



DISEÑO DE UNA ONTOLOGÍA DE SEMÁNTICA LÉXICA EN LOS PROYECTOS MULTIGENERA Y MULTICOMB

DESIGN OF A LEXICAL SEMANTICS ONTOLOGY IN THE MULTIGENERA AND MULTICOMB PROJECTS

Rosa María Martín Gascueña
Universidad Complutense de Madrid
rosamartingascuena@filol.ucm.es

RESUMEN

El trabajo se enfoca en la presentación del diseño, desarrollo y funcionalidad de una ontología en los proyectos *MultiGenera*¹ y *MultiComb*², que se llevan a cabo en múltiples idiomas (español, alemán y francés) con el propósito de generar de manera automática frases nominales (FN) y sus contextos oracionales. La ontología juega un papel fundamental en estos proyectos al formalizar las propiedades esenciales de los elementos léxicos siguiendo principios lingüísticos teóricos de la gramática y de la lexicografía valencial. Esta ontología proporciona etiquetas de marcado semántico para las clases semánticas que conforman el entorno sintagmático de veinte sustantivos de diversas áreas de conocimiento. Su objetivo es organizar los datos léxicos para su almacenamiento en una base de datos y su posterior recuperación a través de las aplicaciones prototipo *Xera* y *Combinatoria*. En las que se generan automáticamente las estructuras nominales de los sustantivos y su combinación en el contexto oracional.

El proceso de elaboración de la ontología se divide en dos fases: en la primera, se utilizan las mismas clases semánticas que en *Portlex* para clasificar los actantes de cada nombre empleados para la generación automática de FN en las tres lenguas en *Xera* 1.0. En la segunda fase, se reconfigura la ontología al conectar automáticamente con *synsets* de WordNet y otras ontologías relacionadas para aumentar la base de datos léxica de los proyectos. Esta versión posterior de la ontología presenta una mayor granularidad, lo que facilita la programación para generar los entornos de los nombres y sus combinaciones en contextos oracionales. La ontología se aplica para aprender idiomas gracias a su interfaz amable en las diversas aplicaciones. Los usuarios pueden utilizarla para seleccionar posibles combinaciones léxicas en FN. Es importante señalar que la ontología está disponible en línea y sigue en constante desarrollo.

Palabras clave: lexicografía, ontologías, categorías semánticas, frase nominal.

ABSTRACT

The aim of this paper is to present the design, development, and functionality of the ontology in the *MultiGenera* and *MultiComb* multilingual projects, (Spanish, German, and French) for the automatic generation of the noun phrases (FN) and its sentence contexts. It is important to highlight the primordial role of this ontology for the development of these, as it involved a formalisation of the fundamental properties of lexical items in accordance with valential grammar and lexicography. The ontology provides semantic tagging labels for the semantic classes that make up the syntagmatic environment of the twenty nouns, belonging to different areas of knowledge. Its purpose is to organise the lexical data for storage in the database and subsequent retrieval through the prototype applications *Xera* and *Combinatoria*, which automatically generate the nominal structures of the nouns and their combination in sentence contexts.

The ontology development process is divided into two phases: in the first phase, the same semantic classes as in *Portlex* to classify the actants of each noun used for the automatic generation of FN in the *Xera* 1.0. in the three languages. In the second phase, the ontology is reconfigured by automatically connecting to the WordNet database *synsets* and other related ontologies to increase the lexical database of the projects. The latter ontology presents greater granularity, which favours programming for the generation of name environments and their sentence combination. The application of the ontology in language learning is shown by the user-friendly interface in various applications to select the possible lexical items combinations in FN. It should be noted that the ontology is still under development and is available online.

Keywords: lexicography, ontologies, semantic categories, noun phrase.

¹ *MultiGenera*. Generación multilingüe de estructuras argumentales del sustantivo y automatización de extracción de datos sintáctico-semánticos. Fundación BBVA. Ayudas Fundación BBVA a Equipos de Investigación Científica - Humanidades Digitales. 2017-2020. <http://portlex.usc.gal/multigenera/>

² *MultiComb*. Generador multilingüe de estructuras argumentales del sustantivo con aplicación en la producción en lenguas extranjeras. FI2017-82454-P: Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia, Generación de Conocimiento. MCIN/ AEI/ FEDER "Una manera de hacer Europa" (EXCELENCIA 2017, 2017-PN091). 2018-2021. <http://portlex.usc.gal/multicomb/>



1. INTRODUCCIÓN

El objetivo de este trabajo es presentar la ontología de los proyectos lexico-gráficos, multilingües (español, alemán y francés), consecutivos y complementarios *MultiGenera* y *MultiComb*, disponibles en la red. Al comenzar el estudio, las primeras cuestiones que se plantearon fueron definir el concepto ontología en esta investigación, cómo sería su elaboración y cuál sería su utilidad en el desarrollo de los proyectos.

Se revisaron diferentes definiciones: etimológicamente, el término *ontología* significa el estudio del Ser, aunque a finales del siglo XX, adquirió otra significación en el ámbito informático y en la documentación. En el primero, una ontología designa una colección de objetos diferentes relacionados (Lenat & Guha, 1990)³ y se define como “An *ontology* is an explicit specification of a conceptualization” (Gruber, 1993, p. 1), para otros autores “una ontología es una especificación explícita y formal de una conceptualización compartida” (Borst, 1997; Studer, Benjamins & Fensel, 1998, p. 25); una ontología es un tipo especial de objeto de información o artefacto computacional (Guarino, Oberle & Staab, 2009). En el Procesamiento del Lenguaje Natural (PLN) y la Inteligencia Artificial (IA), las ontologías se emplean para modelizar formalmente la estructura de un sistema, las entidades y sus relaciones. Y para algunos autores, la base de datos WordNet (Miller, Beckwith, Fellbaum, Gross & Miller, 1990) es una ontología léxica utilizada como modelo para crear otras ontologías como EuroWordNet⁴ (Vossen, 1998) (vid. 3). Por otro lado, en el ámbito documental, una ontología es una tecnología

³ Lenat y Guha (1990) en el proyecto CYC, sobre inteligencia artificial que combina bases de datos y ontología: “the ontology of Cyc is organized around the concept of categories. We also refer to them as classes or collections. The categories are organized in a generalization-specialization hierarchy” (1990, p. 42).

⁴ La ontología en EuroWordnet (Vosse, 1998) se define como un modelo abstracto, una estructura interpretable por el computador, que define las relaciones e implicaciones necesarias para modelar conceptualmente un área de conocimiento.

para los sistemas de información. Algunos conciben la ontología como un tesoro: “una lista controlada y estructurada de términos para el análisis temático y la búsqueda de documentos”⁵ o como una jerarquía taxonómica cuyos componentes principales son las clases (p. e. *mamíferos*), las subclases (p. e. *gatos*) y los individuos o las instancias (p. e., *siamés*) más sus relaciones, que deben estar codificadas en un lenguaje informático con una base lógica para especificar las relaciones e inferencias entre ellas, y además, pueden centrarse en un área de conocimiento concreto (Codina & Pedraza-Jiménez, 2011).

La definición adoptada aquí considera la ontología como un modelo conceptual abstracto que describe las relaciones e implicaciones léxicas necesarias para modelar áreas de conocimiento y, es interpretable computacionalmente. En otras palabras, una ontología es un constructo de categorías semánticas relacionadas dentro en un dominio.

La motivación para elaborar y desarrollar una ontología propia estaba justificada, porque, en principio, ninguna otra ontología de libre acceso podía dar cuenta de las restricciones combinatorias de los sustantivos en la frase nominal (FN), tal como se plantea en la gramática y lexicografía valencial (Domínguez Vázquez, 2011; Domínguez Vázquez, Valcárcel Riveiro & Lindemann, 2018) base de la investigación. Así pues, se decidió crear una ontología a medida para el tratamiento computacional de los datos léxicos, que, además, pudiera emplearse con fines didácticos en las aplicaciones lexicográficas de los proyectos y que en un futuro sirva como índice interlingüístico entre las tres lenguas.

La ontología inicial es diseñada *ad hoc*; a partir de las categorías semánticas definidas en el proyecto anterior⁶. Se emplea para organizar el léxico que

⁵ Thesaurus de la Unesco <http://vocabularies.unesco.org/browser/thesaurus/en/?clang=es> [consulta 20/05/2021]

⁶ Proyecto anterior: *Portlex*.

nutre las aplicaciones de generación automática de las frases nominales. De este modo, se concibe una estructura de clases o categorías semánticas que contribuyen a formalizar las relaciones sintagmáticas en la FN y generar automáticamente la combinatoria del sustantivo a partir de esas categorías semánticas. Posteriormente, la ontología se enlaza automáticamente con WordNet y otras ontologías relacionadas con ella para incrementar los elementos léxicos de las clases semánticas, como se explicará en el apartado 4. Esto conlleva un cambio en la configuración inicial de la ontología: favorecer la formación del léxico en las aplicaciones *Xera* y *Combinatoria* como se explicará en el apartado 4.2.

El empleo de ontologías en aplicaciones informáticas es frecuente, aunque no tanto el hecho de que las clases semánticas de la ontología se muestren en la interfaz del usuario, como en los prototipos *Xera* y *Combinatoria*, para que este elija entre las posibles opciones y así se generen automáticamente frases nominales y su contexto oracional en varias lenguas. Esta propuesta novedosa es de gran utilidad para la enseñanza de lenguas y la traducción automática.

Este trabajo se divide en varios apartados, el 2 presenta el marco teórico del que parte la ontología; el 3 indica las ontologías léxicas con las que se enlazarán automáticamente y presenta trabajos relacionados; en el 4 se expone la metodología empleada para el diseño de la ontología y las aplicaciones; el 5 expone las conclusiones y, para finalizar; en el 6 están las referencias bibliográficas.

2. FUNDAMENTOS LINGÜÍSTICOS DE LA ONTOLOGÍA

En este apartado se resumen los fundamentos lingüísticos teóricos de esta investigación de los que se parte para la creación de la ontología. Estos residen en la gramática y lexicografía valencial (Domínguez Vázquez, 2011; Domínguez Vázquez, Valcárcel Riveiro & Lindemann, 2018), los principios

de categorización, la teoría del prototipo (Rosch, 1978), los corpus de Sketch Engine⁷ y el concepto de *synsets* de WordNet (1987).

2.1. EL ANÁLISIS VALENCIAL DEL SUSTANTIVO

La valencia del sustantivo se identifica con sus características sintáctico-semántico particulares. Esto implica que cada sustantivo selecciona unos elementos léxicos, denominados *actantes*, y cada actante desempeña una función semántica (agente, experimentante...) en consonancia con unas propiedades semánticas determinadas; por ejemplo, el sustantivo *dolor* tiene un significado que se relaciona con un actante (experimentante) con el rasgo semántico [animado] como *el dolor de su madre*. La ontología se concibe para formalizar semánticamente el análisis valencial de los sustantivos (Domínguez Vázquez, 2011).

En estos proyectos se ha trabajado con veinte sustantivos de cinco áreas léxico-conceptuales diferentes (Tabla 1): *Expresión*, *Clasificación*, *Afección*, *Locación-situación*, *Locación-dirección*, según la clasificación de un proyecto anterior para un diccionario multilingüe sobre la valencia del nombre (*Portlex*).

Expresión	Clasificación	Afección	Locación-situación	Locación-dirección
1. Pregunta	1. Olor	1. Amor	1. Presencia	1. Huida
2. Respuesta	2. Sabor	2. Aumento	2. Ausencia	2. Viaje
3. Conversación	3. Color	3. Dolor	3. Estancia	3. Mudanza
4. Discusión	4. Ancho	4. Muerte		
5. Texto				
6. Video				

TABLA 1: Áreas léxico-conceptuales para los sustantivos en español

La valencia de cada sustantivo en algunos casos coincide en las tres lenguas, pero en otros puede presentar peculiaridades formales distintas, aunque

⁷ <https://www.sketchengine.eu/>

compartan la misma perspectiva onomasiológica. El comportamiento actancial de cada sustantivo se analiza indicando los roles semánticos que desempeñan sus argumentos. En el esquema actancial de los sustantivos se indican los actantes o argumentos que se completan con unidades léxicas con unos rasgos categoriales determinados: [Animado] [Humano]... Así, por ejemplo, el sustantivo *amor*, que pertenece al *área de la afección*, presenta una valencia de tres posibles actantes A1 (1), A2 (2), A3 (3), cuyos roles semánticos son para A1: aquel / aquello que experimenta un nuevo estado o situación, agente; A2: aquel/aquello que tiene o dispone de algo; A3: aquel/aquello no afectado, tema.

- (1) El intenso amor de Pepe. (A1) [Animado] [Nombre Propio]
- (2) Un sincero amor por la humanidad. (A2) [Animado] [Humano]
- (3) Un profundo amor a los libros. (A3) [Inanimado][Objeto]

Después de analizar el comportamiento actancial de todos los sustantivos se hicieron búsquedas en los corpus de Sketch Engine para encontrar las palabras más frecuentes que cubrían los diferentes actantes y a partir de ahí categorizar los prototipos léxicos. Este proceso fue muy importante para la construcción de las categorías y relaciones semánticas de la ontología (vid. 4.1.1).

2.2. EL PROCESO DE CATEGORIZACIÓN SEMÁNTICA

Lakoff (1987, p. 5) indica que: “There is nothing more basic than categorization to our thought, perception, action, and speech. Every time we see something as a kind of thing, for example, a tree, we are categorizing”. La categorización es una de las operaciones lingüísticas de conceptualización que intervienen en la construcción del significado. Es una actividad mental que implica organizar, estructurar y agrupar elementos que comparten rasgos de significado conceptual mediante propiedades y funciones. Al categorizar se produce un ajuste de nuestro foco de atención hacia ciertas características ignorando otras y así, creamos categorías conceptuales que se definen como

construcciones teóricas abstractas formadas por unidades léxicas con propiedades comunes. Así, *discusión*; y *texto*; se incluyen dentro del área conceptual de la *expresión* (Tabla 1). Estas clasificaciones representan la forma en que articulamos nuestra experiencia del mundo para hacerlo manejable. Una categoría conceptual incluye conceptos y a su vez, en sí misma, es un concepto, desde la perspectiva onomasiológica (Martín-Gascueña, 2013, p. 90).

La categorización es un fenómeno cognitivo relacionado desde la lógica, con la extensión e intensidad del significado. El punto de vista extensional indica continuidad del significado, relacionado con la categorización, y parte de presupuestos lógicos para definir las relaciones de inclusión de significados (hiponimia y meronimia). La relación por excelencia en la creación de taxonomías es la hiponimia, caracterizada por ser implicativa, transitiva y asimétrica (Lyons, 1977, p. 274; Cruse, 1986, 2002, 2004; Brown, 2002). Por ejemplo, si *texto* está incluido en la categoría⁸ de *expresión*, entonces todos los *textos* son *expresión*, pero no toda *expresión* es un *texto*; esto indica que ser un *texto* implica ser una expresión, aunque no al revés, los *textos* son un grupo dentro de la expresión, luego es una relación implicativa y asimétrica. En cuanto a la propiedad transitiva, está basada en la capacidad de contención de los significados y está condicionada por el contexto (Croft & Cruse, 2004). Por ejemplo, si *texto* es un hipónimo de *expresión* y *expresión* es un hipónimo de *comunicación*, entonces *texto* es un hipónimo de software *comunicación*. Asimismo, la intensidad está relacionada con la composicionalidad del significado. La intensidad implica una inclusión de significados de las unidades léxicas de las propiedades de niveles superiores o hiperónimos en las de niveles inferiores o hipónimos, caracterizados por heredar sus características y distinguirse por tener algún rasgo más. La combinación de extensión e intensidad permite generar inferencias e implicaturas de significación.

⁸En este trabajo *categoría* y *clase semántica* son equiparables y se utilizarán indistintamente.

Las categorías conceptuales o semánticas en la teoría clásica se definen por una lista de condiciones necesarias y suficientes que deben cumplir cualquiera de sus miembros para pertenecer a ella. Sus límites están definidos y la relación entre sus miembros es lineal. En la teoría del prototipo (Rosch, 1978), las categorías interesan por su organización interna en torno al prototipo, que es el mejor representante, cuya referencia sociocultural puede corresponderse con un ser real o ideal y se identifica con el Modelo Cognitivo Ideal (MCI) que es una estructura abstracta que interviene en los procesos de construcción del significado (Lakoff, 1987). Esta organización categorial determina las relaciones intercategorial e intracategorial de las unidades léxicas. Las clases semánticas organizan el conocimiento léxico en la ontología.

2.3. LA BASE DE DATOS WORDNET

La base de datos del inglés WordNet se creó en 1985 por un grupo de psicolingüistas de la universidad de Princeton dirigidos por Miller y es considerada por algunos investigadores como una ontología léxica a pesar de que no fuera concebida como tal. Esta base de datos léxica está fundamentada en los modelos de memoria léxica basados en la semántica de Lyons (1977). La idea original era probar que el uso de diccionarios conceptuales era mejor que el simple orden alfabético. La característica más ambiciosa de WordNet es el intento de organizar la información léxica, desde la perspectiva onomasiológica, por los significados de las palabras, más que por su forma o visión semasiológica.

En esta base de datos, las palabras se diferencian por categorías gramaticales: sustantivos, adjetivos, verbos; su significante, forma física, y significado, referido a un concepto que puede estar representado por una o varias palabras. De este modo, el significado de una palabra se representa por un grupo de sinónimos denominados *synsets* que simbolizan conceptos, son *nodos* en la red léxica unidos por relaciones semánticas horizontales, sinonimia y verticales, hiperonimia y meronimia. El modelo de red de significados agrupados

en *synsets* por relaciones de inclusión y sinonimia también está presente en nuestra ontología.

WordNet representa un modelo de organización léxica, donde los sustantivos se clasifican en jerarquías, lo que implica un sistema de herencia de significados. Así pues, se establecen varios niveles de categorización: en el nivel más alto de la jerarquía está el significado más abstracto, la {entidad}, en el nivel inmediatamente inferior se incluye la categoría {objeto, cosa} e {idea}, a continuación, la clasificación entre *animados* e *inanimados* y posteriormente, se identifican con alguno de los veinticinco primitivos semánticos que funcionan como categorías conceptuales o semánticas cerradas (Tabla 2).

Primitivos Semánticos		
{act, action, activity}	{food}	{process}
{animal, fauna}	{group, collection}	{quantity, amount}
{artifact}	{location, place}	{relation}
{attribute, property}	{motive}	{shape}
{body, corpus}	{natural object}	{state, condition}
{cognition, knowledge}	{natural phenomenon}	{substance}
{communication}	{person, human being}	{time}
{event, happening}	{plant, flora}	
{feeling, emotion}	{possession}	

TABLA 2: Los primitivos semánticos en WordNet (Miller, Beckwith, Fellbaum, Gross & Miller, 1990)

La posibilidad de descargar y utilizar la base de datos WordNet con todo su entramado de relaciones semánticas ha permitido que muchos proyectos la utilizaran como una ontología léxica, aunque presentase algunas inconsistencias en las categorizaciones (Gangemi, Navigli & Velardi, 2003), que se fueron corrigiendo en sucesivas versiones como WordNet 1.7 y posteriores. En consecuencia, se ha transformado en una ontología léxica para representar el conocimiento. Distingue entre relaciones de subclases e individuos o instancias, y asocia identificadores únicos a cada categoría o clase (Martin,

2003). WordNet es considerada por algunos una ontología superior, porque las categorías propuestas en sus grupos de *synsets* y sus relaciones son modelos compartidos con otras ontologías que además han compartido datos. En consecuencia, WordNet funciona como un Índice Interlingüístico (ILI) entre otras ontologías de diferentes idiomas, facilitando la interoperabilidad semántica en la definición de términos.

Muchos proyectos utilizan WordNet como fuente para recuperar información y también para desambiguar significados. En muchos casos, sus categorías se han convertido en una especificación formal para extraer automáticamente asociaciones de información léxica de WordNets en distintas lenguas e interpretarlas como un conjunto de relaciones conceptuales (inclusión y sinonimia), formalmente definidas en la ontología. Este es el caso del proyecto del Istituto di Scienze e Tecnologie della Cognizione DOLCE⁹ (Descriptive Ontology for Linguistic and Cognitive Engineering), creado por Masolo, Borgo, Gangemi, Guarino y Oltramari (2003).

WordNet ha sido un elemento clave para el desarrollo computacional de algunas ontologías de semántica léxica que se verán a continuación y con las que se ha enlazado automáticamente la ontología. Ninguna se consideró totalmente apropiada para basar los proyectos en ellas porque en principio, no respondía a las necesidades del análisis valencial (vid. 2.1) y se optó por diseñar y elaborar una propia, que partía de la investigación de *Portlex*.

3. ONTOLOGÍAS DE SEMÁNTICA LÉXICA

WordNet ha sido un elemento clave para el desarrollo de algunas ontologías de semántica léxica de código abierto con las que se ha enlazado la ontología de esta investigación, a través de la WordNet del gallego: Galnet (Solla Portela & Gómez Guinovart, 2015). Así las clases semánticas

⁹ DOLCE <https://www.istc.cnr.it/it/content/dolce-descriptive-ontology-linguistic-and-cognitive-engineering>

han incrementado su léxico al conectar automáticamente con las ontologías que presentamos a continuación por orden cronológico de creación (vid. 4.1.2).

3.1. TOP ONTOLOGY

Top Ontology (TOP) es una jerarquía independiente de clases, diseñadas para agrupar, comparar e intercambiar conceptos entre idiomas en EuroWordNet¹⁰ (Rodríguez, Climent, Vossen, Bloksma, Peters, Alonge, Bertagna & Roventini, 1998). Está basada en relaciones semánticas de inclusión y sinonimia como en WordNet 1.3, aunque la categorización semántica del léxico se basa en el lexicon generativo de Pustejovsky (1995, p. 76). En concreto, en la estructura de *qualia* o modos de explicación, formada por diferentes roles o *qualia: formal*, representa la información que distingue el contenido de una palabra de otras relacionadas con ella, dentro de su dominio correspondiente; *constitutivo*, recoge la información sobre la entidad y sus partes; *télico*, especifica la finalidad de la entidad y *agentivo*, detalla los factores que originan la existencia de la entidad.

La TOP es un repositorio de información de semántica léxica que ha aumentado con más rasgos los *synsets* de WordNet y el resultado es WordNet 1.6. La TOP está programada con el lenguaje de marcado OWL (Web Ontology Language), y es una jerarquía que consta de 63 características organizadas en tres tipos diferentes de entidades: 1stOrderEntity: cosas físicas (imagen); 2ndOrderEntity: eventos, estados y propiedades (imagen); 3rdOrderEntity: entidades no observables (Vossen, 1998).

Esta ontología no respondía a la división en área conceptuales que se plantea en la ontología de los proyectos para facilitar el tratamiento computacional.

¹⁰ EuroWordNet (1996) es una base de datos léxica, multilingüe para el neerlandés, el italiano y el español, cada lengua posee sus propios WordNets enlazados con la base de datos WordNet 1.5, unidos por un índice interlingüístico ILI de significados que es la ontología superior, denominada Top Ontology.

3.2. SUGGESTED UPPER MERGED ONTOLOGY

Suggested Upper Merged Ontology (SUMO), diseñada por Teknowledge Corporation (Pease, Niles & Li, 2002), es una ontología de comunicación computacional básica para distintos sistemas informáticos de procesamiento de información. SUMO es un esquema jerárquico de clases o categorías, reglas de relación y relaciones que establece enlaces con los *synsets* de WordNet 1.3. En principio, era una ontología de nivel general, que incluía niveles medios de especificación. Se fue ampliando con otras muchas ontologías de dominio específico. Un hecho que hay que resaltar es que la nomenclatura empleada en la ontología sigue estándares que le permiten emplear los mismos significados y maximizar su compatibilidad. Por este motivo la SUMO permite la interoperatividad con otras aplicaciones de razonamiento automatizados.

Esta ontología de carácter general fue concebida para la comunicación entre máquinas, no para establecer clases o categorías con mayor especificidad que permitieran la programación con mayor granularidad.

3.3. WORDNET DOMAINS

WordNet Domains, desarrollado en el Instituto per la Ricerca Scientifica e Tecnológica (Bentivogli, Forner, Magnini & Pianta, 2004), es un recurso léxico creado de forma semiautomática. Posee un conjunto de unas doscientas etiquetas de *dominios básicos* con un grado de granularidad adecuado para su procesamiento computacional aplicado a la categorización de textos y a la desambiguación del sentido de las palabras. Asimismo, WordNet se incrementa al anotar sus *synsets* con las etiquetas de dominio semántico de WordNet Domains Hierarchy (WDH). Estos dominios pueden incluir *synsets* de distintas categorías gramaticales, de diferentes niveles jerárquicos de WordNet y reunir varios sentidos de la misma palabra en clusters homogéneos, para reducir la polisemia de palabras en WordNet (Bentivogli, Forner, Magnini & Pianta, 2004).

WordNet Domains realiza el mapeo entre sus dominios básicos, los temas de WordNet 1.5 y las categorías emergentes de Wikipedia. De este modo, se

consigue una alineación aproximada entre WordNet y Wikipedia, útil para producir corpus multilingües y específicos de dominio. El multilingüismo se logra a través de enlaces entre las categorías de Wikipedia en diferentes idiomas. WordNet Domains se ha integrado en la base de datos léxica multilingüe MultiWordNet. En un principio enlazaba la WordNet del italiano con la versión 1.3 de WordNet, posteriormente se han incluido más WordNets del español, portugués, hebreo rumano y latín (Pianta, Bentivogli & Girardi 2002) y se distribuye bajo una licencia Creative Commons Attribution.

Enlazar con los dominios de WordNet excedía las dimensiones de *Multi-Genera* y *MultiComb*, ya que no se centraba en ningún dominio conceptual específico, sino que tenía un carácter de clasificación de léxico general en varias lenguas.

3.4. GALNET

Galnet es una WordNet para el gallego que, mediante un índice interlingüístico enlaza con WordNet 3.0 en inglés y con otras ontologías como la TOP, la SUMO y WordNet Domains. Hay una transferencia de conocimiento de WordNet a Galnet, se estructura igual mediante relaciones de inclusión (hiponimia y meronimia) y sinonimia. A los *synsets* nominales se les denomina *epinónimos* y representan la categoría de un área semántica, a la que se asignarán automáticamente otros *synsets* mediante algoritmos que evaluarán su proximidad a través del tratamiento terminológico de las relaciones léxico-semánticas (Solla Portela & Gómez Guinovart, 2015). Galnet facilita la interoperatividad semántica entre las distintas ontologías indicadas, mediante la conversión de los formatos con el lenguaje de marcado LMF (Lexical Markup Framework) y la compatibilidad con los modelos OWL existentes, con lo que se consigue la conexión entre diferentes capas de información, permitiendo reutilizar datos.

La ontología de esta investigación ha expandido el léxico de sus clases semánticas gracias a Galnet, que ha permitido mediante su índice interlingüístico enlazar con todas las ontologías antes mencionadas.

En los siguientes apartados presentamos cómo se han utilizado estas ontologías y la influencia que han ejercido en la evolución de la ontología semántica de los proyectos lexicográficos cuyas aplicaciones *Xera* y *Combinatoria* crean ejemplos a partir de patrones argumentales valenciales y son prototipos para nuevos modelos de diccionarios plurilingües automáticos y dinámicos.

4. METODOLOGÍA

La cuestión que se planteó era cómo diseñar una ontología para estos proyectos que diera cuenta, desde la perspectiva semasiológica y onomasiológica, de las características semánticas de los prototipos léxicos para los diferentes sustantivos, en alemán, francés y español. La dificultad radicaba en establecer las etiquetas de las clases semánticas para almacenar y recuperar los datos léxicos, y de este modo, crear paradigmas semánticamente coherentes para la generación automática de la FN con cierta autonomía del contexto. Y, además, que, en un futuro, sirvieran como índice interlingüístico entre las distintas lenguas de los proyectos, de tal forma, que a través de la ontología se pudiera establecer equivalencias entre las lenguas para una aproximación comparativa y contrastiva del funcionamiento de estos sustantivos en alemán, español y francés.

Gangemi, Navigli y Velardi (2003) proponen distintos modelos para la construcción de una ontología, dependiendo de su utilidad; el enfoque *top down*, descendente, es el más adecuado para elaborar ontologías generales. En una estructura jerárquica se parte de los nodos superiores independientes de un dominio. El enfoque *bottom up*, ascendente, es para ontologías terminológicas, se intenta llegar a un nodo más general a partir de nodos locales. Y el modelo *híbrido* que intenta aprovechar los dos enfoques anteriores, descendente y ascendente. Asimismo, la construcción de la ontología puede completarse con otras ya existentes, según sea su finalidad conectando con ellas automáticamente. Hay ontologías construidas con modelos de funciones

léxicas para unir diferentes bases de datos y ontologías para comunicarse computacionalmente¹¹.

La ontología del proyecto es de carácter general y se pensó como una estructura conceptual que sirviera para formalizar el tratamiento computacional, tanto para el almacenaje como la recuperación y la programación de los sustantivos y sus actantes.

En el proceso de elaboración de la ontología, se siguió un modelo *híbrido*; en la primera fase, se siguió un modelo ascendente, *bottom up*, a partir de los datos léxicos se crearon y denominaron las clases semánticas en la ontología 0.1. En la segunda fase, el hecho de conectarse con otras ontologías para aumentar el léxico de los paquetes semánticos de los sustantivos obligó a revisar las clases de la ontología 0.1 para favorecer la conexión computacional. El modelo de referencia fue WordNet para incrementar los niveles categoriales y las clases semánticas, por este motivo, lo denominamos enfoque descendente, *top down*.

4.1. ENFOQUE ASCENDENTE

La ontología se concibió como una estructura jerárquica de categorías semánticas para organizar el léxico, facilitar el almacenaje, la recuperación y la programación de las aplicaciones. Siguiendo este enfoque *bottom up*, se decidió construir las clases semánticas a partir de los datos obtenidos en los corpus

¹¹ Actualmente, se trabaja en la creación de estándares internacionales para la representación de lexicones que posibiliten la comunicación entre máquinas y así permitan, unir datos y compartir recursos. Este es el caso del proyecto LEMON, Lexicon Model for Ontologies, que propone un modelo para modelar el léxico y hacer diccionarios legibles para la máquina, no contiene datos lingüísticos, sino que establece definiciones formales para la creación de diccionarios monolingües, inicialmente, en formato LMF, Lexical Markup Framework. A su vez, el proyecto Ontolex-Lemón (Bosque-Gil & García, 2019) propone un modelo de estandarización para la representación de lexicones computacionales, unido a las tecnologías OWL (Web Ontology Language) para obtener significado y RDF (Resource Description Framework), que relaciona datos por ejemplo triples <books> <haswriter> <writer> Así se consigue compartir los recursos léxicos fácilmente y construir otros recursos léxico-semánticos multilingües de acuerdo con estas tecnologías de Web Semántica y la nube de datos abiertos vinculados, para establecer equivalencias entre ellos para que puedan ser compartidos y reutilizados en las comunidades científicas internacionales (Bosque-Gil, 2019).

de Sketch Engine y así, establecer los niveles categoriales generales y las etiquetas semánticas, de acuerdo con el esquema básico de catorce rasgos (Domínguez Vázquez, 2011) utilizados en *Portlex* como puede verse en la Tabla 3.

Abreviaturas	Significado categorial
<i>mat</i>	material (concretos)
<i>anim</i>	animado
<i>hum</i>	humano
<i>zool</i>	animal
<i>inst</i>	instituciones
<i>inanim</i>	inanimado
<i>mas</i>	masa (concretos no contables, como por ejemplo el <i>agua</i>)
<i>obj</i>	objeto (inanimado contable)
<i>plant</i>	planta
<i>inmat</i>	inmaterial (abstractos)
<i>intel</i>	conceptos delimitables y contables (por ejemplo, <i>idea</i>) y sistemas reglados jerárquicos (por ejemplo, los <i>días</i> , <i>meses</i> , <i>comunismo</i> , etc.)
<i>situ</i>	situación
<i>situest</i>	situación estática
<i>situdin</i>	situación dinámica

TABLA 3: Categorías semánticas propuestas en *Portlex*

Los veinte sustantivos de las tres lenguas se analizaron valencialmente para obtener su esquema actancial (vid. 2) con el que se realizaron las búsquedas en los corpus de Sketch Engine (Figura 1), para extraer información mediante las *queries* o consultas sobre los actantes de cada sustantivo y así obtener las unidades léxicas más frecuentes y representativas de cada casilla funcional, o sea, sus prototipos semánticos. Los resultados obtenidos se descargaban en un fichero CSV (Comma-Separated Values), o XML y se etiquetaban semánticamente según las catorce categorías de la Tabla 3, de este modo se configuran los niveles categoriales de la ontología.

CONCORDANCE Spanish Web 2011 (esTenTen11, Eu + Am)

CQL [lemma="amor"] [word="de"] [tag="N.C."] • 909
0.08 per million tokens • 0.0000083%

Details Left context KWIC Right context

1	elmundo.es	a su hermana Blanca. Pero esta historia de	amor de final	abierto es sólo es primero de los proyectos que tiene
2	unav.es	dio. La Radio ha sido trezada en el aire con	amores de oyentes	y profesionales". Y lo sigue haciendo.
3	uclm.es	rupu Los Bravos. Se trata de una historia de	amor de final	trágico, y con cierto tono orie basada en la novela de
4	mtv.es	inglés, Fly/Forget . En este álbum se une el	amor de Ivri	por la música electrónica, con unos cálidos temas acú
5	guzmanurrero.es	, de la otra o, lo más seguro, del cuento. Los	amores de Stendhal	son tema de conversación, aunque retaceado, con su
6	cinetube.es	Rodrigo, desesperado, desprecia la confesión de	amor de Leonarda	y se marcha a Europa, de donde regresará casado co
7	anagrama-ed.es	al devenir del mundo», viajará a encontrarse con un	amor de juven	tud. Fernández, entonces veinteañero, le seg
8	be2.es	pañan al corto en todos los bares. Si un test	amor de Internet	no te convence para buscar pareja o para hacer conte
9	revistadelibros...	decir de Carlos Gardel, esto último no eran más que	amores de estudiante	(ahora una promesa, mañana una traición), la Razór
10	elmundo.es	que si comprensivo... feo como un demonio, pero un	amor de terapeuta	. Un día, sin previo aviso, Eugenio le dio el al
11	buscoenlaces.es	allí sus estudios, quien veía una semejanza entre el	amor de Rizal	por Filipinas y el suyo por Vasconia, El País Vasco. </
12	elmundo.es	inquietante, iba también a poder escapar, gracias al	amor de Milú	, de la vida gris de su familia. Una vida «sin e
13	moonfleet.es	chica numero 6 yo misma" bueno opino que hacer el	amor de parte	de nosotros los hombre a ustedes las mujeres. opino
14	bookscenter.es	el sabor de sus vinos y manjares. Gracias al	amor de Manfredi	por los detalles y a su habilidad para unir pasado y pr
15	20minutos.es	las intituciones publicas de la epoca hasta cartas de	amor de particulares	. Para muchas personas esos papeles signifí

FIGURA 1: Búsqueda del actante 1 (A1) del sustantivo amor

Los nombres y adjetivos se ordenaron por frecuencias y etiquetaron dentro de tres niveles categoriales, el más abstracto, la dicotomía *material* e *inmaterial*. Así se obtuvieron los prototipos, la Tabla 4 muestra un ejemplo de los prototipos léxicos de otro sustantivo (*texto*) para su actante tema (A3) *El texto de la reforma*.

COLOCACIÓN	Género	1 NIVEL	2 NIVEL	3 NIVEL	FRE	LL SCORE
tema	m.	material	inanimado	objeto	4.397	8.36
reforma	m.	material	inanimado	objeto	842	8.62
ley	f.	material	inanimado	objeto	746	8.35
proyecto	m.	material	inanimado	objeto	573	7.55
problema	f.	material	inanimado	intelectual	486	7.40
asunto	m.	inmaterial	inanimado	intelectual	470	7.30

TABLA 4: Categorías semánticas de prototipos del A3 de texto

El siguiente paso, tras destacar los prototipos léxicos, fue agrupar las palabras que compartían los mismos rasgos en ficheros, denominados aquí *paquetes semánticos paradigmáticos*, para almacenarlos en la base de datos. Esta

modelización de los datos en sinónimos y cohipónimos se relaciona con el concepto de *synsets* de WordNet (vid. 3). Estos paquetes contenían sustantivos o adjetivos con características semánticas afines e identificados con dos niveles categoriales.

En definitiva, el enfoque es ascendente, porque las clases semánticas de la ontología se proponen a partir de las características semánticas de los prototipos en las tres lenguas, alemán, español y francés. Esta ontología 0.1 se aplicó en *MultiGenera* en la aplicación *Xera* versión 0.1 del generador monoargumental de la FN (vid. 4.1.1).

4.1.1. Proceso de desarrollo de la ontología 0.1

Los datos léxicos etiquetados y almacenados con la ontología 0.1 resultaban escasos. Entonces se enlazó con otras ontologías de código abierto (vid. 3) para incrementar el vocabulario de los paquetes semánticos de cada uno de los veinte sustantivos, a través de diferentes APIs (Application Programming Interfaces) (Domínguez Vázquez, Solla Portela & Valcárcel Riveiro, 2019), API1 (Lematiza)¹² y API2 (Combina)¹³ creadas para estos proyectos y disponibles en la web (Gómez Guinovart & Solla Portela, 2018).

En el proceso, primero, se subieron los paquetes semánticos de los sustantivos en las tres lenguas a la API 1, para aumentar su número de palabras. Esta API conectaba con otras ontologías: Word Domains, SUMO, TOP, Basic Level Concept, conceptos frecuentes de WordNet y Epinonyms, a través de la WordNet del gallego: Galnet (vid. 3). La Figura 2 muestra la interfaz desde donde se asociaban los paquetes semánticos con estas ontologías para la expansión léxica.


¹² (API1) Conecta las palabras de un fichero a través de Galnet con otras ontologías relacionadas con WordNet. Disponible en <http://portlex.usc.gal/develop/lematiza/>

¹³ (AP2) Disponible en <http://portlex.usc.gal/develop/combina.php>

MultiTools

Lematizador de actantes do corpus

Suba un ficheiro de concordancias ou frecuencias de Sketch Engine en formato xml ou csv ([Help](#))

 [View / edit de Sketch Engine...](#)

Seleccione a lingua de traballo

Deutsch ☐

español ☒

français ☐

[Vai](#)

1 textos de juristas

- Actante: juristas

- Lema actancial: **jurista**

Offsets:

10227985-n a legal scholar versed in civil law or the law of nations

--- WordNet Domains: law subcategories

--- SUMO: hasSkill subcategories

--- Top: 1stOrderEntry subcategories | Function subcategories | Human subcategories | Living subcategories | Object subcategories

--- Epinonyms: 6-30-08441203-n#jurisprudence subcategories | 6-30-09617867-n#expert subcategories

--- Hiperónimo(s): 09617867-n#expert hyponyms subcategories

--- Nivel de hiponimia (substantivos e verbos): 5

--- Ficheiro lexicográfico (substantivos): person

10249950-n a professional person authorized to practice law; conducts lawsuits or gives legal advice

--- WordNet Domains: law subcategories

--- SUMO: OccupationalRole subcategories

--- Top: 1stOrderEntry subcategories | Function subcategories | Human subcategories | Living subcategories | Object subcategories | Occupation

--- subcategories

--- Epinonyms: 6-30-10249950-n#attorney subcategories

--- Hiperónimo(s): 10480253-n#professional | professional_person hyponyms subcategories

--- Nivel de hiponimia (substantivos e verbos): 6

--- Ficheiro lexicográfico (substantivos): person

FIGURA 2: API 1 para expandir el paquete de texto

Los actantes lematizados podían enlazar con los índices u *offsets* de los distintos *synsets* o nodos léxicos de un dominio o área conceptual con los que se correspondía el lema y el nivel categorial al que se asocia en las diferentes ontologías relacionadas con WordNet. Por ejemplo, la Figura 2 muestra un ejemplo del fichero del actante 1 de *texto*, el primer lema obtenido (*jurista*) se puede asociar al *offset*=10227985-n y al *offset*=10249950-n de los *synsets* asociados a distintas áreas de conocimiento y las categorías de diferentes ontologías.

Una vez seleccionados los *offsets*, en las ontologías, se eligieron las categorías más cercanas a los prototipos, se anotaron sus direcciones en la AP2 (Figura 3) para mapear los datos entre las distintas ontologías y se

incrementó el número de palabras con las mismas características semánticas en los paquetes semánticos, correspondientes a los actantes de un sustantivo, por ejemplo [humano-cargo] del sustantivo *texto*.

FIGURA 3: Pantalla de API2 con mapeo a las ontologías

Entonces, los documentos .txt obtenidos en (API2) se subieron a la API3 (Flexiona)¹⁴, un flexionador morfológico, y los resultados se depuraron de agramaticalidades e incoherencias originadas por la regularización morfológica. Todos los paquetes semánticos expandidos, es decir, incrementados con más datos léxicos y etiquetados con la ontología 0.1 se emplearon para programar en la *Xera* 0.1 (Figura 4).

4.1.2. Aplicación de la ontología 0.1 a *Xera* (0.1)

La ontología 0.1 se concibió como una estructura con varios niveles de clases semánticas que podían ser argumentos de un sustantivo concreto. Se muestra la interfaz de la primera versión del generador monoargumental de

¹⁴ (API3) <http://portlex.usc.gal/develop/flexiona.php>

Selección de palabras

Selección a lengua de trabajo	Substantivo nuclear
Deutsch	olor
español	sabor
français	dolor
	presencia
	discusión
	pregunta
	texto
	muerte
	aumento
	huida

Estructura argumental

- [Determinante] + texto + de + [Determinante] + [Actante-N1]
- [Determinante] + texto + de + [Actante-N1]
- [Determinante] + texto + [Actante-A1]
- [Determinante] + texto + de + [Determinante] + [Actante-N2]
- [Determinante] + texto + de + [Actante-N2]
- [Determinante] + texto + [Actante-A2]
- [Determinante] + texto + sobre + [Determinante] + [Actante-N3]
- [Determinante] + texto + sobre + [Actante-N3]
- [Determinante] + texto + de + [Determinante] + [Actante-N3]
- [Determinante] + texto + de + [Actante-N3]

Selección de fuente

Empaquetado léxico

Consulta de WordNet

asociación_tiempoLibre ☐ creencia_religiosa ☐ humano_cargo ☒ humano_educativa ☐ humano_familia ☐ humano_ideologia ☐ humano_origen ☐
humano_profesion ☐ organizaciónGubernamental_autoridad ☐ organización_educativa ☒ organización_militar ☐

Vai →

FIGURA 4: O: *Sustantivo* texto en Xera 0.1

la FN en *Xera1.0*. En la Figura 4, vemos su funcionamiento, al seleccionar *texto* se despliega un menú con su estructura actancial. Aquí aparece seleccionado [Determinante] + texto + [Actante N1] y debajo las clases semánticas que pueden cubrirlo, la elección [humano-cargo] genera las frases nominales mostradas.

4.2. ENFOQUE DESCENDENTE

3) se planteó la necesidad de ampliar y reajustar los niveles categoriales para

establecer equivalencias y favorecer la conexión de datos con otras ontologías. Se vio la necesidad de establecer categorías equiparables para facilitar la conexión de datos léxicos y la interoperabilidad de recursos. Al reelaborar la ontología original se tuvo como referencia WordNet y las ontologías relacionadas, lo que supuso un enriquecimiento de la estructura de esta ontología léxica, se ampliaron las categorías superordinadas y la granularidad en los niveles más específicos, para favorecer la programación en *Xera* y *Combinatoria*. Los niveles superiores aumentaron en tres, desde la categoría más abstracta: *material/inmaterial* de la ontología 0.1 (Figura 1, vid. 4.1.1). El resultado fue la ontología 0.2 con ocho niveles de inclusión, y dividida en ocho áreas de conocimiento o dominios: *Mundo animado*; *Mundo material*; *Mundo intelectual*; *Comunicación, pensamiento y cognición*; *Eventos dinámicos: procesos y actividades*; *Situaciones estáticas y condiciones/estados*; *Locación*; *Tiempo y Cantidad* (Figura 5).

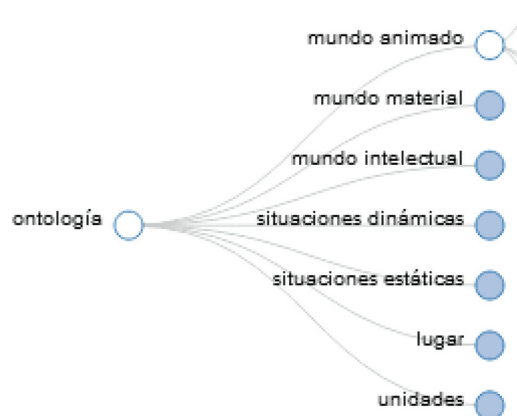


FIGURA 5: Dominios de la ontología de MultiGenera y MultiComb

Los dominios o áreas conceptuales son los hiperónimos que representan los niveles más abstractos e incluyen dos niveles intermedios hasta llegar al marcado como nivel 1, que coincide con el nivel superior de la ontología 0.1, desde el cual aumenta la granularidad hasta en cuatro niveles de subordinación (Figura 6).

Mundo animado	Ser humano		Nivel 1	Nivel2	Nivel 3	Nivel 4	Nivel 5
		Ser humano y condición/relación	animado	humano	general		
			animado	humano	origen		
			animado	humano	familia		
			animado	humano	personaje histórico		
			animado	humano	condición		
			animado	humano	humana	condición	negativa
			animado	humano	humana	condición	positiva
			animado	humano	actor	acto	negativo

FIGURA 6: Detalle de la ontología¹⁵

En definitiva, la ontología 0.2 es el resultado de la evolución de la ontología 0.1 con más niveles categoriales, que facilitan la conexión con otras aplicaciones y restringen la selección de los argumentos semánticos en el procesamiento computacional. De esta manera, se optimizó la modelización y formalización del significado del léxico en la versión *Xera* versión 0.2 y *Combinatoria*.

4.2.1. La ontología 0.2 en *Xera* Versión 2 y *Combinatoria* 2

La ontología 0.2 se emplea en la base de datos de los proyectos, en la programación de las aplicaciones, y sus categorías semánticas se utilizan en la comunicación con el usuario de las aplicaciones *Xera* 0.2 y *Combinatoria* para seleccionar los argumentos que generan automáticamente la FN.

La aplicación *Xera* 0.2 de libre acceso presenta una interfaz más amigable para el usuario que la versión 0.1 y genera FN de un solo argumento. El funcionamiento es similar a la anterior, se elige el idioma, uno de los 20 sustantivos y para la consulta, primero se selecciona la estructura sintáctica formal y después las clases semánticas. Esto generará los ejemplos siguiendo

¹⁵ Disponible en <http://portlex.usc.gal/ontologia/>

el filtrado realizado por los prototipos léxicos, representado por categorías semánticas, adjudicables a cada argumento. La Figura 7 muestra el ejemplo de *texto* con el actante A1 y las posibles combinaciones semánticas con las propiedades [animado humano familiar].

The screenshot shows the Xera web application interface. At the top, there is a blue button labeled 'INFORMACIÓN'. Below it, the 'Idioma:' field is set to 'ES'. The 'Núcleo:' field is set to 'texto'. The 'Estructura:' field is set to 'determinante+adjetivo o+texto+adjetivo o+de+determinante+actante N1'. A red banner indicates '1 paquete seleccionado'. Below this, a list of semantic combinations is shown, each with a checkbox. The combination 'animado humano cargo el (interesante) texto (interesante) de la ministra de agricultura' is selected with a red checkmark. Below the list, there is a slider for 'límite de frases' set to 200. At the bottom, there are three buttons: 'GENERAR', 'EXPORTAR FRASES EN JSON', and 'EXPORTAR FRASES EN CSV'. Below these buttons, a table shows the generated phrases:

frases generadas	
el	texto de la ministra de agricultura
el	texto del sustituto
el	texto del jefe

FIGURA 7: Xera: ejemplo de texto

Fuente: <http://portlex.usc.gal/combinatoria/usuario>.

4.2.2. Combinatoria

La aplicación *Combinatoria* es un generador automático biargumental de la FN en los tres idiomas, disponible online. En la interfaz con el usuario se pueden hacer consultas sobre los sustantivos y sus combinaciones en tres idiomas. Inicialmente, se elige la categoría semántica que completará el argumento 1.

Existen tres niveles de selección para cada uno. En el primero, se selecciona una las ocho categorías de esta ontología: *animado*, *material*, *intelectual*, *lugar*, *estado*, *situación*, *unidad* y *proceso*. Esta elección implica una herencia de significado limitando el número de categorías semánticas del nivel 2. A su vez, la elección de una de ellas restringe más las posibilidades semánticas del nivel 3. En consecuencia, las características sintáctico-semánticas del complemento1 limitan las posibles combinaciones semánticas del argumento, complemento 2. En el nivel 1, el complemento 2 solo muestra las clases semánticas que pueden combinarse con el complemento 1, siguiendo así el mismo proceso restrictivo en los dos niveles siguientes para la selección de la categoría semántica más específica del complemento 2. La Figura 8 muestra un ejemplo de la combinatoria de texto y las frases generadas.

The interface is divided into two main sections for selecting complements. The left section is for 'Primer complemento' (First complement) and the right for 'Segundo complemento' (Second complement). Both sections follow a three-step process: 1. Selection of a primary category (e.g., 'animado' or 'intelectual'), 2. Selection of a secondary category (e.g., 'humano' or 'ideología'), and 3. Selection of a final category (e.g., 'creencia religiosa' or 'política'). Below these steps, a grid of specific categories is displayed. A search bar is located at the top right of the grid. At the bottom, a table lists generated text examples based on the selected categories. The second row of the table is highlighted with a yellow oval.

ejemplo	complemento1	complemento2
el texto de los agnósticos sobre el agotamiento	animado humano creencia religiosa	situación estado físico
el texto de los agnósticos sobre el absolutismo	animado humano creencia religiosa	intelectual ideología política
el texto de los agnósticos sobre el abatimiento	animado humano creencia religiosa	situación estado emocional
el texto de los agnósticos sobre las informaciones confidenciales	animado humano creencia religiosa	intelectual contenido general
el texto de los agnósticos sobre el abominable hombre de las nieves	animado humano creencia religiosa	animado criatura de ficción general
el texto de los agnósticos sobre la actualidad	animado humano creencia religiosa	unidad tiempo período

FIGURA 8: Ejemplo de combinatoria biargumental para texto en Combinatoria
Fuente: <http://portlex.usc.gal/combinatoria/>.

La mayor granularidad de la ontología permite programar con mayor nivel de restricción y generar frases nominales más adecuadas para su aplicación en la enseñanza de lenguas y la traducción. En la actualidad se está trabajando para generar el contexto oracional en un futuro.

5. CONCLUSIONES

En este trabajo, hemos demostrado como la ontología es un recurso léxico-conceptual que facilita la representación del mundo mediante categorías conceptuales relacionadas entre sí. Se ha concebido como un esquema conceptual, una estructura jerárquica cercana a una taxonomía, que organiza las propiedades semánticas del léxico bien a través de relaciones verticales de inclusión y horizontales de identidad. Las relaciones jerárquicas, en general son la base de la ontología, la herencia es una propiedad destacada en la configuración de las categorías semánticas que la conforman. En los proyectos *MultiGenera* y *MultiComb* su utilidad es indiscutible para el almacenamiento y recuperación del léxico de la base de datos y para el desarrollo del software, los niveles categoriales y las clases propuestas en la ontología funcionan como un esquema conceptual básico para la formalización de las restricciones semánticas en la programación de la generación automática de frases nominales con un argumento en *Xera* y con dos en *Combinatoria*.

La motivación para diseñar una ontología propia para los proyectos era dar cuenta de las restricciones semánticas en la combinatoria de los sustantivos en la FN tal como se planteaba en la gramática y lexicografía valencial (Domínguez Vázquez, 2011; Domínguez Vázquez, Valcárcel Riveiro & Lindemann, 2018). Cuando se comenzó la modelización del *MultiGenera*, se pensó que ninguna ontología de libre acceso se adecuaba a la finalidad del proyecto, por lo que se decidió profundizar en las categorías semánticas empleadas en el *Portlex* constituyendo la ontología 0.1. No obstante, al avanzar la investigación, esta ontología se enlazó automáticamente con WordNet y otras

ontologías, para aumentar los datos léxicos. Entonces se vio la necesidad de ampliar las categorías semánticas, sus relaciones y reajustar el etiquetado de las mismas para facilitar la conexión, por esta razón, la ontología evoluciona a la versión 0.2, tomando como referencia WordNet. La ontología 0.2 sigue en periodo de prueba, se espera que pueda ser portable, que permita la reutilización de datos y en un futuro la interoperatividad entre las distintas lenguas de los proyectos¹⁶.

En definitiva, la ontología es una herramienta lexicográfica para la enseñanza de lenguas, con la didactización de las aplicaciones *Xera* y *Combinatoria* que, desde un enfoque onomasiológico, cognitivo organiza el léxico disponible para el aprendiz de lenguas, las categorías semánticas de la ontología que se muestran al usuario funcionan como principio selectivo en las relaciones sintagmáticas de la FN y permiten explorar el sentido dinámico del significado (Croft & Cruse, 2004).

Para concluir, la ontología tiene un valor en sí misma y como herramienta lexicográfica tanto para la lectura humana y como para la lectura computacional. Hay que destacar que de momento sólo se ha probado en estos proyectos. En un futuro, la optimización de esta ontología nos permitirá la interoperabilidad semántica con diferentes aplicaciones y funcionará como índice interlingüístico entre las tres lenguas para su traducción y el análisis contrastivo. La ontología y las demás aplicaciones mencionadas en este trabajo están disponible online en la web de los proyectos¹⁷.

REFERENCIAS BIBLIOGRÁFICAS

- Bentivogli, L., Forner, P., Magnini, B. & Pianta, E. (2004). Revising the Wordnet Domains Hierarchy: Semantics, Coverage and Balancing. En G. Sérasset, S. Armstrong, C. Boitet, A. Popescu-Belis & D. Tufis (eds.), *Proceedings of the Workshop on Multilingual Linguistic Resources MLR2004*. Universidad de Ginebra. <https://doi.org/10.3115/1706238.1706254>

¹⁶ Esta ontología está disponible en la web <http://portlex.usc.gal/ontologia/>

¹⁷ La ontología y todas las herramientas están disponibles en <http://portlex.usc.gal/>

- Borst, W. N. (1997). *Construction of Engineering Ontologies for Knowledge Sharing and Reuse* [Tesis doctoral. Universidad de Twente]. Centre for Telematics and Information Technology (CTIT). <https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and->
- Bosque-Gil, J. (2019). *Linguistic Linked Data for Lexicography* [Tesis doctoral. Universidad Politécnica de Madrid]. Archivo Digital UPM. <https://doi.org/10.20868/UPM.thesis.57887>.
- Bosque-Gil, J. & García, J. (eds.) (2019). *The Ontolex Lemon Lexicography Module Specification*. Ontology Lexica under the W3C Community Final Specification Agreement (FSA). <https://www.w3.org/2019/09/lexicog/>
- Brown, C. (2002). Paradigmatic relations of inclusion and identity I: Meronymy. En A. Cruse, F. Hundsnurscher, J. Michael & P. R. Lutzner (eds.), *Lexikologie: Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen 1. Halbband. Lexicology: An International Handbook on the Nature and Structure of Words and Vocabularies*, vol. I. (pp. 480-485). de Gruyter.
- Codina, L. & Pedraza-Jiménez, R. (2011). Tesoros y ontologías en sistemas de información documental. *El profesional de la información*, 20(5), 555-563. <https://doi.org/10.3145/epi.2011.sep.10>
- Combinatoria = Domínguez Vázquez, M.^a J. (dir.), Valcárcel Riveiro, C., Bardanca Outeiriño, D., Calañas Continente, J. A., Catalá Torres, N., López Iglesias, N., Martín Gascuña, R., Mirazo Balsa, M., Sanmarco Bande, M. T. & Pino Serrano, L. (2020). *Combinatoria. Prototipo online para la generación biargumental de la frase nominal en alemán, español y francés*. Universidade de Santiago de Compostela. Consultado el 30 de mayo de 2023. <http://portlex.usc.gal/combinatoria/usuario>
- Croft, W. & Cruse, D. A. (2004). *Cognitive Linguistic*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803864>
- Cruse, A. (1986). *Lexical Semantics*. Cambridge University Press.
- Cruse, A. (2002). Descriptive models for sense relations II: Cognitive semantics. En A. Cruse, F. Hundsnurscher, J. Michael & P. R. Lutzner (eds.), *Lexikologie: Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen 1. Halbband. Lexicology: An International Handbook on the Nature and Structure of Words and Vocabularies*, vol. I. (pp. 542-549). de Gruyter. <https://doi.org/10.1515/9783110113082.1.15.542>
- Cruse, D. A. (2004). *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press.
- Domínguez Vázquez, M.^a J. (2011). *Kontrastive Grammatik und Lexikographie: spanisch-deutsches Wörterbuch zur Valenz des Nomens*. IUDICIUM Verlag GmbH München.
- Domínguez Vázquez, M.^a J., Solla Portela, M. A. & Valcárcel Riveiro, C. (2019). Resource Interoperability: Exploiting Lexicographic Data to Automatically Generate Dictionary Examples. En I. Kosem, T. Z. Kuhn, M. Correia, J. P. Ferreira, M. Jansen, I. Pereira, J. Kallas, M. Jakubiček, S. Krek & C. Tiberius (eds.), *Electronic Lexicography in the 21st Century: Smart Lexicography. Proceedings of the eLex 2019* (pp. 51-71). Lexical Computing.

- Domínguez Vázquez, M.^a J., Valcárcel Riveiro, C. & Lindemann, D. (2018). Multilingual Generation of Noun Valency Patterns for Extracting Syntactic-Semantical Knowledge from Corpora (*MultiGenera*). En J. Čibej, V. Gorjanc, I. Kosem & S. Krek (eds.), *XVIII EURALEX International Congress. Lexicography in Global Contexts* (pp. 847-854). Ljubljana University Press.
- Gangemi, A., Navigli, R. & Velardi, P. (2003). The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. En R. Meersman, Z. Tari & C. Schmidt (eds.), *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE* (pp. 820-838). Springer. https://doi.org/10.1007/978-3-540-39964-3_52
- Gómez Guinovart, X. & Solla Portela, M. A. (2018). Building the Galician wordnet: methods and applications. *Language Resources and Evaluation*, 52(1), 317-339. <https://doi.org/10.1007/s10579-017-9408-5>
- Gruber, T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2), 199-220. <https://doi.org/10.1006/knac.1993.1008>
- Guarino, N., Oberle, D. & Staab S. (2009). What Is an Ontology? En S. Steffen & R. Studer (eds.), *Handbook on Ontologies. International Handbooks on Information Systems* (pp. 1-17). Springer. https://doi.org/10.1007/978-3-540-92673-3_0
- Lakoff, G. (1987). *Women, Fire and Dangerous Things: what Categories Reveal about the Mind*. The University of Chicago Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Lenat, D. & Guha, R. V. (1990). CYC: A Mid-Term Report. *AI Magazine*, 11(3), 32-59.
- Lyons, J. (1977). *Semántica*. Editorial Teide. <https://doi.org/10.1017/CBO9781139165693>
- Martin-Gascueña, R. (2013). La hiponimia en un área conceptual. *Revista Pragmalingüística*, 21, 86-106 <http://revistas.uca.es/index.php/pragma/issue/view/139/showToc>
- Martin, P. (2003). Correction and Extension of WordNet 1.7 for Knowledge-based Applications. En B. Ganter, A. Moor & W. Lex (eds.), *Conceptual Structures for Knowledge Creation and Communication* (pp. 160-173). Springer. https://doi.org/10.1007/978-3-540-45091-7_11
- Masolo, C., Borgo, S., Gangemi, A., Guarino, N. & Oltramari, A. (2003). *DOLCE: Descriptive Ontology for Linguistic and Cognitive Engineering* <https://www.istc.cnr.it/it/content/dolce-descriptive-ontology-linguistic-and-cognitive-engineering>
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. & Miller, J. (1990). Introduction to WordNet: An on-line Lexical Database. *International Journal of Lexicography*, 3(4), 235-244. <https://doi.org/10.1093/ijl/3.4.235>
- Ontología léxica = Domínguez Vázquez, M.^a J., Valcárcel Riveiro, C. & Bardanca Outeiriño, D. (2021). *Ontología léxica*. Universidade de Santiago de Compostela. Consultado el 30 de mayo de 2023 <http://portlex.usc.gal/ontologia/>
- Pease, A., Niles, I. & Li, J. (2002). The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Application. *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web. July-August*. Edmonton.

- Pianta, E., Bentivogli, L. & Girardi, C. (2002). MultiWordNet Developing an aligned multilingual database. *Proceedings of the 1st International WordNet Conference* (pp. 293-302).
- Portlex = Domínguez Vázquez, M.^a J. (dir.), Valcárcel Riveiro, C., Mirazo Balsa, M., Sanmarco Bande, M.^a T., Simões, A. & Vale, M. J. (2018). *Portlex. Dicionario multilingüe de la valencia del nombre*. Universidade de Santiago de Compostela. Consultado el 30 de mayo de 2023. <http://portlex.usc.gal/portlex/>
- Pustejovsky, J. (1995) *The Generative Lexicon*. The MIT Press Cambridge.
- Rodríguez, H., Climent, S., Vossen, P., Bloksma, L., Peters, W., Alonge, A., Bertagna, F. & Roventini, A. (1998). The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. P. Vossen (ed.) *EuroWordNet: A multilingual database with lexical semantic networks* (pp. 45-80). Springer. https://doi.org/10.1007/978-94-017-1491-4_3
- Rosch, E. (1978). Principles of Categorization. En B. Lloyd & E. Rosch (eds.), *Cognition and categorization* (pp. 27-48). Lawrence Erlbaum.
- Solla Portela, M. A. & Gómez Guinovart, X. (2015). Galnet o WordNet do galego. Aplicacións lexicolóxicas e terminolóxicas. *Revista galega de filoloxía*, 16, 169-201. <https://doi.org/10.17979/rgf.2015.16.0.1383>
- Studer, R., Benhamins, V. R. & Fensel, D. (1998). Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1-2), 161-197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Vossen, P. J. T. M. (1998). EuroWordnet: Building a Multilingual Database with Word Nets for European Languages. *ELRA Newsletter*, 3(1), s.p. https://doi.org/10.1007/978-94-017-1491-4_1
- WordNet = The Trustees of Princeton University (2023). *WordNet*. Princeton University. Consultado el 30 de mayo de 2023. <https://wordnet.princeton.edu/>
- Xera = Domínguez Vázquez, M.^a J. (dir.), Valcárcel Riveiro, C., Bardanca Outeiriño, D., Calañas Continente, J. A., Catalá Torres, N., López Iglesias, N., Martín Gascuña, R., Mirazo Balsa, M., Sanmarco Bande, M.T. & Pino Serrano, L. (2020). *Xera. Prototipo online para la generación automática monoargumental de la frase nominal en alemán, español y francés*. Universidade de Santiago de Compostela. Consultado el 30 de mayo de 2023. <http://portlex.usc.gal/combinatoria/usuario>