales para describir el significado de un término. No se trata de una definición en sentido estricto, sino que dicha estructura proporciona elementos importantes sobre el significado de los términos”.

Destacamos que en el caso de términos de significado cercano, los actantes coinciden, como podemos comprobar entre absorber y absorción:

absorción ₁, n. f.

Estructura actancial: absorción: ~ de Paciente {radiación 1} por Destino {sumidero}

Relaciones lingüísticas de los actantes

paciente
calor ₁ , carbono 1, dióxido de carbono, <u>energía</u> ₁ , <u>gas</u> ₁ , gas de efecto invernadero, <u>radiación</u> ₁
destino
árbol, <u>bosque</u> ₁ , <u>ecosistema</u> ₁ , capa de ozono, planeta, sumidero, superficie

Una vez que hemos completado la estructura actancial, pasamos a la siguiente etapa, las relaciones léxicas.

4.3. Las relaciones léxicas

La penúltima etapa de la redacción de los artículos es confeccionar la lista de relaciones léxicas, compuesta por relaciones paradigmáticas y sintagmáticas. Para codificarlas se utilizan dos niveles: el primer nivel, despojado de metalenguaje técnico, va dirigido al usuario que desea acceder a la información sobre los términos y su combinatoria; y el segundo nivel, dirigido a los lingüistas, lexicógrafos, terminólogos y traductores. Existe un tercer nivel que no aparece en la web, que se refiere a la función léxica siguiendo las pautas de la lexicología y la semántica léxica (Polguère, 2008).

A su vez, esta categoría se compone de cinco subcategorías: significados relacionados, opuestos, otras categorías gramaticales y derivados, tipos de, combinaciones y otros.

4.3.1. Significados relacionados

En primer lugar, identificamos las variantes, sinónimos, cuasisinónimos y los significados relacionados. Si existen sinónimos cuyo grado de sinonimia es total o variantes, se describen en una rúbrica denominada “sinónimos”. Ejemplo:

carbono ₁, n. m.

Sinónimo(s): C

El resto de relaciones de sinonimia se describen en el apartado “Significados relacionados”. A continuación exponemos varios casos de relaciones entre conceptos:

Tabla 2: Relaciones de sinonimia y de significados relacionados

Explicación - término típico	Explicación - rol actancial	Término relacionado
conservación	Sinónimo	preservación
erosionar	significado relacionado	desertificar
hielo	significado relacionado	glaciar

4.3.2. Opuestos

La siguiente etapa es la identificación de las relaciones de antonimia, oposición, etc.

Tabla 3. Relaciones de antonimia y oposición

Explicación - término típico	Explicación - Rol actancial	Término relacionado
conservar 1	Antónimo	dañar
capturar 1	Antónimo	emitir 2.1
disminución	Opuesto	subida
inundación	Contrastivo	sequía

4.3.3. Otras categorías gramaticales y derivados

Continuamos completando las relaciones morfológicas, es decir, términos que pertenecen a la misma familia. Para encontrar estas relaciones, se utiliza la búsqueda truncada. Por ejemplo, buscando atmósfer* obtenemos como resultado atmósfera y atmosférico.

4.3.4. Tipos de

Esta categoría se utiliza principalmente para describir los tipos de un determinado sustantivo encontrados en el corpus. Por ejemplo, los tipos de ecosistema:

ecosistema ₁, n. m.

Relaciones léxicas

Roles actanciales

Explicación - término típico	Explicación - rol actancial	Término relacionado
Tipos de		
Tipo de e.	Tipo de «palabra clave»	<u> río</u> ₁
Tipo de e.	Tipo de «palabra clave»	<u> manglar</u> ₁
Que trata de un lugar específico	Que trata de un lugar específico	~ marino
Que trata de un lugar específico	Que trata de un lugar específico	~ terrestre

4.3.5. Combinaciones

Seguimos completando la información sobre la fraseología de los términos. Como apunta L’Homme (2008: 78-79), muy pocos diccionarios especializados describen las combinaciones léxicas típicas. Sin embargo, la información fraseológica es muy útil para el usuario en tareas de producción de L1 a L2 o en traducción inversa porque se describen los verbos con los que se utilizan los términos sustantivo, por ejemplo para bosque: aprovechamiento de un ~, deforestación de un ~, sostenibilidad de un ~, etc. Ejemplo:

ecosistema ₁, n. m.

Relaciones léxicas

Roles actanciales

Explicación - término típico	Explicación - rol actancial	Término relacionado
Combinaciones		
Un e. se vuelve diferente	Una «palabra clave» se vuelve diferente	<u>cambiar</u> _{1a}
Sustantivo para algo o alguien que produce una variación en un e.	Sustantivo para algo o alguien que produce una variación en una «palabra clave»	<u>cambio</u> ₂ en un ~
Algo o alguien constituye un riesgo potencial para la riqueza de un e.	Algo o alguien constituye un riesgo potencial para la riqueza de una «palabra clave»	<u>amenazar</u> ₁ un ~
-> SUSTANTIVO	-> SUSTANTIVO	<u>amenaza</u> ₂ para un ~
Sustantivo para la e. que se vuelve peor	Sustantivo para la «palabra clave» que se vuelve peor	<u>degradación</u> _{1a} de un ~
Algo o alguien causa que un e. se vuelva peor	Algo o alguien causa que una «palabra clave» se vuelva peor	<u>afectar</u> ₁ a un ~
Algo o alguien conserva un e. en su estado actual	Algo o alguien conserva una «palabra clave» en su estado actual	<u>conservar</u> ₁ un ~
-> SUSTANTIVO	-> SUSTANTIVO	<u>conservación</u> ₁ de un ~
Sustantivo para algo o alguien que conserva un e. en su estado actual	Sustantivo para algo o alguien que conserva una «palabra clave» en su estado actual	<u>preservación</u> ₁ de un ~

4.3.6. Otros

El último apartado de las relaciones léxicas da cabida a todas las relaciones que no encajan en ninguno de los apartados anteriores pero que pueden resultar interesantes para los usuarios. Ejemplo:

atmósfera , n. f.

Relaciones léxicas

Roles actanciales

Explicación - término típico	Explicación - rol actancial	Término relacionado
Otros		
División	División	baja ~
División	División	estratosfera
División	División	troposfera

4.4. Los equivalentes

La última etapa en la confección de las entradas de DiCoEnviro en español es la búsqueda de equivalentes. Para vincular una unidad léxica con otra en inglés o francés, observamos si los significados coinciden y si los actantes en una y otra lengua son similares. Si coinciden, los relacionamos entre ellos. Por ejemplo, “ecosistema” está vinculado con el equivalente *ecosystem* en inglés y *écosystème* en francés.

Una vez completada la ficha, se revisa y se cambia el estatus de 3 a 2, con el fin de que se pueda consultar en línea. La consulta en línea puede realizarse mediante el buscador o mediante un índice alfabético de palabras. El resultado completo de una ficha puede consultarse en el apéndice.

5. CONCLUSIÓN

Como podemos apreciar, DiCoEnviro, diccionario fundamental del medio ambiente, es un diccionario en fase de construcción que pretende describir los términos del medio ambiente (subdominio cambio climático). La novedad de esta obra reside en la descripción de las relaciones léxicas, semánticas y sintagmáticas a partir de la información de un corpus siguiendo los principios de la lexicología combinatoria y explicatoria y la vinculación de equivalentes en tres lenguas: inglés, francés y español.

Sin embargo, está todavía en fase de construcción. Entre las tareas pendientes, resaltamos la inclusión de más términos, la anotación de contextos y la definición de los significados. Además, por ahora los términos incluidos son univerbales, pero sería necesario incluir los términos pluriverbales, como por ejemplo “dióxido de carbono” y ampliar la cobertura a otras áreas del medio ambiente.

Su estatus “en construcción” podría ser una restricción para los usuarios, pero aún sin ser una obra acabada, esta herramienta es muy útil: es fiable (basado en corpus), accesible (disponible en Internet) e incluye información sobre las relaciones léxicas, semánticas y sintagmáticas, que no es fácil encontrar en otros diccionarios de tipología similar.

6. AGRADECIMIENTOS

Al equipo ÉCLECTIK del OLST (Université de Montréal – Canadá), en especial, a su directora, la prof. Marie-Claude L’Homme, por ofrecerme la oportunidad de trabajar en el proyecto de redacción de las entradas en español de DiCoEnviro (la lista completa de investigadores puede consultarse en <http://olst.ling.umontreal.ca/dicoenviro/equipe-es.html/>); y a la Universidad de Valladolid por la Ayuda para Estancias Breves, sin su apoyo económico no habría sido posible la colaboración en el mencionado proyecto.

7. APÉNDICE

calor, s. m.

Estructura actual de calor - creado por Funes/Josete

Relaciones léxicas de los nodos

buscar

actividad. fuente. sol. Tierra

Contexto

Los cambios en la temperatura reflejan, por ejemplo, un incremento de calor que devienen con frecuencia a incrementos en las temperaturas máximas y el número de días considerados muy cálidos. (Source: SA, COARUN, MO)

Pueden producirse cambios importantes (y de hecho ya se aprecian en ciertas medidas en el calor atmosférico por los mares, en el nivel del mar o en las planicies, en la intensidad y dirección de los vientos de circulación en altura, la intensidad y posición de las principales isobaras de altura), en (Source: CC, SUDINTERKANTO, ESPAÑA)

Por un lado, la estructura y evolución del área urbana, la actividad urbana (transporte, economía urbana, etc.) y sus efectos en la generación de calor y emisión de gases de invernadero son sujetos de preocupación por la alta tasa de crecimiento urbano a nivel global y de América Latina. (Source: URBANIZACION, CAMBIOS, AL)

Relaciones léxicas

Palabras relacionadas

Explicación - término riguro	Explicación - ref. contextual	Término relacionado
Opuestos	Antónimo	frío
Otras categorías gramaticales e derivadas		
Que contiene c:	Que contiene a palabra clave:	calorífico
Combinaciones		
La fuente produce c:	La fuente produce la « palabra clave »:	emite, _
La fuente produce c:	La fuente produce la « palabra clave »:	produce, -
-> SUSTANTIVO	-> SUSTANTIVO	generación, de, _
El c. sabe:	La « palabra clave » sabe:	incrementa, _
Sinónimo para la c. que sabe:	Sinónimo para la « palabra clave » que sabe:	avanza, _ de, _
El c. baja:	La « palabra clave » baja:	decreta, _
-> SUSTANTIVO	-> SUSTANTIVO	decrementa, _ de, _
Algo o alguien espera a tener c:	Algo o alguien espera a tener la « palabra clave »:	disfruta, el, _
-> SUSTANTIVO	-> SUSTANTIVO	disfrutar, de, _
Algo o alguien tiene el c:	Algo o alguien tiene la « palabra clave »:	abstrae, el, _
Algo o alguien tiene el c:	Algo o alguien tiene la « palabra clave »:	ataque, el, _

Buscar calor

Edición: TCM NCLM
Última actualización: 05-11-2010

Figura A1.

8. REFERENCIAS BIBLIOGRÁFICAS

- DROUIN, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9 (1), 99-117.
- GARCÍA PALACIOS, J. (2002). El artículo lexicográfico en el diccionario de especialidad. En I. Ahumada (2002). *Diccionario y lenguas de especialidad. V seminario de lexicografía hispánica. Jaén, 21 al 23 de noviembre de 2001*. Jaén: Universidad de Jaén.
- L'HOMME, M.C. (2007). Using Explanatory and Combinatorial Lexicology to Describe Terms. En L. Wanner (Ed.). *Selected Lexical and Grammatical Topics in the Meaning-Text Theory. In Honour of Igor Mel'cuk* (pp. 11-50). Amsterdam/Philadelphia: John Benjamins.
- L'HOMME, M.C. (2008). Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés. *Traduire*, 217, 78-103.
- L'HOMME, M.C. (2009). *DiCoInfo: Dictionnaire fondamental de l'informatique et de l'Internet*. Disponible en <http://olst.ling.umontreal.ca/dicoinfo/manuel-DiCoInfo.pdf> [Fecha de consulta: 21 de marzo de 2011].
- L'HOMME, M.C. Y LANEVILLE, M.E. (2010). *DiCoEnviro: El diccionario fundamental del medio ambiente*. Disponible en http://olst.ling.umontreal.ca/dicoenviro/Dicoenviro_manual_Es.pdf [Fecha de consulta: 21 de enero de 2011].
- MEL'ČUK, I. Y OTROS (1984-1999). *Dictionnaire explicatif et combinatoire du français contemporain*. Recherches lexico-sémantiques I-IV. Montréal: Les Presses de l'Université de Montréal.
- MEL'ČUK, I. Y POLGUÈRE, A. (2007). *Lexique actif du français*. Bruxelles: Duculot.
- PASTOR, V. Y ALCINA, A. (2010). Search Techniques in Electronic Dictionaries: A Classification for Translators, *International Journal of Lexicography* 23 (3), 307-354.
- POLGUÈRE, A. (2008). *Lexicologie et sémantique lexicale. Notions fondamentales*. Montréal: Les Presses de l'Université de Montréal.
- RODILLA, B. M. (1998). *La ciencia empieza en la palabra*. Barcelona: Península.

