

Construcción de marcos semánticos multilingües de conceptos especializados

Construction of multilingual semantic
frameworks of specialized concepts

Beatriz Sánchez Cárdenas

Universidad de Granada
España

ONOMÁZEIN 68 (junio de 2025): 24-47

DOI: 10.7764/onomazein.68.02

ISSN: 0718-5758



Beatriz Sánchez Cárdenas: Departamento de Traducción e Interpretación, Grupo de investigación LexiCon, Universidad de Granada, España. ORCID: 0000-0002-1904-675X. | E-mail: bsc@ugr.es

Fecha de recepción: marzo de 2021

Fecha de aceptación: septiembre de 2021

Resumen

La terminología basada en marcos es un marco lingüístico cuya hipótesis principal es que los conceptos y eventos especializados se organizan conceptualmente en la mente del orador. Utiliza marcos semánticos especializados como representación estructurada de esta organización conceptual de los dominios. Los recursos lexicográficos basados en la terminología basada en marcos representan una rica fuente de información que va más allá de los inventarios de términos estáticos. Esos recursos pueden ser útiles para los expertos de los dominios y los traductores, que necesitan tener acceso a las estructuras conceptuales de los dominios y a las correspondientes realizaciones lingüísticas para poder escribir textos idiomáticos. No obstante, la creación de recursos léxicos que contengan marcos especializados puede llevar mucho tiempo. Por lo tanto, se requieren los instrumentos y el marco metodológico adecuados para realizar consultas de corpus, extraer y analizar ejemplos útiles y obtener la estructura conceptual de un concepto o evento. En este estudio se presenta una metodología destinada a crear marcos semánticos para conceptos especializados. El primer paso de nuestra metodología se basa en la extracción automática del corpus de las estructuras predicativas pertinentes para modelar el evento de DEFORESTACIÓN. A continuación, se realiza una anotación semántica manual de verbos y frases sustantivas, lo que permite inferir las estructuras léxicas que dan cuenta de la estructura argumental del concepto. Sostenemos que, en determinadas condiciones, la integración de los términos compuestos en la extracción del corpus es esencial para alcanzar resultados de alta calidad, que maximicen su utilidad para la construcción de los marcos especializados. El evento de la DEFORESTACIÓN, en el dominio de las ciencias ambientales, se presenta aquí como un ejemplo. Ilustramos todos los pasos de nuestra metodología, incluidas las consultas del corpus y la anotación semántica, hasta la creación de marcos especializados para el evento de DEFORESTACIÓN.

Palabras clave: lingüística de corpus; marcos semánticos; fraseología especializada; medio ambiente.

Abstract

Frame-based terminology is a linguistic framework whose main hypothesis is that specialized concepts and events are conceptually organized in the mind of the speaker. It uses specialized semantic frames as a structured representation of this conceptual organiza-

tion of specialized. Lexicographic resources based on frame-based terminology represent a rich source of information beyond static term inventories. Such resources can be useful for domain experts and translators, who need access to domain conceptual structures and their corresponding linguistic realizations in order to author idiomatic documents. However, the creation of lexical resources containing specialized frameworks can be time-consuming. Therefore, appropriate tools and methodological framework are required to perform corpus queries, extract and analyze useful examples, and elicit the conceptual structure of concepts or events. In this study we present a methodology aimed at creating semantic frameworks for specialized concepts. The first step of our methodology is based on the automatic extraction from the corpus of the relevant predicative structures to model the DEFORESTACIÓN event. This is followed by manual semantic annotation of verbs and noun phrases, which allows us to infer lexical structures that account for the argument structure of the concept. We argue that, under certain conditions, the integration of compound terms in corpus extraction is essential to achieve high quality results, which maximize their usefulness for the construction of specialized frameworks. The DEFORESTACIÓN concept, in the domain of environmental sciences, is presented as an example. We illustrate each step of our methodology, including corpus queries and semantic annotation, leading up to the creation of specialized frames of DEFORESTACIÓN.

Keywords: corpus linguistics; semantic frameworks; specialized phraseology; environment.

1. Introducción

Según una larga tradición en lingüística cognitiva, la mente humana accede al conocimiento mediante estructuras conceptuales llamadas “marcos semánticos” (Fillmore y otros, 2003; Fillmore, 2006). Los recientes desarrollos en el campo de la neurociencia han corroborado esta hipótesis (Faber y otros, 2017). Según esta perspectiva, los eventos se representan según sus participantes y las funciones que estos realizan, reflejando así la estructura profunda del significado.

La terminología basada en marcos (TBM) aplica las premisas de la semántica de marcos al estudio de la organización conceptual que subyace en los dominios especializados (Sánchez-Cárdenas y Ramisch, 2019; Faber, 2012, 2015; L’Homme y otros, 2014; Durán-Muñoz, 2017; L’Homme y Pimentel, 2012, *inter alia*). Los marcos semánticos especializados proporcionan un modelo estructurado y coherente para organizar los conceptos en un dominio especializado, ya que modelan los conceptos especializados de manera independiente a las particularidades de cada lengua (Faber, 2015). Para ello, se agrupan estructuras léxicas relacionadas, reflejando sus dimensiones semánticas, así como las relaciones entre conceptos. Esta representación del conocimiento especializado se basa en la premisa de que las estructuras cognitivas que subyacen en los textos científicos se pueden identificar a través del estudio de sus lexicalizaciones. Por ejemplo, el marco semántico del concepto DEFORESTACIÓN contiene unidades léxicas relacionadas con sus consecuencias medioambientales, tales como *erosión*, *efecto invernadero*, *contaminación del suelo*. El interés que suponen los marcos semánticos multilingües para fines terminológicos y de traducción es indudable. Para traducir, es fundamental conocer la configuración semántica de los conceptos especializados, a la par que sus lexicalizaciones en cada idioma, algo necesario tanto para comprender un concepto como para producir textos idiomáticos. A pesar de algunos avances (Hadouche y otros, 2011; Sánchez-Cárdenas y Ramisch, 2019), no existe aún una metodología sistemática que permita generar el marco semántico de un concepto de manera rigurosa e imparcial. Se trata, pues, de un aspecto que merece ser examinado.

Esta investigación pretende contribuir a establecer una metodología robusta y sin sesgos para representar los marcos semánticos de los conceptos de manera independiente de la lengua en los recursos terminológicos. Para ello, en primer lugar, exploramos nuevas técnicas para la identificación y la extracción automática en los corpus de patrones léxico-gramaticales relevantes. Seguidamente, etiquetamos semánticamente dichos patrones de manera manual. Por último, generamos automáticamente la estructura del marco semántico del concepto objeto de estudio.

Por otro lado, proponemos un modelo de presentación de la equivalencia interlingüística según el marco conceptual, un asunto que ha suscitado, hasta ahora, escasa atención. Entendemos la equivalencia lingüística a nivel de la estructura conceptual de los conceptos.

2. Hipótesis y objetivos

Nuestro estudio parte de los principios teóricos de la terminología basada en marcos, un modelo adecuado para representar el conocimiento especializado. Consideramos que los corpus especializados contienen la información necesaria para crear dichos esquemas conceptuales siguiendo una perspectiva “bottom-up”. Sobre la base de estos principios, formulamos la hipótesis de que es posible extraer automáticamente de los corpus información en forma de predicados-argumentos que puede generalizarse a través de diferentes realizaciones léxicas de predicados y clases de argumentos, guiando así la creación de marcos semánticos especializados independientes del lenguaje.

Como estudio de caso nos centramos en el concepto DEFORESTACIÓN, un evento que representa un reto, dada su complejidad. Desde el punto de vista medioambiental, la deforestación es uno de los principales problemas ecológicos del mundo. Desde una perspectiva conceptual, plantea cuestiones tales como qué entidades y procesos causa este evento y cuáles son sus consecuencias sobre el ecosistema. Desde un punto de vista puramente lingüístico, nos interesa averiguar las estructuras lingüísticas equivalentes ligadas a este concepto en inglés y en español.

Algunos de los recursos terminológicos que tratan este concepto son EcoLexiCon¹, DiCoEnviro² y Gemet³. Si bien todos ellos proporcionan información útil, ninguno representa su marco semántico ni tampoco cómo se lexicalizan estas estructuras en las diferentes lenguas. Además, ninguno ofrece información fraseológica. La imagen 1 muestra la representación de DEFORESTACIÓN en EcoLexicon.

Nuestro estudio busca enriquecer los recursos mencionados, proponer una metodología adecuada para el estudio de los conceptos de marcos conceptuales y de sus correspondencias lingüísticas y, por último, representar dicha información. En concreto, respondemos a estas preguntas:

1. ¿Cuál es la estructura cognitiva del concepto especializado DEFORESTACIÓN?
2. ¿Cómo se puede extraer del corpus la información relevante para construir su marco semántico?
3. ¿Cómo se lexicaliza la información sobre el proceso de la *deforestación* en los textos científicos en español y en inglés?

1 <http://ecolexicon.ugr.es>

2 http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi

3 <https://www.eionet.europa.eu/gemet/en/themes/>

Elaborar una descripción tan detallada y sofisticada desde cero basándose en la observación de los textos es una tarea ardua. El mero estudio de las concordancias no es suficiente, ya que la extracción de conclusiones basadas en la mera observación de las concordancias es no solo una tarea herculeana, sino que, además, un método propenso a errores. Por lo tanto, los marcos semánticos deben construirse de manera rigurosa, partiendo del análisis de corpus especializados y llegando a generalizaciones con alto nivel de inter acuerdo. Esto requiere conocimientos lingüísticos, así como el adecuado manejo de herramientas informáticas apropiadas para realizar consultas y extracciones del corpus. Dichas herramientas pueden apoyar, mejorar y facilitar el análisis de los corpus para confirmar y generalizar la introspección lingüística.

Según la lingüística cognitiva (Lakoff y Johnson, 1997; Rosch y Lloyd, 1978), la estructura del lenguaje y de la cognición emanan de la percepción psicosensores de nuestro entorno. La semántica de marcos (Fillmore, 2006) se enmarca en esta perspectiva. Un marco representa un concepto o evento de manera esquemática. Para definir un marco, es necesario identificar los principales participantes de cada situación. Por ejemplo, el marco “ataque” tiene tres elementos básicos: el AGRESOR, el ARMA y la VÍCTIMA. Puede haber otros elementos del marco no esenciales, como el LUGAR o la MANERA.

En principio, los marcos solo son válidos para una cultura y, por tanto, no son universales. No obstante, pueden generalizarse para un grupo de culturas con características comunes (por ejemplo, la cultura occidental). Lo mismo sucede con las lenguas de especialidad, donde un marco semántico se comparte por todos los expertos de un ámbito. Por ejemplo, el concepto HURACÁN, representado por los términos *huracán*, *hurricane*, *ouragan*, en español, inglés y francés, respectivamente, debería describirse según las entidades que provocan ese evento meteorológico y sus consecuencias sobre otras entidades. Dicha descripción debería contener, al menos, los actores y los verbos que participan en la realización conceptual (y lingüística) de este concepto.

Dado que los marcos semánticos están formados por estructuras argumentales complejas, su creación requiere tanto conocimientos lingüísticos como del ámbito, así como herramientas para extraer esa información del corpus (L'Homme y otros, 2014).

Para llegar a conocer el entramado de redes semánticas y las estructuras lingüísticas de los conceptos especializados es necesario realizar consultas complejas, capaces de modelar patrones de coocurrencia morfosintáctica y sintáctica que reflejen la estructura de los predicados. Las herramientas informáticas pueden apoyar, mejorar y facilitar el análisis de los corpus para confirmar la introspección lingüística y llevar a generalizaciones objetivables.

El trabajo del terminólogo consiste en extraer dichas estructuras complejas a partir de textos especializados. Por ejemplo, en el siguiente extracto se pueden inferir diversas relaciones semánticas. Observamos cómo los huracanes y ciclones comparten rasgos semánticos

en cuanto a que son eventos causados por condiciones atmosféricas que desencadenan eventos hídricos y geológicos. Por ejemplo, la entrada terminológica de HURACÁN se enmarca dentro del concepto EVENTO ATMOSFÉRICO. Deberá contener las entidades y los verbos que participan en la realización de este concepto, como se muestra:

EVENTO ATMOSFÉRICO

- o **causado por:** condiciones atmosféricas (p. ej. *baja presión*) poder *formar/originar/evolucionar* en un evento atmosférico (p. ej. *huracán, ciclón*).
- o **acciones del concepto:** rotar/girar, moverse en una dirección, *impacto/golpe/hacer* contacto con un relieve (por ej. *costa, costa, área*).
- o **afecta:** EVENTOS ATMOSFÉRICOS, poder *disparar/producir/causar*, eventos de agua (p. ej. *inundaciones, olas*) o eventos geológicos (p. ej. *deslizamiento de tierra, flujo de escombros*).

4. Extracción de estructuras verbo-nominales del corpus con Sketch Engine

Nuestro punto de partida son dos corpus, en inglés y en español, que contienen 1 105 718 y 971 788 palabras, respectivamente. Están compuestos por artículos científicos, entradas enciclopédicas y noticias especializadas que versan sobre el tema de la deforestación. El corpus fue preprocesado automáticamente mediante la herramienta Sketch Engine (Kilgarriff y otros, 2014).

Nuestras búsquedas se basan en la idea, ampliamente admitida, de que las estructuras argumentales pertinentes a un marco semántico suelen presentarse en la forma prototípica de “sustantivo-verbo-sustantivo”, conocida como “triple”, compuesta por un sintagma nominal, seguida de un verbo o locución verbal, seguida de uno o más sintagmas nominales. Un *triple* está compuesto por un primer sintagma nominal, seguido de un verbo o locución verbal, seguido por otro sintagma nominal. Por ejemplo [*deforestación, acelerar, calentamiento global*].

Nuestro trabajo se inspira en el protocolo de Sánchez-Cárdenas y Ramisch (2019). Los autores utilizan la herramienta informática MWEtoolkit, diseñada para la extracción de palabras polilexicales (Ramisch, 2015), para identificar triples. Si bien los resultados que obtienen son prometedores, se trata de una herramienta compleja que requiere conocimientos informáticos para su utilización. Proponemos aquí un protocolo que permite alcanzar resultados similares utilizando la herramienta de análisis de corpus Sketch Engine, cuyo uso está ampliamente extendido.

Cabe la posibilidad de que la extracción de estructuras de predicados-argumentos representadas en forma de triples pasen por alto alguna información útil. Por ejemplo, algunos

predicados son, en su mayoría, nominales, en lugar de verbales (*la erosión* en lugar de *la erosión*). Además, algunos predicados pueden tener más o menos de dos argumentos (por ejemplo, “[La expansión de la soja] en el sur del Brasil [contribuyó] a la [deforestación] al [estimular la migración a las regiones de frontera agrícola]”). No obstante, esperamos compensar esta inexactitud de los triples al realizar varias consultas, utilizando variantes terminológicas del concepto estudiado (*forest loss*, *desertification*, *forest shrinkage*...), para así cubrir el patrón fraseológico completo que pretendemos describir.

4.1. Estudio de la variación denominativa de “deforestación”

Según los modelos comunicativos y cognitivistas de la terminología (Cabré, 1993; Temmerman, 2000; Freixa, 2006; León Araúz, 2017), la variación denominativa es un fenómeno común en las lenguas de especialidad. Puede deberse a diversas causas, que dan lugar a distintos tipos de variación: dialectal, funcional, discursiva, interlingüística o cognitiva (Freixa, 2006). Tratamos aquí la variación terminológica de tipo discursivo, en la que influyen parámetros tales como el grado de especialización del texto, la retórica y el estilo, la influencia de otras lenguas, la postura ideológica o el punto de vista. La variación denominativa puede ser monolexical (*desertización*) o polilexical (*calentamiento global*). Se conoce a este tipo de sustantivos con distintas denominaciones como “términos multipalabras” (“*multiword terms*”) (Ramisch, 2015) o “términos compuestos” (Cabezas-García, 2020).

Una característica del evento DEFORESTACIÓN es que involucra a participantes que se lexicalizan como sustantivos complejos, como *pérdida de masa forestal* o *tala de árboles*.

Exploramos las variaciones denominativas del concepto DEFORESTATION en el corpus. Para ello, seguimos un proceso de indagación del corpus, que describimos a continuación. En primer lugar, observamos el término *deforestación* en la herramienta “Thesaurus”, que muestra palabras que comparten una misma distribución en el corpus, por lo que pertenecen al mismo campo semántico. Obtenemos 130 resultados, que corresponden a sustantivos tales como *erosión*, *pérdida* o *degradación*. La imagen 2 muestra una visualización gráfica de los primeros 50 resultados.

A continuación, exploramos las concordancias de cada uno de estos sustantivos para seleccionar aquellos que realmente se refieren al concepto estudiado. Obtenemos una lista de 10 sustantivos: *pérdida*, *degradación*, *desaparición*, *desertización*, *muerte*, *disminución*, *destrucción*, *eliminación*, *reducción*, *tala*. Repetimos la búsqueda con *árbol* y seleccionamos *bosque*, *planta*, *alcornoque*, *castaño*, *ecosistema* y *forestal*.

Por último, combinamos ambos grupos para buscar sintagmas nominales que contengan estos dos elementos, del tipo *desaparición del bosque*. Con el objetivo de localizar en el corpus los compuestos nominales, hacemos las siguientes búsquedas “Corpus Query Language” en Sketch Engine:

THE SAURUS

Deforestación ES

deforestación as noun 557x

visualización by SKETCHENGINE

IMAGEN 3

Ejemplos de variantes denominativas de *deforestación* y de sus concordancias

Lemma	de la desaparición de los impresionantes bosques de a de la desaparición de bosques a través de la lluvia áci ia de desaparición de los bosques se redujo a la mitac able desaparición de bosques naturales en diversos p is, la desaparición de bosques naturales, pero no en t ia de desaparición de los bosques presenta una tende una desaparición casi absoluta de los bosques de las una desaparición casi absoluta de los bosques de las
tala de árbol	
pérdida de bosque	
destrucción de el bosque	
pérdida forestal	
desaparición de bosque	
tala de bosque	brozado de tierras, incluyendo la tala de árboles y deforestación. El agotamiento de los o doméstico como industrial, y la tala de árboles para la exportación de madera en rollo a a la región natural; controlar la tala de los bosques y selvas; promover la rotación de c do al desequilibrio causado por la tala del bosque . Esto no implica que desequilibrios en uicidas, el pastoreo excesivo y la tala indiscriminada de árboles son factores que favore o aumentar, a consecuencia de la tala de bosques y de la irrigación de zonas semiáridas
eliminación de bosque	Deforestación La deforestación o tala de árboles es un proceso provocado generalmente mundial, también participan en la tala de innumerables bosques cada año. Los leñadores
eliminación de el bosque	
disminución de el bosque	Reducción de la deforestación y degradación de bosques (REDD). La REDD formó parte import ovenientes de la deforestación y degradación de bosques y la necesidad de aumentar la remoc ocultas de la deforestación y la degradación de los bosques Los bosques constituyen uno de l ntrolada es un factor clave en la degradación de los bosques , afectados por la lluvia ácida prc n un papel más importante en la degradación de los bosques . Por más controvertidas que sear directas de la deforestación y la degradación de los bosques en muchos países del mundo. Un ros actores La deforestación y la degradación de los bosques ocurre tanto en  íses del Nor - tienen un fuerte impacto en la degradación de los bosques . Algo similar ocurre en el Sur, dc ntes de que la deforestación y la degradación de los bosques no son temas "técnicos". Los bosq
tala forestal	

Repetimos el mismo proceso en inglés y obtuvimos por resultado una lista de variantes denominativas de *deforestación* y de *deforestation*, tal y como muestra la tabla 1.

TABLA 1

Variantes denominativas de *deforestación* y de *deforestation*

VARIANTES DEFORESTACIÓN (ESPAÑOL)	VARIANTES DE DEFORESTATION (INGLÉS)
deforestación, degradación de los bosques, degradación de bosques, degradación del encinar, degradación forestal, desaparición de bosques, desaparición de los árboles, desertización, tala, tala de árboles, tala de los bosques, tala del bosque, tala de bosques, destrucción de árboles, destrucción de los árboles, destrucción de los bosques, destrucción del bosque, destrucción de bosques, pérdida de árboles, pérdida forestal, eliminación de bosque, eliminación del bosque, eliminación de los bosques	deforestation, logging, forest shrinkage, forest clearing, clearing of forest, forest loss, forest fragmentation, loss of forest, forest degradation, destruction of forest, forest scarcity, scarcity of forest, degradation of forest, land clearing, forest cover reduction, reduction of forest

4.2. Verbos asociados con el evento DEFORESTATION

Los verbos desempeñan un papel clave en la configuración de cualquier marco semántico, ya que lexicalizan las acciones realizadas por sus participantes. Para identificar los verbos asociados con el evento de la DEFORESTATION llevamos a cabo esta búsqueda CQL:

```
[lemma="deforestation|logging|shrinkage|clearing|loss|degradation|destruction|scar  
city|degradation|  
reduction"] [[0,3][tag="V.*"&lemma!="be|have|do"]  
Filter context: forest within 2 tokens left/right
```

Este patrón localiza en el corpus una variante terminológica de *degradation* seguida por un verbo, de los que excluimos *be*, *have*, *do*, todo ello en el contexto de la palabra *forest* dos posiciones a derecha o izquierda del segmento de búsqueda. Llevamos a cabo una búsqueda similar en español. La imagen 4 muestra un ejemplo de los resultados obtenidos.

IMAGEN 4

Concordancias de verbos relacionados con DEFORESTATION

w the canopy cover threshold. Moreover, forest	degradation may enhance	susceptibility to fire and may result in substan
on. Edges of forest fragments exposed through	deforestation and logging leaves	the forest susceptible to degradation through
ssessed independently and the areas of forest	loss are matched	with the carbon stocks of the forests undergoi
as of tropical deforestation is also rising. Forest	clearing contributes	to climate change, biodiversity loss, reduced t
apt more intensive pasture technologies. Forest	scarcity drives	up land prices, which make intensive growth r
are the same factors that restrict farmers' forest	clearing also limit	their incomes. Still in Ecuador but in a differen
also be reversed. They hypothesize that forest	scarcity resulting	from past deforestation promotes pasture inte
echanism that generates this sequence. Forest	scarcity leads	to higher land prices. This makes it more attra
ctionality. 4.1. Overall forest degradation Forest	degradation is often associated	with decreasing biomass (Sasaki et al., 2011;
secondary forests, suggesting that forest cover	loss is driving	a retrogressive succession process in the stuc
006–2012 period. • Areas of accelerated forest	loss coincided	with high biodiversity areas. • Land-grabbing i
r aimed to investigate if the rate of global forest	loss and the resulting	loss of biodiversity have altered in association

La tabla 2 resume las extracciones verbales realizadas a partir de las concordancias en ambos idiomas.

TABLA 2

Ejemplos del resultado de la extracción de verbos relacionados con DEFORESTATION

VERBOS (ESPAÑOL) ASOCIADOS CON DEFORESTATION	VERBOS (INGLÉS) ASOCIADOS CON DEFORESTATION
<i>impulsar, sobrevivir, provocar, ocasionar, originar, causar, producir, arrasar, alterar, destruir, afectar, acelerar, generar, desaparecer, eliminar, arrasar, impulsar, revertir, frenar, evitar, reducir, minimizar, conservar, preservar, proteger, detener, frenar, interrumpir, incrementar, conducir, desencadenar, combatir...</i>	<i>lead to, depend on, result from, occur within, contribute to, increase by, influence by, occur in, drive by, clear, conserve, degrade, protect, destroy, restore, log, preserve, maintain, remove, survive, increase, stimulate, encourage, spur, aggravate, decrease, promote, drive, accelerate, reduce, lead, displace, affect, aggravate...</i>

4.3. Participantes del evento DEFORESTATION

La deforestación es uno de los principales problemas ecológicos del mundo. En esta parte del estudio, identificamos en el corpus las entidades que participan en este concepto complejo, provocado por una serie de eventos (ej. *sobreexplotación agraria, lluvia ácida*) que causan la degradación de los bosques, lo que desencadena no solo nefastas consecuencias medioambientales (ej. *erosión, pérdida de biodiversidad*), sino también socioeconómicas (ej. *pobreza, migraciones*). Con el fin de conocer estos elementos, ejecutamos una consulta CQL que extrae del corpus los sintagmas nominales recurrentes que lexicalizan las entidades y eventos asociados con la deforestación:

```
1:[tag="N.*"] [word!="\,|:|."] ([tag="P"]?[tag="N.*"])?([tag="A.*"]?[tag="P.*"]?[tag="N.*"])?  
2:[lemma="revertir|frenar|evitar|reducir|minimizar|conservar|preservar|proteger|detener frenar|interrumpir|impulsar|sobrevivir|provocar|ocasionar|originar|causar|producir|arrasar|alterar|destruir|afectar|acelerar|generar|desaparecer|eliminar|arrasar|impulsar"] []{0,5}  
3:([lemma="deforestación|tala|desertización"]|[lemma="degradación|desaparición|destrucción|pérdida|eliminación"] []{0,2} [lemma="bosque|forestal|árbol"]) within <s/>
```

Esta búsqueda localiza en el corpus en español sintagmas nominales complejos relacionados con la deforestación en posición preverbal. La búsqueda se compone de (1:) un sustantivo que puede estar modificado por una preposición y otro sustantivo, o bien por un adjetivo, seguido o no de una preposición y un sustantivo. A continuación (2:) se insertan los verbos previamente identificados (ver punto 4.2) que lexicalizan causas o consecuencias. La última cadena (3:) hace referencia a las variantes terminológicas de *deforestación*.

Esta búsqueda extrae sintagmas nominales en posición de sujeto, ya que es el elemento que no se especifica. Se repitió una búsqueda similar alternando las posiciones 1: y 3: con el objetivo de localizar los SN significativos en posición de complemento. Estos patrones de búsqueda arrojaron una lista amplia de términos simples y complejos, de los que se seleccionaron los 50 más significativos en inglés y en español. La tabla 3 muestra algunos ejemplos en ambos idiomas⁵.

TABLA 3

Términos asociados con DEFORESTACIÓN y términos asociados con DEFORESTATION

TÉRMINOS (ESPAÑOL) ASOCIADOS CON DEFORESTACIÓN	TÉRMINOS (INGLÉS) ASOCIADOS CON DEFORESTATION
agricultura de roza y quema	agricultural land
agricultura intensiva	arable land
aridificación	banana production
biodiversidad	cattle ranching
cambio climático	climate change
contaminación del agua	commercial ranching
crecimiento demográfico	cropland
dsecación de ríos	damage to ecosystem
efectos ambientales	demand for energy
evapotranspiración	demand for food
explotación forestal	diversity decline
extinción de especies	environmental degradation
fenómenos torrenciales	extinction
ganadería	fire
gases de efecto invernadero	global warming
incendio forestal	growth in demand for food
industria maderera	growth in demand energy
migraciones humanas	intensification of land use increased population
pérdida del hábitat	land degradation
perjuicios económicos	technological change
plagas	temperature change
industrialización	highway construction
inundación	livestock
lluvia ácida	population growth
pobreza	soybean farm
urbanización	species decline
desplazamiento de población	transportation cost
efecto invernadero	unsustainable development

5 Cabe destacar que, al tratarse de dos corpus comparables compuestos por textos originales, no siempre se dan equivalencias exactas entre los términos extraídos, lo cual no es un perjuicio para la construcción del marco semántico (ver apartado 5).

4.4. Extracción de triples “nombre-verbo-nombre” con Sketch Engine

Las búsquedas definidas en el punto anterior están compuestas por tres elementos (nombre-verbo-nombre), donde al menos un elemento queda sin especificar. A partir de las concordancias, se extrajeron automáticamente los triples (*nombre-verbo-nombre*), tal y como se muestra en la tabla 4.

TABLA 4

Ejemplo de concordancias y triples extraídos

ARGUMENTO 1	VERBO	ARGUMENTO 2	EXTRACCIÓN DE TRIPLES (NOMBRE-VERBO-NOMBRE)
este alto consumo de energía	provoca	la deforestación de	consumo de energía/ provocar/deforestación
las sobreexplotación agraria	provoca	una gran desertización	sobreexplotación agraria/ provocar/desertización
la agricultura intensiva	impulsa	la pérdida forestal	agricultura intensiva/ impulsar/pérdida forestal
El uso de combustible fósiles, la deforestación y otras actividades que	producen	el efecto invernadero siguen aumentando	deforestación / producir/ efecto invernadero
La desertización	reduce	la evapotranspiración, y consiguientemente los niveles de humedad atmosférica	desertización /reducir / evotranspiración
La pérdida de bosque	acelera	las inundaciones	pérdida de bosque / acelerar / inundaciones

A continuación, se seleccionaron manualmente los triples que se consideraron relevantes para la construcción del marco, descartando aquellos donde había errores⁶. Los triples no seleccionados fueron descartados por contener distintos errores que sugieren nuevas líneas de investigación para futuros trabajos.

Nuestra propuesta para la construcción de marcos especializados se basa en la anotación semántica de los triples extraídos y su posterior generalización, lo que permite descomponer las etapas en fases sucesivas, en un método estructurado y sistemático, como veremos a continuación.

6 El análisis de los errores en la extracción de triples merece un estudio pormenorizado que excede los límites de esta publicación.

Los triples seleccionados en ambos corpus fueron anotados semánticamente como se ejemplifica en la tabla 5. En primer lugar, los verbos se clasificaron en dominios léxicos, un conjunto de categorías semánticas, atendiendo a la semántica del verbo, como ACCIÓN, CAMBIO o EXISTENCIA (Faber y Mairal, 1999). Después, los argumentos de los verbos se clasificaron de acuerdo a su clase semántica (por ejemplo, ACCIÓN, FLORA, DESASTRE NATURAL), según una tipología de sustantivos diseñada para las ciencias ambientales (Buendía-Castro, 2013, y Gil-Berrozpe y otros, 2018). Por último, se asignó a los argumentos un rol temático de un inventario cerrado, según el papel desempeñado en el enunciado (por ejemplo, *Agent*, *Theme*, *Result*).

TABLA 5

Ejemplo de extracción de triples y anotación semántica en tres niveles (rol temático, clase semántica y dominio léxico)

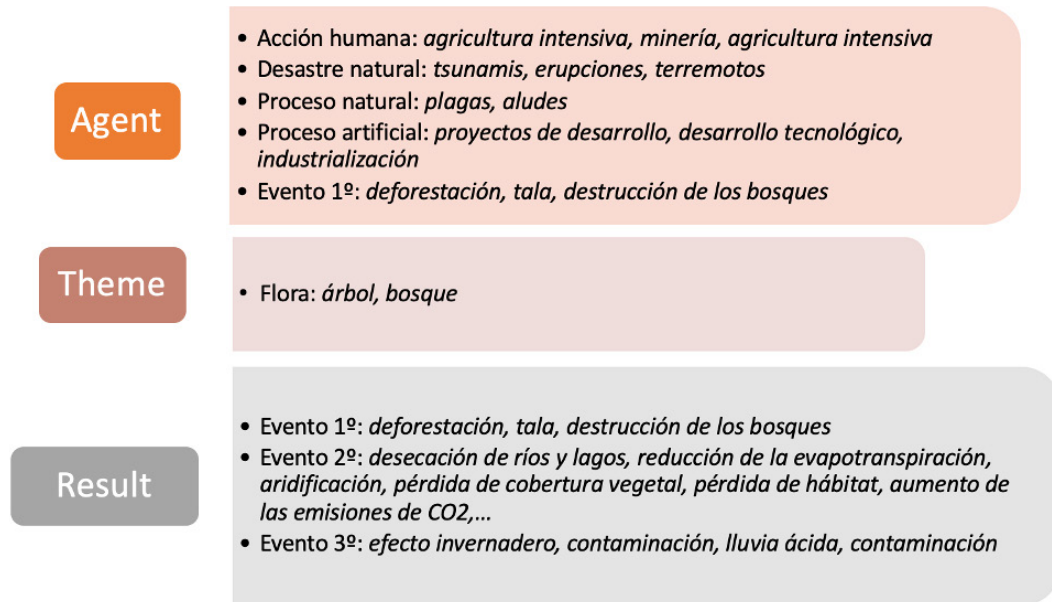
	Argument1	Thematic Role N1	Semantic Class N1	Verb	LEXICAL DOMAIN	Argument 2	Thematic Role N1	Semantic Class N1
1								
2	demand for land	Agent	process>action	spur	CHANGE	deforestation	Patient	attribute
3	Deforestation	Agent	process>change	intensify	CHANGE	natural_flood	Patient	process>loss
4	forest_clearing	Agent	process>loss	contribute_to	CHANGE	change in biodiversity	Patient	process>loss
5	forest_clearing	Agent	process>loss	contribute to	CHANGE	climate_change	Patient	process>loss
6	forest_degradation	Cause	process>change	be a precursor of	EXISTENCE	deforestation	Result	process>loss
7	forest_scarcity	Theme	attribute	drive by	MOVEMENT	land_price	Agent	process>action
8	forest_scarcity	Agent	attribute	lead	MOVEMENT	higher land price	Result	process>action
9	technological_change	Cause	process>change	promote	EXISTENCE	deforestation	Result	process>loss
10	technological_change	Agent	process>change	affect	CHANGE	forest_clearing	Patient	process>loss
11	deforestation	Agent	process>change	influence_by	CHANGE	habitat configuration	Patient	process>loss
12	deforestation	Agent	process>change	drive	MOVEMENT	technological change	Theme	process>action
13	soybean expansion	Agent	process>action	lead_to	MOVEMENT	deforestation	Result	process>loss
14	deforestation	Theme	process>change	drive by	MOVEMENT	transportation	Agent	process>loss
15	technique	Agent	process>action	increase	CHANGE	forest_clearing	Patient	process>loss
16	logging	Agent	process>action	lead	MOVEMENT	deforestation	Result	process>loss
17	cattle ranching	Agent	process>action	lead to	MOVEMENT	deforestation	Result	attribute
18	timber	Agent	flora	lead to	MOVEMENT	deforestation	Theme	process>addition
19	labour	Cause	process>action	provoke	EXISTENCE	deforestation	Result	process>loss
20	transportation_cost	Agent	artificial object	increase	CHANGE	deforestation	Patient	landform

La anotación de los dominios léxicos de los verbos, y los roles temáticos y clases semánticas de los sustantivos que funcionan como argumentos, revela patrones léxico-gramáticos recurrentes en los corpus. Por ejemplo, es muy frecuente la estructura “PROCESO *aumentar* ACCIÓN”, en la que el primer sustantivo actúa como Agente y el segundo como Paciente. Algunos de los sustantivos de la categoría ACCIÓN en esta estructura son *producción de bananas*, *ganadería comercial* o *expansión de la soja*, mientras que la categoría PROCESO se lexicaliza como *deforestación*, *pérdida de bosque* o *erosión de los bosques*. Por tanto, este esquema refleja las actividades que causan la pérdida de masa forestal. La imagen 5 muestra un esquema de los roles temáticos y clases semánticas en el corpus en español.

La anotación semántica de triples “sustantivo-verbo-sustantivo” permite inferir esquemas conceptuales lexicalizados que sería difícil vislumbrar de manera intuitiva. En otras pala-

IMAGEN 5

Ejemplo de la anotación semántica de sintagmas nominales en el corpus español



bras, la anotación semántica de triples extraídos del corpus descompone la creación de marcos especializados en pasos sistemáticos y fáciles de manejar.

5. Agrupación de estructuras semánticas para la construcción de marcos semánticos multilingües

Partiendo de la premisa de que los patrones fraseológicos similares revelan una misma estructura semántica, los triples anotados se agruparon según su similitud, utilizando un *script* en Python que reúne los triples que comparten la misma anotación. Así pues, se agruparon todos aquellos verbos que fueron etiquetados con un mismo dominio léxico. A continuación, se agruparon según su estructura argumental. Por ejemplo, se agruparon las filas 4 y 5 de la tabla 1, ya que el dominio léxico del verbo, así como las clases y roles de los sustantivos, son idénticos.

Los esquemas léxicos agrupados resultantes reflejaron no sólo las pautas léxico-gramaticales del concepto DEFORESTACIÓN, sino también la estructura conceptual de este evento. Basándonos en esto, observamos que el concepto DEFORESTACIÓN se articula en los textos científicos en torno a cuatro dimensiones, representadas por los dominios léxicos de ACCIÓN, EXISTENCIA, CAMBIO y MOVIMIENTO.

5.1. Los marcos semánticos del concepto DEFORESTACIÓN

Hacemos una propuesta de representación de conceptos especializados en recursos lexicográficos y terminológicos. Partiendo de la información lingüística extraída del corpus en forma de triples, se generalizan esquemas conceptuales o marcos semánticos. Estos se articulan en torno a los dominios léxicos a los que pertenecen los verbos de los triples, divididos, a su vez, en subdominios léxicos según estructuras argumentales similares.

Dichas estructuras conceptuales lexicalizan las estructuras cognitivas que subyacen en el concepto. Al tratarse de un conocimiento especializado compartido por todos los expertos del ámbito, son independientes a la lengua. Obviamente, su expresión lingüística sí varía en cada lengua. Tanto es así que no siempre encontramos verbos equivalentes dentro del mismo subdominio, tal y como puede comprobarse en las tablas que presentamos a continuación.

En un primer nivel se sitúan las dimensiones del concepto, en forma de dominios léxicos (Buendía-Castro y Sanchez-Cárdenas, 2016). En un segundo nivel se encuentran los verbos lexicalizados en el concepto, organizados en subdominios. La pertenencia de los verbos a cada subdominio depende de su estructura argumental. Así pues, se agrupan aquellos triples que comparten una misma configuración de roles temáticos y clases semánticas. Las definiciones de los subdominios se basan en la propuesta de Faber y Mairal (1999). Dado que son generalizables a cualquier lengua, se expresan, por convención, en inglés.

Cada tabla refleja una de las cuatro dimensiones semánticas⁷ del marco semántico de DEFORESTACIÓN, a saber, ACTION, CHANGE, EXISTENCE y MOVEMENT. Todas ellas conforman las múltiples dimensiones del concepto, con sus diversas perspectivas y matices, reflejando la complejidad de este concepto.

La definición de cada dimensión se compone de roles temáticos y clases semánticas correspondientes a sus estructuras argumentales. Se ilustra cada dimensión del concepto con extractos de concordancias extraídas del corpus en español. Seguidamente, se describen los subdominios, definidos según la propuesta de Faber y Mairal (1999), junto con los verbos en inglés y español correspondientes extraídos del corpus⁸. Por razones de espacio, no se

7 No ha de confundirse la multidimensionalidad de un concepto con su polisemia (León-Araúz y San Martín, 2012).

8 Debido a la falta de isomorfismo entre las lenguas, las correspondencias entre verbos deben entenderse no uno a uno, sino dentro de la dimensión semántica.

representa en detalle la estructura argumental de cada verbo, algo que, por otro lado, solo sería plenamente satisfactorio mediante una herramienta electrónica.

Dimensión 1 del marco semántico de DEFORESTACIÓN:

Un **AGENT** (ENTIDAD HUMANA: *agricultor*) elimina un **THEME** (FLORA: *árbol*) cortándolo.

- **Los agricultores talan los bosques para obtener espacio para sus cultivos.**

Dominio léxico	Definición	Verbos (ES)	Verbos (EN)
ACTION	To remove (something) from something larger by using a sharp implement	<i>deforestar</i> <i>talar</i> <i>cortar</i>	<i>deforest</i> <i>log</i> <i>cut</i>

Dimensión 2 del marco semántico de DEFORESTACIÓN:

Un **AGENT** (ACCIÓN HUMANA|DESASTRE NATURAL|PROCESO NATURAL|PROCESO ARTIFICIAL) genera un cambio en un **PATIENT** (FLORA: *árbol*) que afecta a un **THEME** (EVENTO: *deforestación*), o bien lo revierte.

- **La expansión agrícola desencadena deforestación.**
- **La tecnología puede minimizar el impacto de la deforestación.**

Dominio léxico	Definición	Verbos (ES)	Verbos (EN)
CHANGE	To cause something	<i>causar</i> <i>desencadenar</i>	<i>stimulate</i> <i>influence</i> <i>encourage</i> <i>provoke</i> <i>promote</i>
	To cause something to be different by changing its characteristics	<i>alterar</i> <i>afectar</i>	<i>affect</i>
	To change something by counteracting a negative process	<i>revertir</i>	<i>restore</i>
	To change something by reducing its consequences	<i>minimizar</i>	<i>slow</i>

Dimensión 3 del marco semántico de DEFORESTACIÓN:

Un **AGENT** (PROCESO NATURAL, PROCESO ARTIFICIAL: *migración, pérdida de biodiversidad*) destruye o protege un **PATIENT** (FLORA: *árbol*), provocando un **RESULT** (EVENTO:

deforestación) que, a su vez, produce otro RESULT (EVENTO: *lluvia ácida*).

- *Los aludes incrementan la deforestación.*
- *La deforestación incrementa la aridificación.*
- *La aridificación reduce el hábitat de numerosas especies.*

Dominio léxico	Definición	Verbos (ES)	Verbos (EN)
EXISTENCE	To begin to exist	<i>ocasionar originar generar producir</i>	<i>result occur</i>
	To continue to exist in despite of difficulties	<i>sobrevivir proteger</i>	<i>protect preserve maintain</i>
	To allow something fragile to continue to exist	<i>conservar preservar</i>	<i>conserve</i>
	To cause something to become bigger	<i>incrementar</i>	<i>increase spur</i>
	To cause something to become smaller	<i>reducir</i>	<i>decrease</i>
	To cause something to become bigger for the worse	<i>agravar</i>	<i>aggravate degrade</i>
	To cause something to stop existing	<i>destruir eliminar interrumpir</i>	<i>eliminate</i>
	To cause something to stop existing in the perception of others	<i>desaparecer</i>	<i>disappear</i>
	To cause something to stop existing by placing it somewhere else	--	<i>remove</i>

Dimensión 4 del marco semántico de DEFORESTACIÓN:

Un **AGENT** (PROCESO NATURAL O ARTIFICIAL) conduce a un **RESULT** (EVENTO: deforestación)

- *La intensificación de las actividades de pastoreo de ganado puede conducir a la deforestación.*
- *Los procesos de erosión se ven acelerados por la deforestación de las riberas.*

Dominio léxico	Definición	Verbos (ES)	Verbos (EN)
MOVEMENT	To cause something to move	<i>impulsar</i>	<i>promote drive</i>
	To cause something to move faster	<i>acelerar</i>	<i>accelerate</i>
	To cause something not to move anymore	<i>detener frenar</i>	<i>reduce</i>
	To cause something to move to a certain place	<i>conducir (a)</i>	<i>lead</i>

Nuestro estudio muestra que el marco semántico del evento de DEFORESTACIÓN conlleva un escenario complejo en el que interactúan numerosos participantes. Este marco semántico representa un proceso causado por una actividad humana o un evento natural. Los resultados indican que las actividades humanas son responsables de este proceso, ya sea directamente (por la tala de árboles) o indirectamente (debido a las emisiones de gases de efecto invernadero que causan desastres naturales que provocan la erosión del suelo). Las consecuencias de la deforestación afectan a todo el ecosistema a muchos niveles y tienen un impacto significativo en los organismos vivos.

6. Conclusión

Los marcos semánticos de los conceptos especializados resultan útiles para comprender su estructura cognitiva. Sin embargo, resultan difíciles de identificar de manera intuitiva. Las herramientas lingüísticas computacionales, unidas a las estrategias de búsqueda en los corpus especializados, pueden ayudar a extraer la información lingüística necesaria para construir dichos marcos. Para ello, extraemos del corpus estructuras fraseológicas, en forma de triples nombre-verbo-nombre que reflejan cómo se lexicaliza la estructura conceptual del concepto. Hemos mostrado cómo realizar dichas búsquedas. Por otro lado, la información recuperada puede utilizarse para mejorar las entradas terminológicas con información fraseológica, como las combinaciones sustantivo-verbo.

Proponemos una representación de la información conceptual, que es independiente a la lengua, unida a la información relevante en cada idioma, de tal manera que se reflejan las correspondencias interlingüísticas. Este tipo de entradas terminológicas son valiosas para los expertos no nativos que codifican en un idioma extranjero, así como para los traductores. Sería interesante llevar a la práctica este modelo de representación en un recurso terminológico electrónico.

7. Agradecimientos

Mi más profundo agradecimiento a Carlos Ramisch por haber desarrollado el programa MWToolkit y haber realizado la programación informática que sustenta esta investigación, así como por su inspiración e inestimable apoyo.

El presente trabajo se ha realizado en el marco del proyecto Integración transversal de la cultura en una base de conocimiento terminológico medioambiental (TRANS-CULTURE, PID2020-118369GB-I00), financiado por el Ministerio de Ciencia e Innovación de España.

8. Bibliografía citada

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

BUENDÍA-CASTRO, Miriam, y Beatriz SÁNCHEZ-CÁRDENAS, 2016: "Using Argument Structure to Disambiguate Verb Meaning" en T. MARGALITADZE y G. MELADZE (eds.): *Proceedings of the XVII EURALEX international congress*, Tbilisi: Ivane Javakhishvili Tbilisi University Press, 482-490.

CABEZAS-GARCÍA, Melania, 2020: *Los términos compuestos desde la Terminología y la Traducción*, Berlin: Peter Lang.

CABRÉ, María Teresa, 1993: *La terminología. Teoría, metodología, aplicaciones*, Barcelona: Antártida.

DURÁN-MUÑOZ, Isabel, 2017: "Producing frame-based definitions", *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 22 (2), 223-249.

FABER, Pamela (ed.), 2012: *Una visión lingüística cognitiva de la terminología y el lenguaje especializado*, Berlín: Walter de Gruyter.

FABER, Pamela, 2015: "Frames as a framework for terminology" en H. J. K. FOCKAERT y F. STEURS (eds.): *Handbook of Terminology* 1, Amsterdam/Philadelphia: John Benjamins, 14-33.

FABER, Pamela, y Ricardo MAIRAL, 1999: *Constructing a lexicon of English verbs*, Nueva York: Mouton de Gruyter.

FABER, Pamela, Juan VERDEJO-ROMÁN, Pilar LEÓN-ARAÚZ, Arianne REIMERINK y Gloria GUZMÁN, 2017: "Specialized knowledge processing in the brain: an fMRI study" en P. FAINI (ed.): *Terminological Approaches in the European Context*, Newcastle-upon-Tyne: Cambridge Scholars Publishing, 168-182.

FILLMORE, Charles, 2006: "Frame semantics", *Cognitive linguistics: Basic readings* 34, 373-400.

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

FREIXA, Judit, 2006: "Causes of denominative variation in terminology: A typology proposal", *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 12 (1), 51-77.

GIL-BERROZPE, Juan-Carlos, Pilar LEÓN-ARAÚZ y Pamela FABER, 2018: "Subtypes of Hyponymy in the Environmental Domain: Entities and Processes" en Cristophe ROCHE (ed.): *Proceedings of the 10th International Conference on Terminology & Ontology: Theories and Applications (TOTh 2016)*, Chambéry: Éditions de l'Université Savoie Mont Blanc, 39-54.

HADOUCHE, Fadila, Guy LAPALME y Marie-Claude L'HOMME, 2011: "Attribution de rôles sémantiques à des actants", *Actes de Traitement automatique des langues TALN 2011*, Montpellier, France.

KILGARIFF Adam, Vít BAISA, Jan BUŠTA, Miloš JAKUBÍČEK, Vojtěch KOVÁŘ, Jan MICHELFEIT, Pavel RYCHLÝ y Vít SUCHOMEL, 2014: "The Sketch Engine: ten years on", *Lexicography* 1 (1), 7-36.

LAKOFF, George, y Mark JOHNSON, 1997: *Metáforas de la vida cotidiana*, Madrid: Cátedra.

LEÓN-ARAÚZ, Pilar, 2017: "Term and Concept Variation in Specialized Knowledge Dynamics" en Patrick DROUIN, Aline FRANCOEUR, John HUMBLEY y Aurélie PICTON (eds.): *Multiple Perspectives on Terminological Variation*, Amsterdam/Philadelphia: John Benjamins, 213-258.

LEÓN ARAÚZ, Pilar, y Antonio SAN MARTÍN, 2012: "Multidimensional Categorization in Terminological Definitions" en Ruth VATVEDT FJELD y Julie MATILDE TORJUSEN (eds.): *15th EURALEX International Congress*, 578-584.

L'HOMME, Marie-Claude, y Janine PIMENTEL, 2012: "Capturing syntactico-semantic regularities among terms: An application of the FrameNet methodology to terminology", 8th International Conference on Language Resources and Evaluation (LREC), 262-268.

L'HOMME, Marie-Claude, Benoît ROBICHAUD y Carlos SUBIRATS RÜGGEBERG, 2014: "Discovering frames in specialized domains", 9th International Conference on Language Resources and Evaluation (LREC), 1364-1371.

RAMISCH, Carlos, 2015: *Multiword Expressions Acquisition: A Generic and Open Framework*, Berlín: Springer.

ROSCH, E., y B. B. LLOYD (eds.), 1978: *Cognition and categorization*, Lawrence Erlbaum.

SÁNCHEZ-CÁRDENAS, Beatriz, y Carlos RAMISCH, 2019: "Eliciting specialized frames from corpora using argument-structure extraction techniques", *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 25 (1), 1-31.

TEMMERMAN, Rita, 2000: *Towards New Ways of Terminology Description. The Sociocognitive-Approach*, Amsterdam/Philadelphia: John Benjamins Publishing Company.