

Presente y futuro de la base de conocimiento terminológica EcoLexicon

*Present and future of the terminological
knowledge base EcoLexicon*

Antonio San Martín

Univ. de Quebec en Trois-Rivières
Canadá

Pilar León-Araúz

Universidad de Granada
España

Melania Cabezas-García

Universidad de Granada
España

Arianne Reimerink

Universidad de Granada
España

Miriam Buendía-Castro

Universidad de Castilla-La Mancha
España

Pamela Faber

Universidad de Granada
España

Beatriz Sánchez-Cárdenas

Universidad de Granada
España

ONOMÁZEIN 49 (septiembre de 2020): 174-202
DOI: 10.7764/onomazein.49.09 | ISSN: 0718-5758



Antonio San Martín: Departamento de Lenguas Modernas y Traducción, Universidad de Quebec en Trois-Rivières, Canadá. | E-mail: antonio.san.martin.pizarro@uqtr.ca

Melania Cabezas-García: Depto. de Traducción e Interpretación, U. de Granada, España. | E-mail: melaniacabezas@ugr.es

Miriam Buendía-Castro: Departamento de Filología Moderna, Universidad de Castilla-La Mancha, España.
| E-mail: miriam.buendia@uclm.es

Beatriz Sánchez-Cárdenas: Depto. de Traducción e Interpretación, U. de Granada, España | E-mail: bsc@ugr.es

Pilar León-Araúz: Depto. de Traducción e Interpretación, U. de Granada, España. | E-mail: pleon@ugr.es

Arianne Reimerink: Depto. de Traducción e Interpretación, U. de Granada, España. | E-mail: arianne@ugr.es

Pamela Faber: Depto. de Traducción e Interpretación, Universidad de Granada, España. | E-mail: pfaber@ugr.es

Fecha de recepción: mayo de 2018 – Fecha de aceptación: marzo de 2019

Resumen

EcoLexicon es una base de conocimiento terminológica multilingüe sobre ciencias medioambientales desarrollada desde 2003 por el grupo de investigación LexiCon de la Universidad de Granada (España) y constituye la aplicación práctica de la teoría de la terminología basada en marcos. El presente artículo describe el funcionamiento de EcoLexicon y presenta sus últimos avances, que incluyen un nuevo corpus y una gramática semántica de *word sketches* en inglés, una reforma del módulo fraseológico, un enfoque flexible a las definiciones terminológicas y la representación conceptual mediante imágenes.

Palabras clave: base de conocimiento terminológica; terminología medioambiental; EcoLexicon.

Abstract

EcoLexicon is a multilingual terminological knowledge base on environmental science developed since 2003 by the LexiCon research group at the University of Granada (Spain). It constitutes the practical application of the Frame-based Terminology theory. This article describes how EcoLexicon works and presents its latest developments, including its new corpus and semantic word sketch grammar in English, a revised and improved phraseology module, a flexible terminological definition approach, and conceptual representation through images.

Keywords: terminological knowledge base; environmental terminology; EcoLexicon.

1. Introducción

EcoLexicon (<http://ecolexicon.ugr.es>) (Faber y otros, 2011; Faber y Buendía-Castro, 2014; Faber y otros, 2016) es una base de conocimiento terminológica (BCT) multilingüe sobre ciencias medioambientales desarrollada por el grupo de investigación LexiCon (<http://lexicon.ugr.es>) de la Universidad de Granada (España). Diseñada y puesta en funcionamiento en 2003, recoge actualmente 4 469 conceptos y 23 520 términos en inglés, español, alemán, francés, ruso y griego moderno.

EcoLexicon es la aplicación práctica de la terminología basada en marcos (TBM) (Faber y otros, 2006; Faber, 2012, 2015), una teoría sobre la representación del conocimiento especializado que emplea algunos aspectos de la semántica de marcos (Fillmore, 1982; Fillmore y Atkins, 1992) para estructurar los dominios especializados y crear representaciones que no dependen de ninguna lengua. La TBM se centra en la organización conceptual, la naturaleza multidimensional de las unidades de conocimiento especializado y la extracción de la información semántica y sintáctica a través del uso de corpus multilingües.

Entre los usuarios a los que va dirigida esta BCT se encuentran tanto profesionales de la lengua y expertos en medioambiente como el público general. Las representaciones que se ofrecen en EcoLexicon están creadas para traductores, redactores técnicos y expertos medioambientales que necesiten alcanzar un mejor conocimiento sobre conceptos especializados del medioambiente con vistas a la redacción o traducción de textos especializados o semiespecializados.

A continuación, en la sección 2, se describen la macroestructura y microestructura de EcoLexicon. En la sección 3 se presentan los últimos avances en EcoLexicon, a saber: su nuevo corpus y su gramática semántica de *word sketches* en inglés, el módulo fraseológico reformado, el enfoque de la definición terminológica flexible y la representación conceptual mediante imágenes. Por último, en la sección 4 se exponen de manera sucinta las conclusiones en cuanto a las contribuciones de EcoLexicon a la terminografía.

2. Macroestructura y microestructura de EcoLexicon

EcoLexicon es un recurso orientado al concepto y, por lo tanto, cada entrada corresponde a un único concepto. Una entrada típica incluye los términos que designan el concepto en las distintas lenguas, una definición en inglés y español (ver sección 4.3) y recursos como imágenes, documentos, direcciones URL, material audiovisual, etc. (ver sección 4.4). Asimismo, todos los términos van acompañados a su vez de información asociada: tipo de término (término principal, sinónimo, variante, acrónimo, etc.), categoría gramatical y género (figura 1). Actualmente, se está trabajando en la inclusión de un módulo fraseológico que se centrará en la combinatoria de cada término (ver sección 4.2).

FIGURA 1

Interfaz de EcoLexicon con la ventana de información asociada al término *litosfera* abierta

The screenshot displays the EcoLexicon web interface. At the top, there is a search bar with 'litosfera' entered and a 'Buscar' button. To the right, there are options for 'Dominios contextuales' (set to 'No dominio'), 'English', and 'Feedback'. Below the search bar, there are tabs for 'Historia', 'Resultados búsqueda', 'A-Z', and 'Camino'. The main content area is divided into several sections:

- Definición:** A text box containing the definition of 'litosfera': 'zona del manto terrestre más superficial, formada por todas las partes sólidas de la Tierra y que contiene la corteza, parte del manto, los continentes y las placas tectónicas.'
- Términos:** A list of terms in different languages: Lithosphere (English), litosfera (Spanish), Lithosphäre (German), Литосфера (Russian), and Λίθοςσφαιρα (Greek).
- Recursos:** A section for additional resources.
- Categorías conceptuales:** A section for conceptual categories.
- Frasesología:** A section for phraseology.

The central part of the interface features a large network diagram with 'Litosfera' as a central node. It is connected to various other terms such as 'Placa tectónica', 'Cámara magmática', 'Asthenosfera', 'Crotón', 'Capa de evian', 'Capa de dispersión profunda', 'Capa de mezcla', 'Helocina', 'Hidrosfera', 'Capa confinante', 'Estrato_capa', 'Capa límite planetaria', 'Capa de densidad', 'Lamina_agua', 'Desertificación', 'Flechaitoral', and 'Huso 30'. A pop-up window titled 'Información del término' is open over the 'Litosfera' node, displaying the following details:

Término:	litosfera
Idioma:	español
Tipo de término:	término principal
Categoría gramatical:	nombre común
Género:	femenino

Below the pop-up window, there is a 'Ver concordancias' button and a legend for relationship types: 'Relaciones topológicas', 'Relaciones meronímicas', and 'Relaciones de jerarquía'.

Todas las entradas de EcoLexicon están conectadas entre sí mediante una estructura conceptual subyacente. Por un lado, se asigna a todos los conceptos de EcoLexicon una o varias categorías conceptuales genéricas (por ejemplo, AGENTE NATURAL, PROCESO ARTIFICIAL, etc.). Por otro lado, en EcoLexicon, los conceptos están asociados entre sí mediante relaciones de tipo tanto jerárquico como no jerárquico, lo que incluye relaciones específicas del dominio del medio ambiente. La inclusión de este tipo de relaciones aporta mayor coherencia y dinamismo a la representación conceptual (Faber y otros, 2009: 7). Las relaciones entre conceptos en EcoLexicon se expresan mediante proposiciones formadas por dos conceptos y la relación que los une: «CONCEPTO *relación* CONCEPTO». El inventario de relaciones conceptuales utilizadas en EcoLexicon es el fruto de un análisis de corpus mediante el cual se verificó la capacidad de cada relación para unir conceptos según la naturaleza de estos (León-Araúz y Faber, 2010; Faber y León-Araúz, 2010). Partiendo de esta base, se estableció la lista de las relaciones junto con sus criterios de uso (León-Araúz, 2009: 301-308; León-Araúz y otros, 2012: 130-140):

- *Tipo_de*: Se trata de la tradicional relación genérico-específica. Representa el resultado de la categorización y estructura las jerarquías taxonómicas. La proposición «OXÍGENO *tipo_de* ELEMENTO QUÍMICO» supone que ELEMENTO QUÍMICO es un concepto superordinado de OXÍGENO o, en otras palabras, que OXÍGENO pertenece a la categoría de ELEMENTO QUÍMICO. Es posible que un concepto tenga varios superordinados debido a la multidimensionalidad, lo que da lugar a la herencia múltiple y, por tanto, a la coexistencia de distintas jerarquías de conceptos (León-Araúz y San Martín, 2012).
- *Parte_de*: Esta es la relación meronímica básica y relaciona una entidad física con sus partes, por ejemplo, «ESTRIBO *parte_de* PRESA», o una disciplina científica con sus subdisciplinas, como «GLACIOLOGÍA *parte_de* HIDROLOGÍA». EcoLexicon incluye además cinco subtipos

de relaciones meronímicas: *fase_de* (une un proceso con sus fases, por ejemplo, «DIAGÉNESIS *fase_de* LITOGÉNESIS»); *compuesto_de* (une tanto objetos artificiales como naturales con el material del que están hechos, por ejemplo, «PEDRAPLÉN *compuesto_de* PIEDRA»); *delimitado_por* (une objetos físicos que se delimitan mutuamente como «MESOSFERA *delimitado_por* ESTRATOSFERA»); *ubicado_en* (se utiliza cuando la ubicación de un objeto físico es esencial en su descripción, es decir, sin esa característica el concepto perdería su identidad, por ejemplo, «VIADUCTO *ubicado_en* DEPRESIÓN GEOGRÁFICA»); y *tiene_lugar_en* (une un proceso con su dimensión temporal, por ejemplo, «PLEAMAR SOSTICIAL MÁXIMA *tiene_lugar_en* SOLSTICIO», o espacial, por ejemplo, «TIFÓN *tiene_lugar_en* OCEANO PACIFICO»).

- *Resultado_de*: Relación no jerárquica que une entidades o procesos con los procesos de los que resultan, por ejemplo, «NIEVE *resultado_de* PROCESO DE BERGERON».
- *Causa*: Relación no jerárquica que une entidades con los procesos que las provocan, como «VIENTO *causa* EROSIÓN EÓLICA».
- *Afecta_a*: Esta relación no jerárquica une entidades o procesos que causan algún tipo de cambio en otra entidad o proceso sin producir un resultado final, por ejemplo, «EROSIÓN ANTRÓPICA *afecta_a* PLAYA». Actualmente se estudia incluir subtipos de esta relación (como *retarda* o *erosiona*).
- *Tiene_función*: Se trata de una relación no jerárquica que une una entidad o un proceso con su función («REVEGETACIÓN *tiene_función* PROTECCIÓN DEL SUELO»). No solo está limitado a las entidades y procesos artificiales, ya que el ser humano puede utilizar entidades y procesos naturales para sus propios fines («GUANO *tiene_función* FERTILIZACIÓN»). EcoLexicon incluye cuatro subtipos de esta relación en su inventario: *se_hace_con* para unir exclusivamente un instrumento con el proceso que lleva a cabo («CONDENSACIÓN *se_hace_con* CÁMARA DE NEBLINA») o la entidad que crea («ECOGRAMA *se_hace_con* ECOSONDA»); *estudia* para unir una disciplina con su objeto de estudio («ZOOLOGÍA *estudia* FAUNA»); *mide* para instrumentos con una función de medición («SALINÓMETRO *mide* SALINIDAD»); y *representa* para gráficos y mapas («MAREOGRAMA *representa* NIVEL DEL MAR»).
- *Atributo_de*: Esta relación une una propiedad con una entidad o un proceso del que es característico, por ejemplo, «IMPERMEABLE *atributo_de* ACUÍFERO CONFINADO».

Las proposiciones conceptuales se presentan al usuario de manera visual en forma de redes en las cuales los conceptos son nodos unidos entre sí por flechas que indican las relaciones. Dada la amplitud del dominio medioambiental, así como su multidimensionalidad inherente, y el gran número de proposiciones conceptuales representadas en EcoLexicon, se ofrece al usuario distintas opciones de visualización para evitar la sobrecarga de información (León-Araúz y Faber, 2010). Concretamente, el usuario puede filtrar por tipo de relación, obtener diferentes modos de visualización de la red conceptual (modo de red, modo de árbol y modo

FIGURA 3

Red de AGUA en el dominio de la ingeniería civil

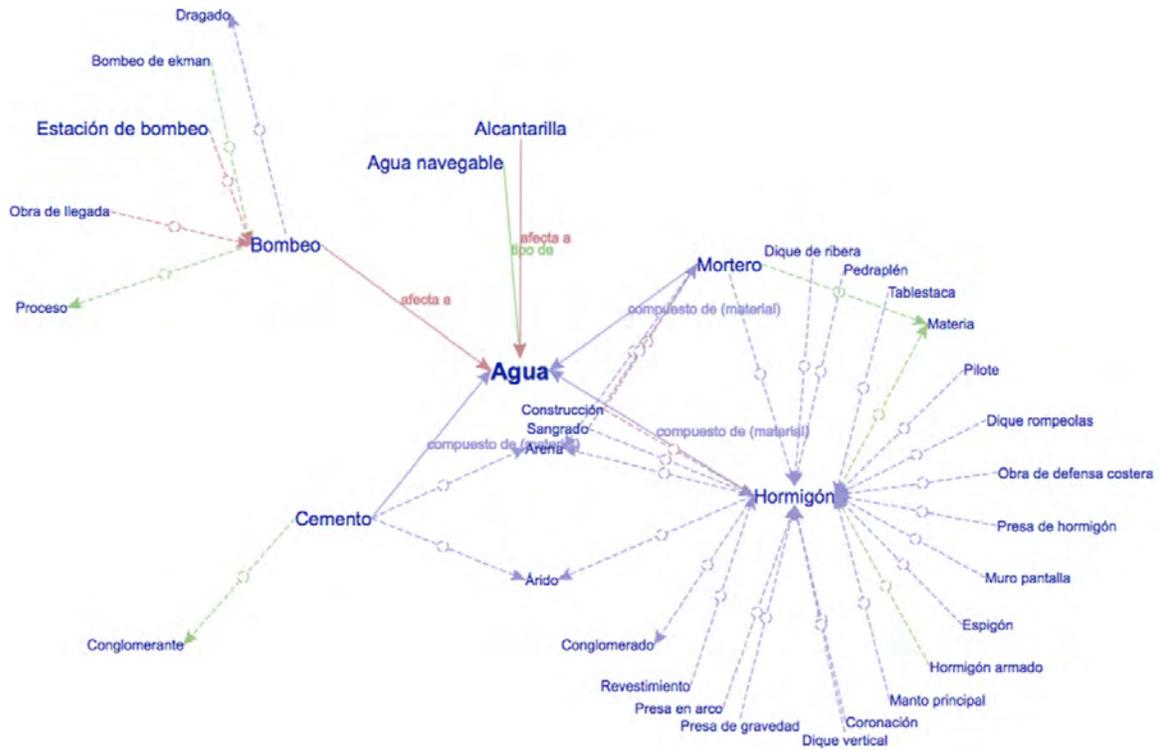


FIGURA 4

Modo de árbol de NUBE

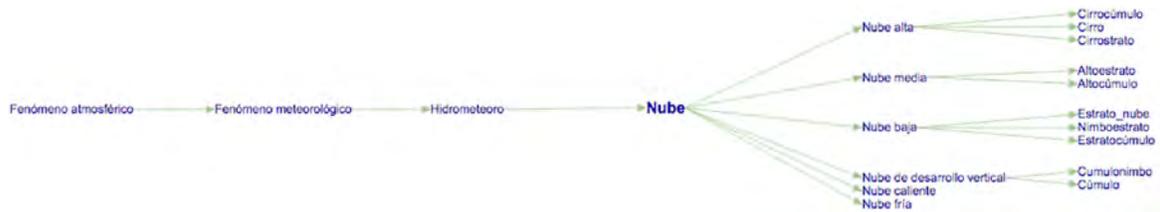


FIGURA 5

Modo de ruta entre AZUFRE y FLORA



3. EcoLexicon frente a otros recursos terminológicos sobre el medio ambiente

Además de EcoLexicon, existen otros recursos terminológicos dedicados al medio ambiente, como GEMET, el *Diccionari de les ciències ambientals*, EnvO y DiCoEnviro.

GEMET (*General Multilingual Environmental Thesaurus*) (www.eionet.europa.eu/gemet) es un tesoro general multilingüe sobre el medio ambiente desarrollado por la Agencia Europea de Medio Ambiente y destinado a la indización, búsqueda y gestión de la información. Para acceder a la información, es necesario elegir una de las 37 lenguas incluidas. A continuación, se pueden utilizar los índices alfabético y temático o el motor de búsqueda, que proporciona una lista con los términos que contienen la secuencia léxica introducida. Al elegir el término en cuestión, se presentan los equivalentes en todas las lenguas (normalmente, uno por idioma), su definición únicamente en inglés y términos relacionados (hiperónimos, hipónimos y otros términos relacionados cuyo vínculo no se precisa). Asimismo, se indica a qué subdominio medioambiental (*theme*) está asociado el concepto y a qué categoría general (*group*) pertenece. Por último, se muestran enlaces a las entradas del concepto en otros recursos (se distingue entre *has close match* y *has exact match*).

El *Diccionari de les ciències ambientals* (<https://cit.iec.cat/DCA>) es un diccionario electrónico de ciencias ambientales elaborado por el Instituto de Estudios Catalanes. Cuenta con un motor de búsqueda en el que se puede elegir el idioma de partida (catalán, español, inglés o francés) y el área temática. Además, se pueden realizar búsquedas truncadas (*comienza por*, *contiene*, etc.). Tras realizar la consulta deseada, aparece el lema en catalán, su definición en esta lengua y los equivalentes en las otras lenguas. Se incluye además el subdominio, el género de los términos y sus sinónimos, indicando cuál es el preferido.

Por su parte, EnvO (<http://environmentontology.org/>) es una ontología del medio ambiente desarrollada por Buttigieg y otros (2016). Se creó principalmente para tareas de anotación semántica. Aunque puede consultarse en línea, su funcionamiento puede resultar poco intuitivo para un usuario medio. Además, tanto la interfaz como los términos de la ontología aparecen únicamente en inglés. Tras introducir la consulta en el motor de búsqueda, se obtiene una lista de las entradas que coinciden con nuestra búsqueda o que guardan alguna relación, sin especificar cuál. En cada entrada se ofrece la definición, los sinónimos, los metadatos, así como la jerarquía de categorías conceptuales (p. ej., *power plant* es una ENTIDAD > ... > EDIFICIO INDUSTRIAL) y la red en la que se enmarca el concepto (p. ej., *geothermal power plant* y *nuclear power plant* son tipos de *power plant*). Las relaciones conceptuales incluidas en el recurso proceden de la Basic Formal Ontology (<http://basic-formal-ontology.org/>) y de la Relations Ontology (<http://www.obofoundry.org/ontology/ro.html>).

DiCoEnviro (*Dictionnaire fondamental de l'environnement*) (<http://olst.ling.umontreal.ca/dicoenviro/>) (L'Homme, 2012) es una base de datos terminológica sobre el medio ambiente en francés, inglés, español, portugués e italiano desarrollada en la Universidad de Montreal. Se

accede a la información mediante un índice alfabético o un motor de búsqueda, en el que se pueden combinar distintas opciones. En cada entrada se muestra la categoría gramatical del término y su definición, basada en la estructura argumental del concepto en forma de roles semánticos. Esta información se enriquece con sinónimos del término, contextos (muchos de ellos anotados sintáctica y semánticamente), equivalentes en las otras lenguas e imágenes, además del estado de desarrollo de la entrada y sus metadatos. Se describen igualmente relaciones léxicas (p. ej., términos relacionados y derivados, argumentos, etc.) basadas en las funciones léxicas de la lexicología explicativa y combinatoria (Mel'čuk y otros, 1995). Es posible obtener una representación visual de dichas relaciones mediante el módulo NeoVisual (L'Homme y otros, 2018). Por último, se ha comenzado recientemente a dotar al recurso de una estructura de marcos basada en FrameNet (Fillmore y otros, 2003) para la versión inglesa y francesa de DiCoEnviro (L'Homme, 2016). Para ello, los marcos relevantes de FrameNet se importan al recurso con adaptaciones de distinto calibre según el caso, como la sustitución de los elementos del marco por los roles semánticos de DiCoEnviro y la creación de nuevos marcos.

Los recursos arriba descritos pueden considerarse complementarios a EcoLexicon, pues algunos de ellos ofrecen lenguas o funcionalidades que este no incluye. Por su parte, EcoLexicon destaca en algunos aspectos como la flexibilidad de su motor de búsqueda, que muestra los resultados al usuario en cualquiera de las lenguas a medida que se teclea la búsqueda. Asimismo, se ofrecen diferentes puntos de acceso y modos de visualizar la información, algunos de ellos solo presentes en EcoLexicon, como el sistema conceptual a modo de red con relaciones tanto jerárquicas como no jerárquicas o la visualización de la ruta más corta entre dos conceptos. A ello se suman definiciones, una inclusión exhaustiva de variantes denominativas, información gramatical y recursos audiovisuales. Asimismo, con la intención de reutilizar la información de otros recursos en la construcción y desarrollo de EcoLexicon se ha decidido publicarlo en la Linguistic Linked Open Data Cloud (LLOD cloud) enlazado a otros recursos de temática medioambiental (León-Araúz y otros, 2011a, 2011b). De esta forma, se mantiene cada recurso por separado, pero se conectan las entradas, de modo que la información consignada en estas sea complementaria entre sí.

La iniciativa del Linked Data tiene como objetivo la creación de un espacio de información compartida a través de la publicación y conexión de recursos estructurados en la Web Semántica (Bizer y otros, 2009). Con el fin de especificar las relaciones semánticas entre diferentes conjuntos de datos, la BCT se convirtió en una ontología en formato RDF y a continuación se definieron las reglas del algoritmo que establecería la equivalencia semántica entre las entradas de EcoLexicon y las de otros recursos de manera semiautomática, empezando por la DBpedia, núcleo de la LLOD. De esta forma, EcoLexicon podrá ofrecerse en el futuro en tres modalidades: (1) la aplicación web, según puede consultarse actualmente; (2) otra aplicación web (EcoLexicon-LD) con la versión *linked data* de EcoLexicon, destinada a usuarios humanos; y (3) un punto de acceso SPARQL con el fin de que otros recursos puedan enlazarse con EcoLexicon y que los datos de EcoLexicon se puedan utilizar en aplicaciones externas.

El algoritmo de enlazado permite semiautomatizar el proceso explotando la semántica contenida en cada conjunto de datos. Las categorías de datos de EcoLexicon utilizadas en el algoritmo son todas aquellas relacionadas con las variantes lingüísticas, las equivalencias multilingües y las relaciones semánticas. El contenido de estas categorías se compara con el contenido de todas las propiedades que contienen texto en la DBpedia. Por ejemplo, el primer paso en el algoritmo de enlazado consiste en comparar todas las variantes en inglés de una misma entrada de EcoLexicon con el contenido de la propiedad *rdfs:label* de DBpedia. Debido a la polisemia, los términos de EcoLexicon pueden coincidir con varias entradas de la DBpedia y por lo tanto llevar a correspondencias incorrectas. Por ejemplo, la cadena *breakwater* coincidiría con la entrada de la obra costera, pero también con un grupo de música que posee la misma denominación. En ese caso, se utilizan los equivalentes multilingües como primer factor desambiguador. No obstante, la polisemia también puede ocurrir a nivel interlingüístico. En esos casos, o cuando no existen en EcoLexicon o en la DBpedia equivalencias multilingües suficientes, se recurre a la información contextual. Si cualquier término relacionado a través de los dominios contextuales con el término de búsqueda aparece en cualquier propiedad textual de la DBpedia (p. ej. *rdfs:comment*; *dbpedia-owl:abstract*, etc.), se considera que los conceptos son equivalentes.

El último paso, todavía en curso, consiste en validar y evaluar la fiabilidad de cada enlace a través de EcoLexicon-LD (figura 6), la versión *linked data* de EcoLexicon. Se trata de que múltiples usuarios voluntarios puedan registrarse en esta plataforma y de que cada enlace sea validado por más de un usuario. De este modo se desarrollará un protocolo de validación del que además podrán extraerse nuevas conclusiones de cara a la mejora del algoritmo y al enlazado de otros recursos asociados.

FIGURA 6

EcoLexicon-LD

The screenshot displays the EcoLexicon-LD interface for the concept 'fan'. At the top, it shows the concept name 'fan' and its URI: <http://manila.ugr.es/r/ecolexicon#concept907>. Below this, there is a section titled 'linked to' which lists several external resources with associated scores: <http://dbpedia.org/page/Fan> (62.8), http://dbpedia.org/resource/Alluvial_fan (70.0), <http://www.freebase.com/m/01k1qvw> (99.0), and <http://www.merriam-webster.com/dictionary/alluvial%20fan> (90.0). There is an 'Add link' button below the list. The 'definitions' section provides two definitions: 'fan-shaped sediment deposit that forms where rapidly flowing water enters a relatively open flat area.' (with a US flag) and 'formación geológica de forma triangular o en abanico constituida por el material detrítico sedimentado que transporta un río.' (with a ES flag). The 'terms' section lists 'abanico' (ES), 'fan' (US), and 'Schuttfächer' (DE). The 'relations' section shows a 'type of' relation to 'geological formation'.

4. Últimos avances en EcoLexicon

4.1. El EcoLexicon English Corpus y la EcoLexicon Semantic Sketch Grammar

El EcoLexicon English Corpus (EEC) (León-Araúz y otros, 2018) es un corpus especializado compilado por el grupo LexiCon durante el desarrollo de EcoLexicon. Contiene 23,1 millones de palabras y está compuesto por textos contemporáneos procedentes del dominio medioambiental. Se puede consultar desde la BCT, en la pestaña *Buscar concordancias* (figura 1).

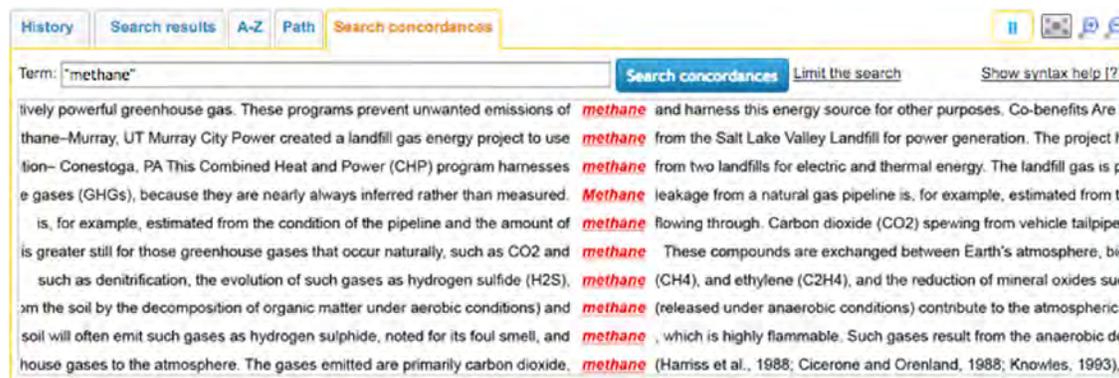
Cada uno de los textos del corpus está etiquetado con una serie de metadatos en XML de los cuales algunos están basados en el esquema Dublin Core y otros se han incluido según las necesidades de la BCT. Las etiquetas proporcionadas por los metadatos se basan en los siguientes parámetros:

- Dominio: El EEC abarca todos los dominios y subdominios de las ciencias ambientales (ej., biología, meteorología, ecología, ingeniería ambiental, derecho ambiental, etc.).
- Usuario: El EEC incluye textos destinados a tres tipos de usuarios según su nivel de especialización: experto, semiexperto, público general.
- Variante geográfica: El EEC incluye textos en inglés americano, británico y europeo.
- Género: El EEC cubre una gran variedad de géneros textuales: artículos científicos, libros, páginas web, obras lexicográficas, etc.
- Editor: El EEC distingue entre textos editados por científicos/investigadores, empresas, organismos gubernamentales, etc.
- Año: El EEC incluye textos desde 1973 a 2016.
- País: Los textos del EEC están etiquetados según el país de publicación.

Los metadatos permiten a los usuarios restringir sus consultas según los factores pragmáticos anteriores, como los dominios contextuales o el tipo de lector al que vaya dirigido cada texto. De este modo, los usuarios pueden comparar el uso de un mismo término en diferentes contextos. Por ejemplo, la figura 7 (en la página siguiente) muestra las concordancias de *methane* en textos de ingeniería ambiental.

El EEC también está disponible de forma gratuita en Sketch Engine Open Corpora (<https://the.sketchengine.co.uk/open/>), de manera que los usuarios que estén interesados en consultar el corpus puedan hacerlo a través de todas sus funcionalidades aunque no dispongan de una suscripción con Sketch Engine (Kilgarriff y otros, 2014).

El EEC está etiquetado con el Penn Treebank tagset (versión 3.3 de TreeTagger), que permite realizar consultas complejas en CQL (Corpus Query Language) (Schulze y Christ, 1996). El corpus se ha procesado incluyendo también la gramática de *sketches* en inglés ofrecida por

FIGURA 7Concordancias de *methane* en Ingeniería Ambiental

defecto en Sketch Engine para la compilación de los *word sketches* en conjunción con la EcoLexicon Semantic Sketch Grammar (León-Araúz y otros, 2016; León-Araúz y San Martín, 2018), una nueva gramática creada con el objetivo de ofrecer *word sketches* semánticos (no solo colocacionales) y extraer con ellos varias de las relaciones semánticas más utilizadas en terminología, como son la genérico-específica, la partitiva y las de lugar, causa y función. La EcoLexicon Semantic Sketch Grammar está basada en patrones de conocimiento, que son patrones lingüísticos y paralingüísticos en textos reales que expresan relaciones semánticas concretas (Meyer, 2001) (por ejemplo, los patrones “X es un tipo de Y” o “clases de X como Y” expresan una relación genérico-específica entre X e Y). Gracias a esta gramática, los usuarios pueden acceder a *word sketches* como los que muestra la figura 8 (en la página siguiente).

Por último, cabe destacar que EcoLexicon también dispone de un corpus español accesible por el momento solo desde la interfaz de la BCT. Sigue en fase de compilación y en el futuro también será accesible a través de Sketch Engine, para lo que también se está desarrollando una gramática en español para la identificación de *word sketches* semánticos.

4.2. Fraseología

EcoLexicon cuenta además con un módulo fraseológico que incluye colocaciones verbales. Se decidió comenzar por las colocaciones verbales porque, a pesar de tratarse de una categoría central del discurso (Fellbaum, 1990), existen aún pocas herramientas terminológicas que incorporen verbos. La TBM y EcoLexicon adoptan un enfoque amplio del concepto de *colocación*. Así, por *colocaciones verbales* entendemos combinaciones frecuentes de dos o más unidades léxicas formadas por un sustantivo + verbo o verbo + sustantivo en las que el colocativo (el verbo) está limitado por el significado de la base (el sustantivo), pero a su vez el verbo restringe los tipos de nombres con los que se puede combinar (Buendía-Castro, 2013: 115). Por ejemplo, en la colocación ‘el fuego arde’, el predicado ARDER admite únicamente ar-

gumentos que puedan estar literalmente en combustión (ej., el fuego), y el argumento ‘fuego’ necesita un verbo que se refiera a un proceso de combustión (ej., ARDER) (Montero Martínez y Buendía-Castro, 2017).

FIGURA 8

Word sketches extraídos mediante la EcoLexicon Semantic Sketch Grammar

"bacterium" is a type of...			"bacterium" is the generic of...		
	1,007	0.12		1,028	0.12
organism	158	10.00	coli	17	8.94
microorganism	88	10.92	plant	14	6.85
micro-organism	28	9.64	Pseudomonas	10	8.24
agent	18	8.09	Escherichia	10	8.22
decomposer	15	8.83	fungus	9	7.60

"rock" has part...			"rock" is part of...		
	3,029	0.09		2,055	0.06
mineral	213	10.54	crust	44	9.09
quartz	65	9.17	soil	34	7.97
fragment	47	8.79	belt	27	8.52
feldspar	45	8.79	continent	23	8.30
plagioclase	41	8.67	part	22	7.96

"volcano" is located at...			"volcano" is the location of...		
	318	0.04		71	0.01
plate	17	10.11	cone	7	11.10
island	14	9.42	ocean	3	8.23
boundary	11	9.38	type	3	6.74
Pacific	8	8.71	area	3	6.59
margin	7	8.87	precursor	2	9.66

"tsunami" is the cause of...			"tsunami" caused by...		
	196	0.04		1,057	0.20
damage	18	7.54	earthquake	177	11.31
destruction	12	8.74	landslide	68	10.73
erosion	7	6.70	eruption	36	9.34
devastation	6	9.08	water	33	7.70
death	6	6.67	movement	23	8.65

"energy" has function...			"energy" is the function of...		
	2,151	0.03		999	0.02
water	57	8.71	fuel	23	8.96
produce	41	8.83	carbon	14	8.12
make	33	8.45	biomass	13	8.44
process	22	7.99	waste	13	8.20
electricity	21	8.17	light	12	8.28

Los verbos se clasifican en EcoLexicon según su significado. En consonancia con esta perspectiva y de acuerdo con las premisas del modelo de la Gramática-Léxica (Faber y Mairal Usón, 1999), los verbos se organizan en dominios léxicos (según su significado nuclear) y en subdominios (según su significado específico). La figura 9 muestra un extracto de la entrada fraseológica de *incendio*. Esta dimensión de *incendio* activa en concreto el significado nuclear de *to cause sth/sb to change for the worse* con el significado nuclear de CAMBIO.

FIGURA 9

Extracto de la entrada de *incendio* en EcoLexicon

Información del término	
Término:	incendio
Idioma:	español
Tipo de término:	término principal
Contexto:	ev-incendio1-jose.txt
Categoría gramatical:	nombre común
Género:	masculino

Ver concordancias

Phraseological section

ACTION

to produce fire arder2

CHANGE

to cause sth/sb to change for the worse afectar arder1 arrasar
asolar calcinar
castigar dañar
demoler derribar
derruir derrumbar
destrozar destruir

Entrada fraseologica

Para averiguar cuál es la especificación de la dimensión activada, se analizaron los argumentos verbales y se los asignó a una categoría semántica. Según nuestro enfoque, las categorías semánticas son generalizaciones de un conjunto de términos con un comportamiento semántico y sintáctico similar (Buendía-Castro, 2013; Sánchez-Cárdenas, 2010). Dicho de otro modo, cuando diversas realizaciones lingüísticas comparten el mismo tipo de argumento, se considera que tienen un significado parecido, que poseen la misma estructura conceptual subyacente, que tienen un comportamiento semántico y sintáctico similar y que pueden pertenecer a la misma categoría semántica. De igual modo, cuando distintos verbos seleccionan el mismo tipo de argumentos que activan las mismas categorías semánticas, se prevé que activarán el mismo tipo de significado. Como se muestra en la figura 9, los verbos que comparten la dimensión de significado “TO CAUSE STH/SB TO CHANGE FOR THE WORSE” son *afectar*, *arder*, *arrasar*, *asolar*, *calcinar*, *castigar*, *dañar*, *demoler*, *derribar*, *derruir*, *derrumbar*, *destrozar* y *destruir*. Todos estos verbos tienen como primer argumento la categoría semántica de *DESASTRE NATURAL*, y como segundo argumento, *CONSTRUCCIÓN*, *ÁREA GEOGRÁFICA* O *SER HUMANO*. Al hacer clic en cada verbo, el usuario accede a ejemplos de uso, así como a una sección de notas con información sobre restricciones de significado. La extracción del corpus de esta información se realiza mediante la identificación semiautomática en el corpus de unidades lingüísticas que constituyen triples (*edificio-arrasar-incendio*); para ello se utilizan herramientas informáticas tales como el MWEtoolkit (Sánchez-Cárdenas y Ramisch, 2019) y Sketch Engine.

Por otro lado, se está remodelando actualmente el módulo fraseológico de EcoLexicon para incorporar compuestos nominales (CN) (p. ej., *recurso eólico*), unas unidades fraseológicas que reciben un tratamiento poco sistemático en los recursos terminográficos, a pesar de su eleva-

da frecuencia en los textos especializados. En español, se trata de expresiones cuyo núcleo es un sustantivo seguido de uno o varios modificadores (adjetivos o sintagmas preposicionales). Los CN se caracterizan por contener proposiciones subyacentes, que resultan patentes en los procesos de formación de términos que señala Levi (1978): supresión del predicado (p. ej., *central eléctrica*, en lugar de *central que produce electricidad*) y nominalización del predicado (p. ej., *evaluación del impacto ambiental* en lugar de *el impacto ambiental se evalúa*).

Estos predicados subyacentes se entienden a menudo como la relación semántica entre los formantes del CN (p. ej., *central eléctrica* > *central tiene_función electricidad*), cuya omisión constituye una de las principales dificultades de los CN. Así pues, se emplean diferentes métodos para clarificar el significado de los CN, como los inventarios de relaciones semánticas (p. ej., *función, resultado_de*). Estas se pueden refinar por medio de paráfrasis verbales, como *excitar* o *magnetizar*, en el caso de *generador de imanes permanentes*, especificando así la relación entre el núcleo (*generador*) y el modificador (*imanes permanentes*) (Nakov y Hearst, 2006; Cabezas-García y Faber, 2017a).

Por otra parte, hemos combinado recientemente el uso de patrones de conocimiento, paráfrasis y anotación semántica de los argumentos con categorías conceptuales (Cabezas-García y León-Araúz, 2018). Es decir, para *erosión de la línea de costa*, se pueden extraer líneas de concordancia del tipo *la erosión afecta a la línea de costa* o *la línea de costa sufre la erosión* por medio de búsquedas con patrones de conocimiento. Para identificar la relación semántica implícita se utilizan también paráfrasis verbales (p. ej., *la erosión degrada la línea de costa*, *la línea de costa retrocede debido a la erosión*) y se anotan los formantes del CN con categorías conceptuales (p. ej., *erosión de la línea de costa* se puede anotar como PROCESO + ACCIDENTE GEOGRÁFICO) (Cabezas-García y San Martín, 2017).

Por lo tanto, la nueva sección de CN del módulo fraseológico de EcoLexicon proporcionará información sobre la formación sintáctico-semántica de estos términos compuestos. Asimismo, los usuarios podrán encontrar soluciones a los problemas de traducción que a menudo plantean estas unidades, especialmente en el caso de los CN extensos (ej., *generador de inducción de rotor bobinado*). Es de esperar que el uso de patrones de conocimiento, paráfrasis y categorías conceptuales permita desarrollar un procedimiento de inferencia de relaciones semánticas (Cabezas-García y León-Araúz, 2018) que facilite la organización conceptual de los CN (Cabezas-García y Faber, 2017b), así como la semiautomatización de las estrategias y la elaboración de pautas de traducción. Asimismo, la variación característica de los CN también aparecerá reflejada en esta sección (Cabezas-García y Faber, 2017c).

El módulo fraseológico de EcoLexicon constituye, en definitiva, un continuo entre el submódulo verbal y el submódulo de CN, pues muchas colocaciones cuentan con un equivalente nominal en forma de CN (ej., *el volumen de sedimentos se conserva* y *conservación del volumen de sedimentos*). Por ello, ambas secciones se vincularán por medio de paráfrasis verbales y de la anotación semántica de los argumentos con categorías conceptuales (Cabezas-García

y San Martín, 2017; Cabezas-García y León-Araúz, 2018). Por último, este nuevo módulo fraseológico ofrecerá diferentes modos de visualización adaptados a las necesidades del usuario. Por ejemplo, habrá una vista de término, una visualización semántica (Cabezas-García y Gil-Berrozpe, 2018) o una representación bilingüe (inglés y español). De este modo, los usuarios podrán encontrar información lingüística y semántica de las unidades fraseológicas que les ayudará a producir textos idiomáticos en el dominio del medioambiente.

4.3. Definiciones terminológicas flexibles

Otra de las novedades en EcoLexicon consiste en que sus definiciones terminológicas se están adaptando con arreglo al enfoque de la definición terminológica flexible (EDTF) (San Martín, 2016), que resulta de la aplicación de los principios de la lingüística cognitiva (Lakoff, 1987; Croft y Cruse, 2004; Evans y Green, 2006; entre otros) y de la TBM a las definiciones. Dicho enfoque puede resumirse en cinco principios.

1. *Las definiciones deben reflejar la organización de los conceptos en la mente:* En la mayoría de recursos terminológicos, la redacción de definiciones se ha basado tradicionalmente en la delimitación de las características necesarias y suficientes del concepto en cuestión. Es decir, en la delimitación del conocimiento llamado *definicional* frente al conocimiento enciclopédico, que sería el resto de conocimiento acerca del concepto. Sin embargo, cualquier terminólogo puede atestiguar que resulta difícil, si no imposible, determinar el conocimiento definicional para la mayoría de conceptos. Basta con observar la divergencia en las definiciones del mismo concepto en distintos recursos para darse cuenta de ello. Esto es así, entre otros motivos, porque los conceptos presentan efectos de prototipicidad (Rosch, 1975), es decir, se asemejan más bien a categorías que codifican un análisis estadístico de las características que sus miembros suelen poseer (las llamadas características prototípicas) (Laurence y Margolis, 1999: 27). A diferencia de lo que ocurre en el enfoque clásico con las características suficientes y necesarias, no es preciso que los miembros de la categoría posean todas las características prototípicas para ser considerados miembros. Por ejemplo, una característica prototípica de TSUNAMI es que ocurre en el mar; sin embargo, hay tsunamis que ocurren en lagos y no por ello dejan de pertenecer a esta categoría; son simplemente casos menos prototípicos de TSUNAMI. Por otro lado, las características prototípicas varían según el contexto en el que se active el concepto (entendiendo contexto como cualquier factor que afecte a la interpretación real de un signo o una expresión [Kecskes, 2014]). Por ejemplo, un COMBUSTIBLE produce energía, prototípicamente, a partir de una reacción química. Sin embargo, si el concepto se activa en un contexto temático de ingeniería nuclear, un COMBUSTIBLE produce energía, prototípicamente, mediante una reacción nuclear. El EDTF defiende que la definición terminológica debe reflejar la variación de los rasgos activados por los conceptos según el contexto para así satisfacer plenamente las necesidades del usuario. Por ejemplo, un concepto como METANO recibirá una definición distinta en un recurso terminológico sobre ingeniería energética (se destacará su papel como fuente de energía) o en ciencias atmosféricas (es un gas de efecto invernadero).

Así pues, se incluyen en la definición las características prototípicas relevantes, así como la descripción de los marcos en los que el concepto participa (San Martín y León-Araúz, 2013); por ejemplo, para la definición de *metano* en un recurso de ciencias atmosféricas será preciso integrar en la definición una explicación sobre el EFECTO INVERNADERO y el papel de METANO dentro de este para que el destinatario pueda satisfacer sus necesidades de información.

2. *Las definiciones codifican un presignificado*: Un presignificado está a medio camino entre el potencial semántico de un término y su significado en contexto. El potencial semántico de un término es la suma del contenido conceptual que puede activar en cualquier evento de uso (Allwood, 1999; Evans, 2009). Una definición no puede representar el potencial semántico porque se trata de una cantidad ingente de información que nunca se activa simultáneamente. Tampoco el significado es el objeto de la definición porque estrictamente las unidades léxicas solo poseen significado en contexto, es decir, cuando son utilizadas en situaciones reales. Así pues, el fin de la definición es representar un presignificado: un subconjunto del potencial semántico limitado por restricciones contextuales. El terminólogo impone a la definición dichas restricciones (por ejemplo, restricciones según el dominio de conocimiento, según la ideología, según la cultura, etc.) según dónde se insertará y a quién irá dirigida.

3. *Las definiciones tienen una dimensión contextual, ontológica y funcional*: La dimensión contextual concierne las restricciones contextuales aplicadas a la definición. En línea con la recontextualización de las redes conceptuales, la aplicación del EDTF en EcoLexicon se centra en representar mediante la definición cómo varían las características de un concepto dado según el dominio en el que se activan. No obstante, al seleccionar el contenido de una definición se debe tener en cuenta también la dimensión ontológica del concepto que se define (como entidad, evento o atributo y sus subtipos [cf. Seppälä, 2015]), y la dimensión funcional (los factores que afectan a la selección del contenido de la definición en función de los destinatarios y de sus necesidades [cf. Tarp, 2008]).

4. *Un concepto puede tener más de una definición*: Si bien es habitual crear más de una definición para términos polisémicos, no es una práctica habitual hacerlo para un término que varía contextualmente¹. Por ello, las definiciones terminológicas tienden a ser demasiado ge-

1 La variación contextual (también llamada vaguedad o indeterminación conceptual) es el fenómeno que ocurre cuando un concepto activa distintos rasgos con diferentes niveles de relevancia según el contexto de activación. Un ejemplo de variación contextual es *ozono*, que se conceptualiza como un desinfectante en el dominio del tratamiento de aguas y como un gas que se encuentra de manera abundante en la atmósfera en el dominio de la gestión de la calidad del aire, pero en ambos casos se trata de la misma sustancia y, por tanto, del mismo concepto (*ozono*). *Metano*, anteriormente mencionado, es otro ejemplo de variación contextual. Este fenómeno se opone a la polisemia, la cual ocurre cuando un término está asociado a más de un concepto. Por ejemplo, *planta* es polisémico porque designa varios conceptos como PLANTA_1 (ser vivo autótrofo y fotosintético) o PLANTA_2 (instalación industrial).

nerales o demasiado específicas. Para resolver este problema, el EDTF propone múltiples definiciones para el mismo concepto en recursos que abarquen amplias áreas de conocimiento. El número de definiciones para un mismo concepto lo determina el terminólogo de acuerdo con los datos provenientes del análisis de corpus y las necesidades de los usuarios potenciales del recurso. Este principio se está implementando en EcoLexicon mediante una definición general medioambiental acompañada de definiciones según los distintos subdominios del medio ambiente (según la jerarquía de dominios de EcoLexicon). La tabla 1 muestra un extracto de la definición múltiple de CLORO.

TABLA 1

Extracto de la definición múltiple de CLORO

GENERAL MEDIOAMBIENTAL

Elemento químico no metálico que pertenece a la familia de los halógenos y que se presenta como un gas amarillo verdoso a temperatura y presión estándares. Debido a su alta reactividad, solo se encuentra de forma natural en compuestos como el [...].

GESTIÓN DE LA CALIDAD DEL AIRE

Contaminante atmosférico emitido principalmente como clorofluorocarbonos, hidroclo-rofluorocarbonos, tetracloruro de carbono y metilcloroformo, que son sustancias que agotan la capa de ozono. Cuando estas sustancias alcanzan la [...].

QUÍMICA

Elemento halógeno no metálico con número atómico 17 que se presenta como un gas amarillo verdoso a temperatura y presión estándares. Solo se encuentra naturalmente en compuestos como el cloruro de sodio (sal común) en agua de mar y en la halita [...].

ABASTECIMIENTO Y TRATAMIENTO DE AGUAS

Desinfectante de agua que, debido a su poder oxidante, se agrega al agua, generalmente en un tanque de contacto, para matar o inactivar patógenos como parte del tratamiento de aguas residuales. El cloro se usa en forma pura o como hipoclorito [...].

5. *La redacción de la definición debe estar basada en corpus*: Dado que el medio privilegiado para la transmisión de conocimiento especializado son los textos escritos por expertos (Bourigault y Slodzian, 1999), la redacción de definiciones debe basarse principalmente en el análisis de corpus, mediante técnicas tales como la extracción de contextónimos (San Martín, 2016: 272-274) y de *word sketches* (tanto los originales como los semánticos).

4.4. Representación conceptual con imágenes

La TBM propone una descripción conceptual multimodal en la que la información contenida en las definiciones terminográficas se complementa con información visual para facilitar la comprensión de sistemas conceptuales complejos y dinámicos (Faber, 2009). Por esta razón, se compiló un corpus visual para enriquecer la descripción conceptual en la BCT. Hasta la fecha, las imágenes se seleccionan según las proposiciones conceptuales contenidas en la definición del concepto. Por ejemplo, la definición de BARRANCO según EcoLexicon es un *tipo_de*

ACCIDENTE GEOGRÁFICO que consiste en una depresión profunda en el terreno, provocada por corrientes de agua. La figura 10 es una fotografía en color que muestra un barranco con un río en el fondo, por lo que se puede incluir en la BCT en la entrada de BARRANCO por la relación conceptual *tipo_de*.

FIGURA 10

Imagen que representa BARRANCO *tipo_de* ACCIDENTE GEOGRÁFICO
(autor: Stab Shebs, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=4995549>)



La plantilla definicional permite seleccionar las imágenes de forma sistemática para cada entrada conceptual en la BCT, pero conlleva que cada imagen se considere un todo y se vincule a un solo concepto.

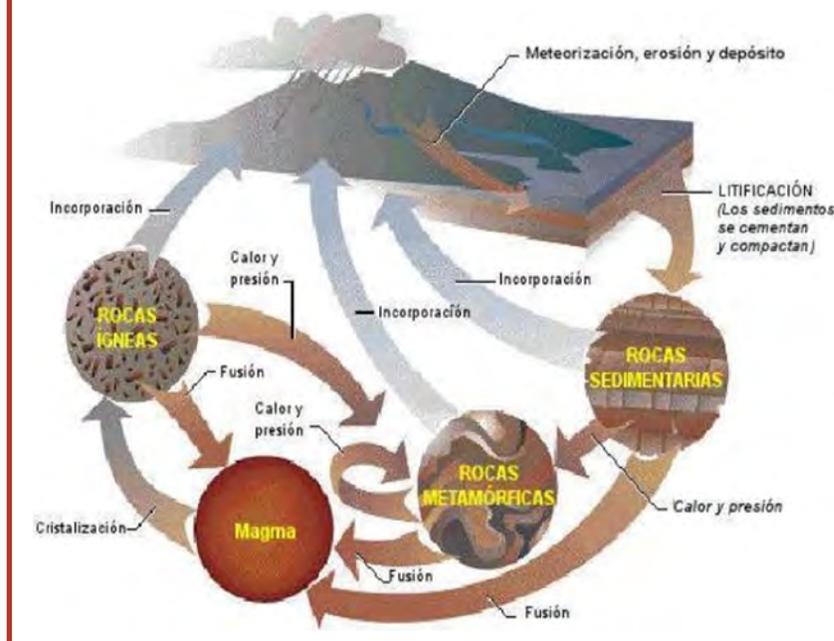
En investigaciones anteriores se estudió cómo el mismo concepto se puede (y se debería) representar a través de diferentes imágenes según la perspectiva o el contenido semántico (Faber y otros, 2007; Reimerink y otros, 2010). Sin embargo, la misma imagen puede servir igualmente para representar visualmente otras entradas conceptuales (por ejemplo, una entidad y el proceso por el que se ha formado, o un concepto y sus partes).

Al contrario de lo que sucede con las entidades físicas, como la de la figura 10, los procesos se describen sobre todo mediante relaciones meronímicas como *tiene_lugar_en*, *fase_de* y *resultado_de*. De ahí que las imágenes que muestran procesos suelen representar más de una sola relación conceptual. Por ejemplo, la figura 11 representa el CICLO GEOLÓGICO, un proceso extremadamente complejo con varias fases en un diagrama de flujo.

FIGURA 11

Imagen que representa el CICLO GEOLÓGICO

(fuente: <https://es.slideshare.net/Jesucitohq954018530/tema-05-ggciclos-en-geologia>)



Los conceptos LITIFICACIÓN, FUSIÓN, CRISTALIZACIÓN e INCORPORACIÓN *tienen_lugar* debajo de la superficie terrestre. Al mismo tiempo, son *fases_del* CICLO GEOLÓGICO. En la imagen también se representa la relación *resultado_de*, como por ejemplo MAGMA *resultado_de* FUSIÓN y ROCA SEDIMENTARIA *resultado_de* LITIFICACIÓN. Algunas de las partes de la imagen muestran una gran similitud con sus referentes en el mundo real como la montaña y la nube con lluvia. Otras entidades y procesos complejos llevan una etiqueta como ROCAS SEDIMENTARIAS, ROCAS METAMÓRFICAS, LITIFICACIÓN y CRISTALIZACIÓN. El uso de colores similares, pero diferentes, enfatiza la similitud y, al mismo tiempo, delimita conceptos similares o aquellos que comparten proximidad espacial. Por ejemplo, el MAGMA es de un rojo más intenso, mientras que los diferentes tipos de rocas se representan con diferentes tonos de marrón. El uso de los diferentes tonos de marrón y los dibujos que forman además permite ver en qué se distinguen los diferentes tipos de roca. Las flechas añaden dinamismo a la imagen al representar cómo algunas fases siguen a otras y cómo se afectan mutuamente.

Llamamos patrones de conocimiento visuales (PCV) a los rasgos morfológicos, como los descritos para la figura 11, que ayudan a la representación del conocimiento, tales como las flechas, el uso de colores y las etiquetas. Al describir una imagen, resulta imprescindible señalar la función que realiza cada uno, ya que los PCV pueden ser polisémicos. Por ejemplo, las flechas se pueden usar para añadir dinamismo a la imagen cuando señalan la dirección del movimiento, o bien los diferentes pasos de un proceso. Sin embargo, también se pueden

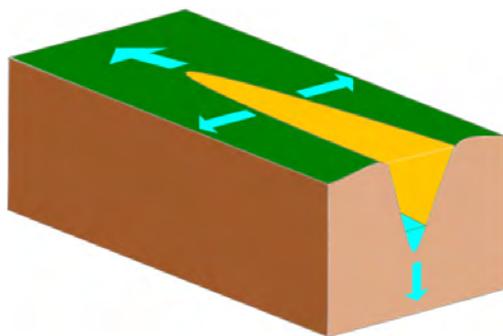
usar para señalar la ubicación de una entidad en la imagen o para vincular esa entidad con su etiqueta textual o con las partes de las que esté compuesta. Por último, el uso de los colores puede desempeñar funciones diferentes tales como aumentar el realismo de la imagen o realzar el contraste entre conceptos similares.

Con el objetivo de diseccionar las imágenes en función de todos los parámetros descritos anteriormente, se ha creado una aplicación específica: Manzanilla. Esta aplicación, desarrollada en el entorno de Camomile (<https://camomile.limsi.fr>) (Poignant y otros, 2016), permite anotar las imágenes en varias interfaces consecutivas. En la primera, se define el tipo de imagen (fotografía, dibujo o diagrama de flujo); en la segunda, se anotan los conceptos representados en la imagen; y en la tercera, las relaciones entre los conceptos presentes y los PCV que las representan y su función. En cada paso, se proponen opciones al anotador basadas en la información contenida en EcoLexicon. Tras anotar todas las imágenes del corpus visual, se guardarán en un módulo separado que se vinculará a EcoLexicon a través de las proposiciones conceptuales representadas en cada imagen. De esta manera, aumentará la coherencia interna de la BCT y se aliviará la carga de trabajo a la larga (Reimerink y otros, 2016; León-Araúz y Reimerink, 2016; Reimerink y León-Araúz, 2018).

Así pues, la figura 10 se mostrará al usuario final no solamente en la entrada conceptual BARRANCO, sino también cuando se indagara en la proposición conceptual BARRANCO *resultado_de* EROSIÓN FLUVIAL. En esta última, el usuario también se encontrará, por ejemplo, con la figura 12, que muestra la interrelación de un río con el proceso de EROSIÓN, que a su vez surgiría en búsquedas basadas en proposiciones conceptuales como EROSIÓN *tiene_lugar_en* RÍO.

FIGURA 12

Imagen que representa EROSIÓN *tiene_lugar_en* RÍO
(autor: Kayau, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=25267645>)



Imágenes complejas como la de la figura 12 se podrán vincular con un gran número de conceptos y proposiciones conceptuales en la base de datos. Además, las imágenes que representan procesos complejos interrelacionando muchos conceptos diferentes son especialmente útiles para entender un dominio especializado de forma eficaz.

Dada la complejidad y profundidad de nuestro sistema de anotación de imágenes, a día de hoy no es posible realizarla de forma automática. ImageNet (Deng y otros, 2009; Russakovsky y otros, 2015), por ejemplo, incluye sobre todo fotografías que no pueden representar conocimiento especializado complejo y la anotación automática a partir de los *synsets* de WordNet que se aplica no permite la anotación detallada que proponemos aquí. BabelNet (Navigli y Ponzetto, 2012) es otro ejemplo de una gran base de conocimiento que relaciona de forma automática imágenes de ImageNet y Wikipedia con su contenido. Sin embargo, no se especifican las relaciones semánticas contenidas en esas imágenes. En ciencias de la computación, la investigación en anotación automática de imágenes ha experimentado un progreso sustancial (Jeon y otros, 2003; Li y Wang, 2008), pero se centra principalmente en fotografías de objetos y en conceptos más bien generales sin tener en cuenta la interacción entre los elementos semánticos (Mei y otros, 2008). En la comunidad de las redes neuronales también se han centrado especialmente en el reconocimiento automático de objetos (Russakovsky y otros, 2015), pero tampoco abordan las relaciones entre diferentes entidades o procesos en una misma imagen. Para nuestro propósito, la relación entre las entidades y/o procesos y el contexto o el fondo en el que se encuentran o desarrollan es esencial para representar el conocimiento multidimensional complejo del dominio del medioambiente.

5. Conclusiones

EcoLexicon es un robusto recurso terminológico basado en una sólida base teórica. Aunque ya ha cumplido 15 años, continúa creciendo y adquiriendo nuevos dominios, conceptos, información lingüística e incluso nuevas funcionalidades. Por lo tanto, estamos ante un trabajo en continuo desarrollo, ya que la innovación es constante. Con el paso de los años, se ha convertido además en una plataforma para la creación de productos lingüísticos como el EcoLexicon English Corpus y la EcoLexicon Semantic Sketch Grammar y su gramática semántica o EcoLexiCAT (León-Araúz y otros, 2017; León-Araúz y Reimerink, 2018), una herramienta web de traducción asistida por ordenador de textos especializados del medio ambiente, que explota principalmente los datos contenidos en EcoLexicon.

Este artículo ha descrito brevemente algunos de los desarrollos más recientes en EcoLexicon, que son la aplicación práctica de avances teóricos innovadores. Cabe destacar que la mayor parte de ellos son directamente implementables en otros recursos terminográficos sin necesidad de mayor adaptación, por lo que constituyen propuestas de eficacia probada para la mejora de otros recursos terminográficos existentes.

La excelente calidad e indudable efectividad de EcoLexicon para tareas de comprensión y adquisición del lenguaje especializado se ha demostrado en estudios como el de Giacomini (2014). Por lo tanto, además de una fuente de innovaciones terminográficas, resulta una herramienta valiosa tanto para profesionales de la lengua y expertos en medio ambiente como para el público en general.

6. Bibliografía citada

ALLWOOD, Jens, 1999: "Semantics as meaning determination with semantic-epistemic operations" en Jens ALLWOOD y Peter GÄRDENFORS (eds.): *Cognitive Semantics. Meaning and Cognition*, Ámsterdam/Filadelfia: John Benjamins, 1-17.

BIZER, Christian, Tom HEATH y Tim BERBERS-LEE, 2009: "Linked data - the story so far", *International Journal on Semantic Web and Information Systems* 5, 1-22.

BOURIGAULT, Didier, y Monique SLODZIAN, 1999: "Pour une terminologie textuelle", *Terminologies Nouvelles* 19, 29-32.

BUENDÍA-CASTRO, Miriam, 2013: *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources*. Tesis de doctorado, Universidad de Granada.

BUTTIGIEG, Pier Luigi, y otros, 2016: "The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation", *Journal of Biomedical Semantics* 7 (1), 57.

CABEZAS-GARCÍA, Melania, y Pamela FABER, 2017a: "Exploring the Semantics of Multi-word Terms by Means of Paraphrases" en Miguel Ángel CANDEL-MORA y Chelo VARGAS-SIERRA (eds.): *Temas actuales de terminología y estudios sobre el léxico*, Granada: Comares, 193-217.

CABEZAS-GARCÍA, Melania, y Pamela FABER, 2017b: "A Semantic Approach to the Inclusion of Complex Nominals in English Terminographic Resources" en Ruslan MITKOV (ed.): *Computational and Corpus-Based Phraseology*, Berlín: Springer, 145-59.

CABEZAS-GARCÍA, Melania, y Pamela FABER, 2017c: "The role of micro-contexts in noun compound formation", *Neologica* 11, 101-118.

CABEZAS-GARCÍA, Melania, y Juan Carlos GIL-BERROZPE, 2018: "Semantic-based retrieval of complex nominals in terminographic resources" en *Proceedings of the 18th EURALEX International Congress*, Ljubljana: Euralex, 269-281.

CABEZAS-GARCÍA, Melania, y Pilar LEÓN-ARAÚZ, 2018: "Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study" en Nicoletta CALZOLARI y otros: *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki: European Language Resources Association (ELRA), 2511-2518.

CABEZAS-GARCÍA, Melania, y Antonio SAN MARTÍN, 2017: "Semantic annotation to characterize contextual variation in terminological noun compounds: a pilot study" en *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, Valencia: Association for Computational Linguistics, 108-113.

CROFT, William, y Alan CRUSE, 2004: *Cognitive linguistics*, Nueva York: Cambridge University Press.

DENG, Jia, Wei DONG, Richard SOCHER, Li-Jia LI, Kai LI y Li FEI FEI, 2009: "ImageNet: A Large-Scale Hierarchical Image Database", *IEEE Computer Vision and Pattern Recognition (CVPR)*, Miami, FL.

EVANS, Vyvyan, 2009: *How Words Mean: Lexical Concepts, Cognitive Models, and Meaning Construction*, Oxford: Oxford University Press.

EVANS, Vyvyan, y Melanie C. GREEN, 2006: *Cognitive Linguistics: an Introduction*, Edimburgo: Edinburgh University Press.

FABER, Pamela, 2009: "The cognitive shift in terminology and specialized translation", *MonTI. Monografías de Traducción e Interpretación* 1, 107-34.

FABER, Pamela, 2012: *A Cognitive Linguistics View of Terminology and Specialized Language*, Berlín, Boston: De Gruyter Mouton.

FABER, Pamela, 2015: "Frames as a framework for terminology" en Hendrik KOCKAERT y Frieda STEURS (eds.): *Handbook of Terminology*, Ámsterdam, Filadelfia: John Benjamins, 14-33.

FABER, Pamela, y Miriam BUENDÍA-CASTRO, 2014: "EcoLexicon" en Andrea ABEL y otros (eds.): *Proceedings of the XVI EURALEX International Congress*, Bolzano: EURALEX, 601-607.

FABER, Pamela, Pilar LEÓN-ARAÚZ, Juan Antonio PRIETO VELASCO y Arianne REIMERINK, 2007: "Linking Images and Words: the description of specialized concepts", *International Journal of Lexicography* 20 (1), 39-65.

FABER, Pamela, Pilar LEÓN-ARAÚZ y Juan Antonio PRIETO VELASCO, 2009: "Semantic Relations, Dynamism, and Terminological Knowledge Bases", *Current Issues in Language Studies* 1, 1-23.

FABER, Pamela, y Pilar LEÓN-ARAÚZ, 2010: "Dinamismo conceptual en las bases de conocimiento terminológico: el caso de EcoLexicon", *Íkala, revista de lenguaje y cultura* 15 (25), 75-100.

FABER, Pamela, Pilar LEÓN-ARAÚZ y Arianne REIMERINK, 2011: "Knowledge representation in EcoLexicon" en Noa TALAVÁN y otros (eds.): *Technological Innovation in the Teaching and Processing of LSPs: Proceedings of TISLID'10*, Madrid: Universidad Nacional de Educación a Distancia, 367-386.

FABER, Pamela, Pilar LEÓN-ARAÚZ y Arianne REIMERINK, 2016: "EcoLexicon : New Features and Challenges" en Ilan KERNERMAN y otros (eds.): *GLOBALLEX 2016: Lexicographic Resources for Human Language Technology in conjunction with LREC 2016*, Portorož: Globalex, 73-80.

FABER, Pamela, y Ricardo MAIRAL USÓN, 1999: *Constructing a Lexicon of English Verbs*, Berlín: Mouton de Gruyter.

FABER, Pamela, y otros, 2006: "Process-oriented terminology management in the domain of Coastal Engineering", *Terminology* 12 (2), 189-213.

FELLBAUM, Christiane, 1990: "English Verbs as a Semantic Net", *International Journal of Lexicography* 3 (4), 278-301.

FILLMORE, Charles J, 1982: "Frame Semantics" en LINGUISTIC SOCIETY OF KOREA: *Linguistics in the Morning Calm*, Seúl: Hanshin, 111-137.

FILLMORE, Charles J., y Sue ATKINS, 1992: "Toward a frame-based lexicon: The semantics of RISK and its neighbors" en Adrienne LEHRER y Eva KITTAY (eds.): *Frames, Fields and Contrasts*, Hillsdale: Lawrence Erlbaum Assoc, 75-102.

FILLMORE, Charles, Christopher JOHNSON y Miriam R. L. PETRUCK, 2003: "Background to FrameNet", *International Journal of Lexicography* 16 (3), 235-250.

GIACOMINI, Laura, 2014: "Testing user interaction with LSP e-lexicographic tools: A case study on active translation of environmental terms" en *Proceedings of the 12th Edition of the Konvens Conference*, German Society for Computational Linguistics: Hildesheim, 77-85.

JEON, Jiwoon, Victor P. LAVRENKO y Raghavan MANMATHA, 2003: "Automatic image annotation and retrieval using crossmedia relevance models" en *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, 119-126.

KECSKES, Istvan, 2014: *Intercultural Pragmatics*, Oxford, Nueva York: Oxford University Press.

KILGARRIFF, Adam, y otros, 2014: "The Sketch Engine: ten years on", *Lexicography* 1 (1), 7-36.

LAKOFF, George, 1987: *Women, fire, and dangerous things: what categories reveal about the mind*, Chicago: University of Chicago Press.

LAURENCE, Stephen, y Eric MARGOLIS, 1999: "Concepts and Cognitive Science" en Eric MARGOLIS y Stepehn LAURENCE (eds.): *Concepts: Core Readings*, Cambridge, MA: MIT Press, 3-81.

LEÓN-ARAÚZ, Pilar 2009: *Representación multidimensional del conocimiento especializado: el uso de marcos desde la macroestructura hasta la microestructura*. Tesis de doctorado, Universidad de Granada.

LEÓN-ARAÚZ, Pilar, y Pamela FABER, 2010: "Natural and contextual constraints for domain-specific relations" en Verginica BARBU MITITELU y otros (eds.): *Proceedings of the Workshop Semantic Relations. Theory and Applications*, La Valletta, 12-17.

LEÓN-ARAÚZ, Pilar, Pamela FABER y Pedro MAGAÑA REDONDO, 2011a: “Linking Domain-Specific Knowledge to Encyclopedic Knowledge: an Initial Approach to Linked Data” en *2nd Workshop on the Multilingual Semantic Web (The 10th International Semantic Web Conference)*, Bonn, 68-73.

LEÓN-ARAÚZ, Pilar, Pedro MAGAÑA REDONDO y Pamela FABER, 2011b: “Integrating Environment into the Linked Data Cloud” en W. PILLMAN, S. SCHADE y P. SMITS (eds.): *Proceedings of the 25th International Conference Environmental Informatics. EnviroInfo Ispra 2011*, Shaker Verlag, 370-379.

LEÓN-ARAÚZ, Pilar, Pamela FABER y Silvia MONTERO MARTÍNEZ, 2012: “Specialized language semantics” en Pamela FABER (ed.): *A Cognitive Linguistics View of Terminology and Specialized Language*, Berlín, Boston: De Gruyter Mouton, 95-175.

LEÓN-ARAÚZ, Pilar, y Arianne REIMERINK, 2016: “Evaluation of EcoLexicon Images” en Fahad KHAN y otros (eds.): *Joint Second Workshop on Language and Ontology & Terminology and Knowledge Structures (LangOnto2 + TermiKS)*, Portorož, 16-22.

LEÓN-ARAÚZ, Pilar, y Arianne REIMERINK, 2018: “Evaluating EcoLexiCAT: a Terminology-Enhanced CAT Tool” en Nicoletta CALZOLARI y otros (eds.): *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki: European Language Resources Association (ELRA), 2374-2381.

LEÓN-ARAÚZ, Pilar, Arianne REIMERINK y Alejandro GARCÍA-ARAGÓN, 2013: “Dynamism and context in specialized knowledge”, *Terminology* 19 (1), 31-61.

LEÓN-ARAÚZ, Pilar, Arianne REIMERINK y Pamela FABER, 2017: “EcoLexiCAT: a Terminology-enhanced Translation Tool for Texts on the Environment” en Iztok KOSEM y otros (eds.): *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, Leiden: Lexical Computing, 321-341.

LEÓN-ARAÚZ, Pilar, y Antonio SAN MARTÍN, 2012: “Multidimensional Categorization in Terminological Definitions” en R. V. FJELD y J. M. TORJUSEN: *Proceedings of the 15th EURALEX International Congress*, Oslo: EURALEX, 578-84.

LEÓN-ARAÚZ, Pilar, y Antonio SAN MARTÍN, 2018: “The EcoLexicon Semantic Sketch Grammar: from Knowledge Patterns to Word Sketches” en Ilan KERNEMAN y Simon KREK (eds.): *Proceedings of the LREC 2018 Workshop “Globalex 2018 – Lexicography & WordNets”*, Miyazaki: Globalex, 94-99.

LEÓN-ARAÚZ, Pilar, Antonio SAN MARTÍN y Pamela FABER, 2016: “Pattern-based Word Sketches for the Extraction of Semantic Relations” en *Proceedings of the 5th International Workshop on Computational Terminology*, Osaka, 73-82.

LEÓN-ARAÚZ, Pilar, Antonio SAN MARTÍN y Arianne REIMERINK, 2018: "The EcoLexicon English Corpus as an open corpus in Sketch Engine" en *Proceedings of the 18th EURALEX International Congress*, Ljubljana: Euralex.

LEVI, Judith, 1978: *The Syntax and Semantics of Complex Nominals*, Nueva York: Academic Press.

LI, Jia, y James Z. WANG, 2008: "Real-Time Computerized Annotation of Pictures", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (6), 985-1002.

L'HOMME, Marie-Claude, 2012: "Using ECL (Explanatory Combinatorial Lexicology) to discover the lexical structure of specialized subject fields" en Jurij APRESJAN y otros (eds.): *Words, Meanings and other Interesting Things. A Festschrift in Honour of the 80th Anniversary of Professor Igor Alexandrovic Mel'cuk*, Moscú: RCK, 378-390.

L'HOMME, Marie-Claude, 2016: "Terminologie de l'environnement et Sémantique des cadres" en *Congrès mondial de linguistique française (CMLF 2016)*, Tours.

L'HOMME, Marie-Claude, Benoît ROBICHAUD y Nathalie PRÉVIL, 2018: "Browsing the Terminological Structure of a Specialized Domain : A Method Based on Lexical Functions and their Classification" en *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki: ELRA, 3079-3086.

MEI, Tao, Yong WANG, Xian-Sheng HUA, Shaogang GONG y Shipeng LI, 2008: "Coherent image annotation by learning semantic distance" en *IEEE Conference on Computer Vision and Pattern Recognition*, Piscataway, NJ, 1-8.

MEL'ČUK, Igor, André CLAS y Alain POLGUÈRE, 1995: *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve: Duculot.

MEYER, Ingrid, 2001: "Extracting Knowledge-Rich Contexts for Terminography" en Didier BOURIGAULT, Christian JACQUEMIN y Marie-Claude L'HOMME: *Recent Advances in Computational Terminology*, Ámsterdam/Filadelfia: John Benjamins, 279-302.

MONTERO MARTÍNEZ, Silvia, y Míriam BUENDÍA-CASTRO, 2017: "Clasificación semántica de colocaciones verbales para la adquisición y codificación de conocimiento experto", *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics* 30 (1), 240-72.

NAKOV, Preslav, y Marti A. HEARST, 2006: "Using Verbs to Characterize Noun-Noun Relations", *Artificial Intelligence Methodology Systems and Applications* 4183, 233-44.

NAVIGLI, Robert, y Simone Paolo PONZETTO, 2012: "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network", *Artificial Intelligence* 193, 217-250.

POIGNANT, Johann, y otros, 2016: "The CAMOMILE Collaborative Annotation Platform for Multi-modal, Multi-lingual and Multi-media Documents" en Nicoletta CALZOLARI y otros (eds.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož: European Language Resources Association (ELRA), 1421-1425.

REIMERINK, Arianne, Mercedes GARCÍA DE QUESADA y Silvia MONTERO MARTINEZ, 2010: "Contextual information in terminological knowledge bases: A multimodal approach", *Journal of Pragmatics* 42 (7), 1928-1950.

REIMERINK, Arianne, y Pilar LEÓN-ARAÚZ, 2018: "Manzanilla: An Image Annotation Tool for TKB Building" en Nicoletta CALZOLARI y otros (eds.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki: European Language Resources Association (ELRA), 1053-1059.

REIMERINK, Arianne, Pilar LEÓN-ARAÚZ y Pamela FABER, 2016: "Image selection and annotation for an environmental knowledge base", *Language Resources and Evaluation* 50 (3), 443-474.

ROSCH, Eleanor, 1975: "Cognitive representations of semantic categories", *Journal of Experimental Psychology* 104 (3), 192-233.

RUSSAKOVSKY, Olga, Jia DENG, Hao SU, Jonathan KRAUSE, Sanjeev SATHEESH, Sean MA, Zhiheng HUANG, Andrej KARPATHY, Aditya KHOSLA, Michael BERNSTEIN, Alexander C. BERG y Li FEI-FEI, 2015: "ImageNet Large Scale Visual Recognition Challenge", *International Journal of Computer Vision* 115, 211.

SAN MARTÍN, Antonio, 2016: *La representación de la variación contextual mediante definiciones terminológicas flexibles*. Tesis de doctorado, Universidad de Granada.

SAN MARTÍN, Antonio, y Pilar LEÓN-ARAÚZ, 2013: "Flexible terminological definitions and conceptual frames" en Selja SEPPÄLÄ y Alan RUTTENBERG (eds.): *International Workshop on Definitions in Ontologies (DO 2013)*, Montreal: Concordia University, 1-7.

SÁNCHEZ-CÁRDENAS, Beatriz, 2010: "Les restrictions sémantiques des arguments verbaux: une question de fréquence d'usage", *Synergies France* 6, 41-50.

SÁNCHEZ-CÁRDENAS, Beatriz, y Carlos RAMISCH, 2019: "Eliciting specialized frames from corpora using argument-structure extraction techniques", *Terminology* 25 (1).

SCHULZE, Bruno Maximilian, y Oliver CHRIST, 1996: *The CQP user's manual*, Stuttgart: Universität Stuttgart.

SEPPÄLÄ, Selja, 2015: "An ontological framework for modeling the contents of definitions", *Terminology* 21 (1), 23-50.

TARP, Sven, 2008: *Lexicography in the borderland between knowledge and non-knowledge general lexicographical theory with particular focus on learner's lexicography*, Tübingen: Max Niemeyer Verlag.