

Categorías semánticas para describir estructuras argumentales en un ámbito de especialidad¹

Los estudios sobre tipologías nominales de sustantivos no abundan en las lenguas de especialidad. Algo que resulta sorprendente si tenemos en cuenta el poder descriptivo que una tipología nominal puede tener en un ámbito de especialidad, donde existe una mayor restricción del significado de las unidades lingüísticas y de sus combinaciones léxicas. En este artículo exponemos, por un parte, los primeros pasos de una tipología nominal en el ámbito de especialidad del medio ambiente y en el subdominio de los desastres naturales. Por otro lado, planteamos aquí una hipótesis según la cuál si los términos que aparecen en los argumentos de los verbos de un ámbito de especialidad se clasifican y describen según los roles semánticos que suelen ocupar –basándonos en un estudio de corpus– es posible predecir los verbos asociados con una estructura argumental dada. Se trata de una idea que cobra especial interés en el ámbito de la traducción, ya que estas predicciones tienen un alcance interlingüístico.

El objetivo de este artículo es plantear las bases necesarias para sistematizar la descripción de la estructura argumental, requisito previo para verificar la hipótesis señalada. En concreto, exponemos un protocolo para establecer las categorías semánticas de los conceptos del subdominio de los desastres naturales. Utilizamos la estructura conceptual de EcoLexicon (ecolexicon.ugr.es), un tesoro visual que representa los conceptos especializados del medio ambiente en redes semánticas. Para llevar a cabo la clasificación de conceptos en categorías, nos basamos en las relaciones semánticas de EcoLexicon. En este artículo tomamos como punto de partida los términos en español pero dado que la equivalencia entre términos está establecida en EcoLexicon, la clasificación podría ser fácilmente extrapolable a otras lenguas.

1. La estructura argumental como predictor léxico interlingüístico

Los estudios sobre clasificación de sustantivos en categorías abundan. Algunos adoptan una perspectiva estrictamente lingüística, como las clases de objetos de Gross² (1994), las clases léxicas de Bosque (1999) o la tipología nominal de Flaux et

¹ Esta investigación ha sido realizada en el marco del proyecto de investigación RECORD: Representación del Conocimiento en Redes Dinámicas [Knowledge Representation in Dynamic Networks, FFI2011-22397], financiado por el Ministerio de Ciencia e Innovación de España.

² Es cierto que las clases de objetos tienen como perspectiva una aplicación computacional.

Van Velde (2000). Otros, desde una perspectiva computacional, han dado lugar a ontologías tales como WordNet Miller 1990, FrameNet (Baker *et al* 1998), VerbNet (Kipper 2005) o ADESSE³ (García-Miguel *et al* 2010).

Ninguno de estos estudios, sin embargo, es directamente transferible a la clasificación de conceptos en un área de especialidad. En ese sentido, el grupo LexiCon está trabajando en un sistema de clasificación que permita estructurar la base de conocimientos EcoLexiCon en categorías semánticas nominales. EcoLexicon representa de forma visual el conocimiento especializado en el ámbito de las Ciencias Ambientales. Hasta ahora, la organización de los conceptos de esta base de datos se ha basado en roles semánticos (AGENTE, PROCESO, PACIENTE, RESULTADO), pero dicha clasificación resulta insuficiente, ya que no los estructura según sus rasgos semánticos.

El objetivo último de la clasificación de los conceptos de EcoLexiCon en categorías semánticas es doble. Por una parte, nos permitirá mejorar la información fraseológica de la base de conocimiento. Hasta ahora, los recursos lexicográficos especializados han prestado poca atención a la combinatoria de los términos. Resulta sin embargo sorprendente si tenemos en cuenta que cada término tiene unas preferencias léxicas que varían de una lengua a otra. Por ejemplo, los verbos que se combinan con un término en una lengua no pueden traducirse utilizando equivalencias de la lengua general, puesto que en los lenguajes de especialidad, cada idioma cuenta con reglas de combinatoria léxica propias que a menudo se basan en la semántica de sus argumentos. Desde esta perspectiva, una tipología de las clases semánticas de los distintos ámbitos de las Ciencias Ambientales nos permitiría alcanzar un mayor poder descriptivo de la fraseología propia de cada ámbito. La segunda utilidad de esta clasificación sería la posibilidad de predecir la traducción multilingüe de verbos basándonos en las estructuras actanciales. Esta idea se basa en una hipótesis de que los verbos equivalentes en distintas lenguas *próximas* comparten una misma estructura actancial (Buendía Castro 2013). En el siguiente ejemplo puede comprobarse cómo, a pesar de que los verbos *spew*, *éjecter* y *expulsar* no siempre son equivalentes directos en la lengua general, sí funcionan como equivalentes dentro del subdominio de la volcanología dado que, como puede observarse, estos verbos comparten una misma estructura actancial. Es decir, sus argumentos tienen el mismo rol semántico y función sintáctica. Además, los términos que actúan como argumentos (*volcán*, *lava*) pertenecen a una misma clase conceptual (ACCIDENTE GEOGRÁFICO, MATERIAL GEOLÓGICO).

Campo: Ciencias Ambientales Ámbito: Volcanología Marco: RELEASE frame			
Rol semántico	agente	verbos	TEMA
<u>Sintaxis</u>	Sujeto		Complemento Directo
<u>Clase semántica</u>	ACCIDENTE GEOGRÁFICO		MATERIAL GEOLÓGICO
<u>Lexicalizaciones</u>	<i>The volcano</i>	spewed	lava and ashes
	El volcán	expulsó	lava y cenizas
	Le volcan	a éjecté	de la lave et des cendres

 Tabla 1. Estructura argumental de los verbos *spew*, *expulsar* y *éjecter*

Nuestra hipótesis es que si conseguimos clasificar todos los conceptos de la base de conocimiento EcoLexiCon en clases conceptuales, seremos capaces de establecer un mecanismo semiautomático que permita averiguar la traducción de un verbo dado en contexto dentro de un ámbito de especialidad. Esto es lógico si tenemos en cuenta que cada verbo selecciona ciertas categorías en cada uno de sus argumentos y que, a su vez, cada estructura argumental lleva asociada unos verbos concretos. Por ejemplo, dentro del subdominio de la volcanología, a partir del enunciado (1) se deduce la estructura argumental (2) y las posibles traducciones de ese verbo en francés e inglés (3 y 4 respectivamente). Como se observa en la tabla 1, la estructura argumental de los verbos *spew*, *expulsar* y *éjecter* coincide en cuanto a roles semánticos (AGENTE, TEMA) y categorías semánticas (ACCIDENTE GEOGRÁFICO, MATERIAL GEOLÓGICO). Esto nos permite decir que es muy probable que los verbos que comparten una estructura argumental dada sean equivalentes. Si esta hipótesis es cierta, una vez que hayamos estudiado las estructuras argumentales de los verbos de cada subdominio será posible determinar la traducción de un verbo de un subdominio de especialidad a partir de su estructura argumental.

1. El volcán sigue expulsando lava y cenizas.
2. [AGENTE/S/ACCIDENTE GEOGRÁFICO/] V [TEMA/COD/MATERIAL GEOLÓGICO]
3. Verbos en francés correspondientes a esta estructura: *éjecter*, *rejeter*, *cracher*, *éjecter*
4. Verbos en inglés correspondientes a esta estructura *expel*, *eject*, *spit*, *erupt*

2. Hacia una tipología de las clases semánticas del Medio Ambiente

Dado que cada ámbito de especialidad tiene unos patrones lingüísticos propios, distintos a la lengua general, el establecimiento de categorías semánticas de un campo de especialidad debe hacerse atendiendo a la idiosincrasia de cada ámbito, en este caso el Medio Ambiente. Presentamos aquí la metodología que seguimos para establecer estas categorías semánticas. El primer paso consiste en la constitución de un corpus sobre el subdominio que estudiamos, por ejemplo el de la volcanología, seguido de su análisis utilizando herramientas semiautomáticas como SketchEngine³ o AntConc⁴ y, por último, la representación de los resultados en la base de datos Eco-LexiCon. Después, explotamos los resultados siguiendo varios procedimientos que explicamos en los apartados siguientes.

2.1. Constitución y análisis del corpus

El primer paso de nuestro estudio es la constitución corpus comparable en inglés, francés y español con el fin de establecer equivalencias interlingüísticas. Para ello existen dos métodos.

El primer método consiste en la recopilación de artículos científicos, de divulgación y periodísticos de un sub-ámbito de especialidad dado, por ejemplo la volcanología, que pertenece al ámbito más general de la Sismología, los desastres naturales meteorológicos (Meteorología) o los movimientos de ladera de cadenas montañosas (Geología). Extraemos los artículos de revistas especializadas a través de bibliotecas electrónicas y los convertimos a formato txt. Al ser ámbitos tan restringidos, un corpus pequeño de unas doscientas mil palabras es a menudo suficiente para que sea representativo. En este caso, hemos constituido corpus sobre desastre naturales meteorológicos.

Otra opción complementaria para constituir un corpus especializado es la herramienta automática WebBootCat, integrada en SketchEngine. Esta herramienta permite buscar en la web de manera automática textos en los que aparezcan distintas combinaciones de número variable de un conjunto de palabras clave. Por ejemplo, para obtener un corpus de los desastres naturales causados por fenómenos meteorológicos hemos hecho una búsqueda a partir de esta lista de palabras clave: *tifón, huracán, tornado, tsunami, precipitación, vientos, huracán, terremoto, tormenta tropical, fenómeno, intensidad, destrucción, daños, costa, desastres*. Hemos combinado de manera automática estas palabras en grupos de tres palabras para buscar textos que las contengan. Los tipos de texto que hemos obtenidos son artículos divulgativos o artículos de prensa sobre fenómenos meteorológicos.

Al aunar estos dos procedimientos, hemos obtenido un corpus de 300.000 palabras, que resulta representativo si tenemos en cuenta lo restringido que es este campo.

³ <<http://www.sketchengine.co.uk/>>

⁴ <<http://www.antlab.sci.waseda.ac.jp/software.html>>

2.2. Categorías léxicas: del término al verbo y del verbo al término

El verbo selecciona los sustantivos con los que se combina, por ejemplo el verbo *solucionar* requiere en su COD sustantivos como *problema, situación*. Pero también sucede lo contrario, y es que los sustantivos, sobre todo en las lenguas de especialidad, también imponen restricciones léxicas al verbo con el que se combinan. El corpus se procesa con el programa SketchEngine (Kilgarriff *et al* 2004). La función “Word List” nos permite acceder a una lista de las palabras clave más frecuentes del corpus, de donde extraemos los términos más recurrentes. Por ejemplo, en el corpus de desastres naturales obtenemos términos como *huracán, tsunami, tormenta tropical*. Observamos en la figura 1 cómo el término *huracán* aparece en el corpus con una serie de verbos prototípicos tales como *golpear, arrasar, destruir, azotar*.

densas </p><p> Myanmar. En 2008, un intenso	huracán golpeó	la costa de Myanmar (antes llamada Birmania
casi imposible, tanto, quizá, como que un	huracán arrase	una costa. Uno puede tomar medidas y precaverse
El poderoso viento y la lluvia de un	huracán golpean	un edificio Florida . Los huracanes son
Louis-Georges Arsenault. "Normalmente, un	huracán destruye	las frágiles viviendas de los más pobres
las huellas de la hecatombe; impetuosos	huracanes derribaron	altas torres, y no faltan, más bien sobran
Pero lo peor no queda ahí, pues ciclones y	huracanes arrasarán	todo lo que encuentren a su paso. Esta
mal tiempo también a causado desastres: el	huracán azotó	la región, arrancó techos y rompió árboles
cinco años, el 29 de agosto del 2005, ese	huracán provocó	devastadoras consecuencias en Nueva Orleans
sin hogar. Las fuertes precipitaciones del	huracán causaron	abundantes inundaciones en varias partes
de última hora. El huracán Katrina	golpeó	ayer la costa de Florida causando importates

Figura 1. Concordancia de [huracán + V]

Estudiamos los verbos asociados con cada término gracias a la función “Word Sketch”, que permite una visualización más rápida y directa que la lista de concordancias. Tal y como observamos en la tabla 2, obtenemos una lista de los verbos que se combinan de manera más frecuente con cada uno de estos términos en posición de sujeto y objeto, dentro de estructuras como “<Term> V N”, donde <Term> representa cualquiera de los términos sobre los que se consulta el “Word Sketch”. Estas dos estructuras indican que el AGENTE puede aparecer en primera o segunda posición argumental.

	“<Term> V N”	“N V <Term>”
huracán	<i>tocar, pasar, afectar, producir, azotar, atravesar</i>	<i>formar, ser, causar, acercar, impulsar, afrontar, originar</i>
tsunami	<i>viajar, golpear, devastar, provocar, alcanzar</i>	<i>generar, provocar, llegar, producir, causar</i>
tormenta tropical	<i>tocar, dañar, producir, causar, arrojar, originar, ocasionar</i>	<i>desarrollar, pasar, originar, producir</i>

Tabla 2. Verbos asociados con los términos *huracán, tsunami, tormenta tropical*

Observamos que estos términos seleccionan verbos que expresan las relaciones causa-efecto (*causar, producir, originar, ocasionar*), verbos de movimiento (*pasar, tocar, alcanzar, atravesar*), verbos de destrucción (*azotar, dañar*) y verbos de existencia (*originar, formar, ser, producir*).

A su vez, esta información sirve, de manera inversa, para estudiar los sustantivos que aparecen en las posiciones clave de cada uno de estos verbos. Esto nos da información sobre las categorías semánticas con las que se combina cada verbo. Desde esta perspectiva, es la propia combinatoria del verbo, dentro de cada lenguaje de especialidad, lo que nos guía en el establecimiento de las categorías léxicas. Así, las categorías son más significativas y alcanzan un mayor poder predictivo que si las estableciéramos independientemente del corpus.

Una vez que obtenemos la lista de verbos, estudiamos la estructura argumental de cada uno de ellos prestando especial atención a los sustantivos que aparecen como argumentos. Estos pueden agruparse en grandes categorías según la similitud de sus rasgos. Por ejemplo, es evidente que *cometa* y *meteorito* pertenecen a una misma categoría diferente a la de conceptos como *partícula* o *huracán*. El siguiente paso consiste en extraer las categorías léxicas que se combinan con cada verbo. Así, el verbo *golpear* aparece en nuestro corpus combinado con los siguientes tipos de sustantivo en las posiciones de Agente y Paciente:

Categoría léxica	N del Agente
CUERPO CELESTE	cometa, meteorito, asteroide,
PARTÍCULA ELEMENTAL	partícula, neutrinos, electrones, fotones
RADIACIÓN	sol, rayo de luz, radiación
DESASTRE NATURAL	huracán, tornado, tifón, tsunami, tormenta tropical, terremoto, inundación

Tabla 4. Algunos agentes del corpus sobre desastres naturales meteorológicos

Categoría léxica	N del Paciente
LUGAR NATURAL	suelo, tierra, bosque, costa, playa
LUGAR ARTIFICIAL	edificio, colegio, casa
LUGAR POLÍTICO	región, continente, país, ciudad

Tabla 5. Algunos pacientes del corpus sobre desastres naturales meteorológicos

Estas listas de los sustantivos más comunes en los argumentos de cada verbo nos llevan a constituir una primera clasificación de las categorías semánticas como CUERPO CELESTE o PARTÍCULA ELEMENTAL. Se trata sin embargo de una primera clasificación que es necesario verificar mediante otros procedimientos que

nos permitan tener criterios lingüísticos más fiables. Para ello, utilizamos tests distribucionales basados en las relaciones conceptuales propias de cada categoría como veremos en el siguiente apartado.

2.3. Tests distribucionales basados en relaciones conceptuales

Con el fin de asentar cada categoría y sus miembros sobre criterios sólidos, establecemos para cada categoría una serie de tests distribucionales basados en las relaciones semánticas que cada concepto activa dentro del propio corpus y en la base de conocimiento EcoLexicon. Si tomamos el ejemplo de la categoría DESASTRE NATURAL en EcoLexiCon, vemos que está caracterizada por las proposiciones conceptuales siguientes:

- Un DESASTRE NATURAL causa PÉRDIDAS HUMANAS/MATERIALES/ECONÓMICAS
- Un DESASTRE NATURAL daña el ENTORNO
- Un DESASTRE NATURAL ocurre en un periodo de tiempo corto
- Un DESASTRE NATURAL ocurre de manera violenta

Los sustantivos *huracán*, *tornado*, *tifón*, *tsunami*, *tormenta tropical*, *terremoto*, *inundación* cumplen la veracidad de estas proposiciones. Todos ellos causan pérdidas humanas, materiales y/o económicas, dañan el entorno y ocurren rápido y de forma violenta.

- Un *huracán/tornado/tifón/tsunami/tormenta tropical/terremoto/inundación* causa PÉRDIDAS HUMANAS/MATERIALES/ECONÓMICAS.
- Un *huracán/tornado/tifón/tsunami/tormenta tropical/terremoto/inundación* daña el MEDIO AMBIENTE.
- Un *huracán/tornado/tifón/tsunami/tormenta tropical/terremoto/inundación* ocurre en un periodo de tiempo corto.
- Un *huracán/tornado/tifón/tsunami/tormenta tropical/terremoto/inundación* ocurre de manera violenta.

Sin embargo, este procedimiento no soluciona enteramente el problema de la constitución de las clases semánticas y su justificación. Por una parte, no podemos tomar como único punto de referencia la base de datos EcoLexicon puesto que no todos los términos que encontramos en los argumentos de los verbos están representados en ella. Esto no es de extrañar, puesto que la base de datos EcoLexiCon no reúne aún la totalidad de los términos relacionados con el medio ambiente. Por otra parte, algunos sustantivos que aparecen en los argumentos de los verbos, como *edificio* o *país*, no pertenecen directamente a este ámbito y por lo tanto no están representados en EcoLexiCon. Para salvar este escollo, planteamos tests distribucionales basados en las propiedades sintáctico-semánticas de los sustantivos en cuestión. Nos inspiramos en trabajos sobre clases léxicas que aplican estos procedimientos como los de Gross (1994) o Flaux y Van Velde (2000).

Este es el caso de la clase LUGAR ARTIFICIAL. Estos sustantivos se caracterizan por necesitar de la acción humana como agente de su creación, por eso son compatibles con esta estructura:

- LUGAR ARTIFICIAL fue construido por...
- El *edificio/colegio/casa* fue construido por...

Los referentes de estos sustantivos se caracterizan por tener un modelo ideal previo a su existencia que constituye su máxima expresión. Es decir, que para que un ente llegue a la categoría de *edificio* debe cumplir unas características, entre ellas la de tener una finalidad propia. Desde el punto de vista lingüístico esto se manifiesta porque el sustantivo es compatible con estructuras que indican su funcionalidad:

- LUGAR ARTIFICIAL sirve para V
- El *edificio/colegio/casa* ha sido construido para trabajar/estudiar/vivir

Estos sustantivos suelen tener un poseedor, por lo que pueden insertarse en el sintagma nominal:

- EL LUGAR ARTIFICIAL de N (poseedor)
- El *edificio/colegio/casa* de Juan

Además, en ciertos casos, pueden actuar por metonimia como sujetos de verbos que requieren un agente humano:

- EL LUGAR ARTIFICIAL ha pensado/decidido/ordenado
- El *colegio/establecimiento/juzgado* ha pensado/decidido/ordenado

Estos N aparecen también con verbos de movimiento:

- Ir a/venir de LUGAR ARTIFICIAL
- Voy al/vengo del *colegio/establecimiento/juzgado*

Este tipo de tests son eficaces para descartar la inclusión de ciertos elementos dudosos en el grupo. Por ejemplo, los sustantivos *tienda de campaña* o *paseo marítimo* no verifican la totalidad de las proposiciones anteriores y por lo tanto no forman parte de la clase léxica LUGAR ARTIFICIAL:

- *La tienda de campaña fue construida por
- La tienda de campaña de Juan
- ?La tienda de campaña ha decidido que
- Voy a la /vengo de la tienda de campaña

- El paseo marítimo fue construido por
- *El paseo marítimo de Juan
- *El paseo marítimo ha decidido que
- Voy al /vengo del paseo marítimo

No obstante, es importante señalar que las fronteras entre clases léxicas no están tan claramente delimitadas como pueda parecer en un principio. Los límites entre una clase y otra son difusos. Es el caso de los sustantivos *tienda de campaña* y *paseo marítimo*, que se encuentran en la periferia de la clase LUGAR ARTIFICIAL. Además, hay que tener en cuenta que un sustantivo puede pertenecer a dos clases léxicas diferentes a la vez puesto que el significado es polifacético. Es el contexto lo que activa un aspecto semántico u otro de cada sustantivo. Por ejemplo, el sustantivo *playa* puede funcionar como LUGAR NATURAL o FORMACIÓN GEOLÓGICA en los ejemplos siguientes:

- Lisa ha tomado el sol en la playa.
- La playa se formó en una erupción volcánica.

Tampoco el procedimiento mediante el que se establecen los tests distribucionales es infalible puesto que estos no dejan de estar basados en cierta subjetividad del lingüista que los “crea” a partir de su propia intuición o conocimientos enciclopédicos. Con el fin de obtener una categorización lo más exacta posible y, sobre todo, de obtener unos resultados adaptados a la categorización que el propio discurso especializado establece de manera interna, combinamos este procedimiento con otro mediante el cual la categorización se extrae del propio corpus.

2.4. Categorías semánticas basadas en patrones de conocimiento

Los patrones de conocimiento (*knowledge patterns*) constituyen según algunos autores (Condamines 2002; Barrière y Abago 2006; Cimiano y Staab 2006) uno de los métodos más fiables para establecer relaciones semánticas entre conceptos. Nos basamos en la hipótesis de que los términos pertenecientes a una clase semántica atienden a los mismos patrones de relaciones conceptuales. Por ejemplo, dentro del campo de la sismología, los términos *erupción* y *terremoto* se comportan de manera similar dentro de la estructura causal “La erupción/el terremoto provoca N”, como en el enunciado “La erupción/el terremoto provoca daños/destrozos”. De esta manera, si localizamos todos los términos lexicalizados en una relación conceptual dada, podremos decir que pertenecen a la misma clase léxica. En el ejemplo anterior, los sustantivos *daños* y *destrozos* comportan los mismos rasgos semánticos y distribucionales y por lo tanto pueden agruparse dentro de una misma clase léxica.

El primer paso de este procedimiento es por lo tanto la búsqueda de los patrones de conocimiento dentro del corpus en los que aparece cada término. En primer lugar, estudiamos las relaciones conceptuales que el hiperónimo del término en cuestión mantiene con otros conceptos dentro la base de datos EcoLexiCon. Por ejemplo, una de las relaciones conceptuales del término *huracán* es la de hiperonimia (*is_a*) con la categoría léxica EVENTO EXTREMO (Huracán *is_a* extreme event). A continuación, observamos la representación en EcoLexiCon que las relaciones conceptuales de la categoría EVENTO EXTREMO mantiene con otros conceptos de la ontología son las siguientes:

- AN EXTREME EVENT causes HUMAN/ECONOMIC/MATERIAL LOOSES
- AN EXTREME EVENT affects THE ENVIRONMENT
- AN EXTREME EVENT occurs in a SHORT PERIOD OF TIME

De ahí, deducimos que estas relaciones conceptuales se lexicalizan en el corpus a través de enunciados como estos:

- Un huracán causa pérdidas humanas/económicas/ materiales.
- Un huracán afecta al medio ambiente.
- Un huracán arrasa de manera rápida.

El objetivo de este procedimiento no es otro que detectar cómo se lexicalizan en el corpus esas relaciones, con el fin de agrupar todos los términos que atienden a los mismos patrones de conocimiento bajo una misma categoría léxica. Por ejemplo, dentro de la meteorología, todos aquellos daños que provoca un *huracán*, como por ejemplo *muertes*, *pérdidas*, *destrucción* pertenecerán a una misma categoría, en este caso la de PÉRDIDAS HUMANAS/ECONÓMICAS/MATERIALES. Para averiguar cuáles son los sustantivos que forman parte de esta categoría, realizamos una búsqueda en el corpus del patrono [An EXTREME EVENT causes HUMAN/ECONOMIC/MATERIAL LOOSES]. Para realizar la búsqueda, debemos estudiar primero cómo se lexicaliza la causa dentro del corpus que estamos estudiando. El verbo de causalidad por antonomasia es *causar*. Este patrono se representa en SketchEngine de la siguiente manera: [lemma="tifón"] []{1,2} [lemma="causar" []{1,2} [tag="N.*"]. De esta manera, obtendremos una concordancia (figura 2) en la que observamos cuáles son los N que resultan de la acción de *tifón*. Podemos hacer lo mismo con otros sustantivos de comportamiento similar como *huracán*.

[lemma="tifón"] []{1,2} [lemma="causar" []{1,2} [tag="N.*"]		
lluvias en la isla de Hokkaido. En China et	tifón Talim ha causado importantes daños	y un centenar de víctimas mortales. </p>
trae entre manos... </p><p> "Parece que el Morakot" llegó el domingo por la tarde, el	tifón no causará ningún efecto	durante este fin de semana, ¡gracias a
Así se derrumbaba el hotel Chin Shuai. El	tifón ha causado la muerte	de tres personas y un desaparecido en las
paso por la isla filipina de Luzón </p><p> El	tifón Morakot ha causado en Taiwán	las peores inundaciones de los últimos
decretase la alerta roja y advirtiéndose de que	tifón Ketsana ha causado un centenar	de muertos y centenares de miles de desplazados
el	tifón podría causar severos daños	en varias áreas costeras de la zona. </p>
fuerte vendaval y las lluvias que arrastra et	tifón Shanshan ha causado hoy la muerte	a ocho personas y heridas a más de doscientas

[lemma="huracán"] []{1,2} [lemma="causar" []{1,2} [tag="N.*"]		
Sandy no pasó directamente sobre Haití, el	huracán ha causado estragos	en la empobrecida isla caribeña. Varios
antes del paso de 'Sandy'. Anteriormente, el	huracán había causado un muerto en Jamaica	. El ciclón avanza en dirección norte-noreste
los datos reflejados en su página web. El	huracán ha causado daños por un importe	de unos 50.000 millones de dólares (39.000
ha cobrado al menos 39 vidas. Asimismo el	huracán ha causado grandes daños	materiales. La elevación del nivel de las
Nueva York y Nueva Jersey. En Nueva York,	huracán ha causado muchas víctimas	, la mayoría de ellas fallecidas por árboles
et	huracán ha causado inundaciones	en cinco departamentos del sur de Haití
afectado muy gravemente las inundaciones.	huracán ha causado inundaciones en cinco	departamentos
El	huracán ha causado gran destrucción	del sur de Haití y en otras regiones como
afectado muy gravemente las inundaciones.	huracán ha causado pérdidas por más	en la ciudad más grande del mundo.
El	huracán ha causado ya las cancelaciones	Inundaciones
, pero incluso tras amainar su fuerza el	huracán, ha causado la muerte	de mil 100 millones de dólares en la zona
las redes sociales. Hasta el momento el		de numerosos vuelos y la suspensión de
todo el domingo para Nueva Inglaterra". El		de siete personas en la costa del Pacífico
tropical 'Jova', que hasta hace poco era un		

Figura 2. Concordancia de “tifón causa N” y “huracán causa N”

También es posible realizar una búsqueda más amplia en el corpus para observar los patrones de causalidad. Mediante la búsqueda: [tag="N.*"] [lemma="causar" []{1,2} [tag="N.*"] obtenemos una concordancia como la que se muestra en la figura 3.

el tercio norte de Filipinas, donde sus	lluvias torrenciales causaron las peores inundaciones
perdieron la vida ahogados en riadas y	corrimientos de tierra causados por los aguaceros
Las aguas de la inundación se llevaron	vehículos y causaron destrozos en edificios

Figura 3. Concordancia de “N causa N”

Además de esto, es necesario ampliar las estructuras de causalidad a otros verbos como *causar, provocar, originar, propiciar, favorecer, activar*. El resultado de estas búsquedas es que, a partir de estos patrones, observamos cuáles son los sustantivos que tienen una misma distribución y obtenemos así una visión más precisa de los miembros conforman cada categoría como observamos en la siguiente tabla.

DESASTRE NATURAL	huracán, lluvias torrenciales, corrimientos de tierra, inundaciones, aguaceros
PÉRDIDAS HUMANAS/ECONÓMICAS/MATERIALES	estragos, muertos, inundaciones, destrucción, pérdidas, destrozos

Tabla 6. Miembros de dos categorías conceptuales

En un proceso circular de retroalimentación, una vez que las categorías semánticas están más claramente definidas y sabemos cuáles son los sustantivos que las integran, podemos avanzar en el conocimiento del corpus para obtener nuevas categorías y patrones. Esto quiere decir que, si sabemos cuáles son los miembros de la categoría PÉRDIDAS HUMANAS/ECONÓMICAS/MATERIALES podemos estudiar de manera rápida qué tipo de pérdidas causa cada tipo de desastre natural mediante una búsqueda como esta:

```
*DUAL
=cuse_of/effect_of
1: [tag="N.*" [] {0,5} [lemma="causar"|"provocar"|"ocasionar" []]{0,5} 2: [tag="N.*"
1: [tag="N.*" [] {0,5} [lemma="causar"|"provocar"|"ocasionar" []]{0,5}
2[lemma="daño.*"|"víctima.*"|"muerte.*"|"muerto.*"
|"estrago.*"|"destrucción.*"|"inundaciones.*"|"desplazados.*"]
```

Tabla 8. Búsqueda de N que causan “PÉRDIDAS HUMANAS/ECONÓMICAS/MATERIALES”

Esta búsqueda nos da acceso, de manera automática, a una lista de las causas de cada tipo de desastre natural como vemos en la tabla 6 donde se muestran los N más

frecuentes de la estructura <Term> causa N. Esto nos permite obtener una mayor precisión que la descrita en la tabla 6 sobre los miembros que constituyen la categoría PÉRDIDAS HUMANAS/ECONÓMICAS/MATERIALES. En concreto, esta búsqueda añade a esta categoría los términos: *movimiento de terreno, tragedia, erosión, inundación, daño, alud, lodo, desbordamiento, daños*.

PÉRDIDAS HUMANAS/ECONÓMICAS/MATERIALES		
<i>tsunami</i>	<i>tormenta tropical</i>	<i>inundación</i>
movimiento de terreno	alud	desbordamiento
tragedia	lodo	daños
erosión	inundación	
inundación	muerte	
daño		
agua		
muerte		
viento		

Tabla 9. Resultados de <Term> causa N

Este tipo de información es útil desde al menos dos puntos de vista. Por una parte, nos da una información muy útil para profundizar en la semántica de cada término a través de su combinatoria léxica. Por ejemplo, observamos que los N que aparecen en el corpus como causas de *tsunami* son más frecuentes que los de *tormenta tropical* y estos, a su vez, más frecuentes que los de *inundación*. Esto indica que el espectro semántico de estas palabras es distinto y nos conduce a una descripción del significado de cada término más preciso:

Un *tsunami* suele actuar sobre el *terreno*, provocando *daños* materiales como *erosión, inundación, muerte* y daños no materiales como *tragedia* y *muerte*, por la acción del *agua* y el *viento*.

Una *tormenta* causa daños como materiales *inundaciones, muertes, aludes*.

Una *inundación* causa daños sobre todo debido *lluvias* y *desbordamientos*.

Desde la perspectiva de la lingüística aplicada, esta información nos permite recopilar los miembros de cada categoría semántica y, por lo tanto, definirlas con mayor precisión. A medio plazo, esperamos que esto nos ayude a predecir las traducciones de los verbos; si tenemos un repertorio de los sustantivos que aparecen más frecuentemente en los argumentos de los verbos un subdominio dado y conocemos las categorías a las que pertenecen esos sustantivos, será posible aislar de manera automática, a partir de un enunciado, los N de los argumentos, la categoría a la que pertenecen y, por lo tanto, su estructura actancial. A partir de ahí, y puesto que las estructuras

actanciales son equivalentes entre las lenguas (tabla 1), podremos obtener una lista de los verbos equivalentes en otras lenguas.

4. Conclusión y perspectivas

Hemos partido de la hipótesis de que es posible establecer equivalencias entre los verbos de distintas lenguas de un ámbito de especialidad tomando como punto de referencia su estructura actancial. Para definir esta estructura y poder verificar nuestra hipótesis, necesitamos en primer lugar contar con un repertorio de las clases conceptuales más frecuentes de cada subdominio de las Ciencias Ambientales.

El objetivo de este artículo ha sido exponer la metodología que seguimos para establecer la clasificación de las categorías semánticas más recurrentes de los diversos ámbitos de las Ciencias Ambientales. En concreto, hemos explicado cómo aislamos los verbos que aparecen con cada término de manera automática mediante la herramienta SketchEngine. Una vez que sabemos cuáles son los verbos asociados a cada verbo, estudiamos en el corpus los sustantivos de sus argumentos. En primer lugar, estas listas de sustantivos se clasifican de manera manual e intuitiva en categorías semánticas. A continuación, con el objetivo de asentar la identidad de cada categoría sobre principios sólidos y objetivos, definimos las características propias de cada una mediante una serie de tests distribucionales. Por último, averiguamos cuáles son los miembros que pertenecen a cada categoría del subdominio. Para ello, nos basamos en la hipótesis de que los sustantivos que comparten una misma distribución pertenecen a la misma clase semántica. Así, buscamos patronos recurrentes en el corpus en los que aparecen sustantivos en cuestión y observamos cuáles se comportan de la misma manera desde un punto de vista sintáctico-semántico. Hemos ilustrado esto con los patronos de causalidad. Las lexicalizaciones de las relaciones conceptuales sirven para averiguar los N que forman parte una estructura argumental. De esta manera, conseguimos tener criterios sobre las características de cada categoría y logramos refinar los sustantivos que pertenecen a cada una de ellas.

Las perspectivas de este trabajo son numerosas. La clasificación de los términos de los subdominios de las Ciencias Ambientales constituye una primera línea de trabajo y una ardua labor. Nuestro objetivo a largo plazo, una vez que obtengamos esta tipología y verifiquemos la hipótesis de la equivalencia de verbos basada en su estructura argumental, será establecer un sistema automático que permita la traducción automática de los verbos dentro de este campo de especialidad.

Miriam BUENDÍA-CASTRO
Pilar LEÓN-ARAÚZ
Beatriz SÁNCHEZ-CÁRDENAS

Referencias bibliográficas

- Baker, Collin F./Fillmore Charles J./Lowe, John B. The Berkeley FrameNet Project. Proceedings of the 17th international conference on Computational Linguistics, volume 1, pages 86-90.
- Bosque, Ignacio/Violeta Demonte (eds), 1999, Gramática descriptiva de la lengua española. Madrid: Real Academia Española, Espasa Calpe, § 1.1-7.
- Buendía, M. 2013. *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources*, Thèse de Doctorat, Granada, Universidad de Granada.
- Flaux Nelly/Danièle Van de Velde, 2000. *Les noms en français : esquisse de classement*, Paris, Ophrys.
- García-Miguel, J.M./F. González Domínguez/G. Vaamonde 2010. «ADESSE. A Database with Syntactic and Semantic Annotation of a Corpus of Spanish», *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, Valletta (Malta), 17-23 de mayo [http://www.lrec-conf.org/proceedings/lrec2010/pdf/859_Paper.pdf]
- Gross, G. 1994. «Classes d'objets et description des verbes», *Langages*, 115, 15-30.
- Kilgarriff A./P. Rychly/P. Smrz/D. Tugwell. «The Sketch Engine», Proceedings EURALEX 2004, Lorient, France.
- Kipper S./K. 2005. *VerbNet: a Broad-coverage, Comprehensive Verb Lexico* <<http://verbs.colorado.edu/~kipper/Papers/dissertation.pdf>>
- Miller G. A. 1990. «WORDNET : An on-line lexical database», *International Journal of Lexicography*, 3(4).