

CONCEPTUALIZACIÓN DE ENTIDADES TERMINOLÓGICAS EN UNA SUBONTOLOGÍA DE DERECHO PENAL: ANÁLISIS DEL CONCEPTO SUPERORDINADO +DRUG_00 EN FUNGRAMKB

Ángel M. Felices Lago

Pedro Ureña Gómez-Moreno

Universidad de Granada, España

Resumen: Este artículo describe las fases de construcción de una ontología terminológica dentro de la arquitectura de FunGramKB, una base de conocimiento léxico-conceptual para el procesamiento computacional del lenguaje natural. La descomposición semántica de la terminología compleja se realiza siguiendo el proceso COHERENT, esto es, una metodología gradual para la formalización de conceptos especializados. Para tratar de ilustrar este proceso hemos seleccionado el concepto +DRUG_00, así como otros conceptos subordinados, tales como \$METHAMPHETAMINE_00, \$CANNABIS_00 y \$COCAINE_00 pertenecientes al dominio de las sustancias estupefacientes (drogas). Las definiciones de los conceptos seleccionados se basan en COREL, un metalenguaje de interfaz inspirado en algunos principios generales de la Gramática del Papel y la Referencia (GPR). Como resultado del proceso de modelado, subsunción y jerarquización, la ruta conceptual superior de la ontología se representa en la Ontología Satélite dentro de FunGramKB según el siguiente esquema: #ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +SOLID_00 > +DRUG_00.

Palabras clave: FunGramKB, ontología nuclear, ontología satélite, terminología, derecho penal, modelado conceptual.

CONCEPTUALIZATION OF ENTITIES IN A CRIMINAL LAW SUBONTOLOGY: ANALYSING THE SUPERORDINATE CONCEPT +DRUG_00 IN FUNGRAMKB

Abstract: This article describes some phases in the process of constructing a term-based Satellite Ontology within the architecture of the Core Ontology integrated in FunGramKB (a lexico-conceptual knowledge base for the computational processing of natural language). The semantic decomposition of complex terminology is implemented following the COHERENT methodology (a stepwise method for formalizing specialized concepts). For that purpose, we have selected the superordinate concept +DRUG_00 as well as other subordinate concepts in the domain of drugs such as \$METHAMPHETAMINE_00, \$CANNABIS_00, and \$COCAINE_00. The definitions of the concepts selected for the study are based on COREL, an interface metalanguage inspired on some general principles of Role and Reference Grammar (RRG). As a result of the modeling, subsumption and hierarchization process the top conceptual path is represented in the Satellite Ontology as follows: #ENTITY > #PHYSICAL > #OBJECT > #SELF_CONNECTED_OBJECT > +ARTIFICIAL_OBJECT_00 > +SUBSTANCE_00 > +SOLID_00 > +DRUG_00.

Keywords: FunGramKB, core ontology, satellite ontology, terminology, criminal law, conceptual modeling.

1. INTRODUCCIÓN

El diseño de una ontología que represente el conocimiento jurídico o que recoja de forma organizada los conceptos fundamentales de una rama concreta del derecho, como por ejemplo el derecho penal (sustantivo y procesal), tiene múltiples aplicaciones en el campo del Procesamiento del Lenguaje Natural (PLN) y, más concretamente, en la resolución de problemas en contextos jurídicos reales. Así pues, tanto el desarrollo conceptual de una ontología jurídica de estas características, así como la población de sus correspondientes lexicones en un número determinado de lenguas distintas, constituyen la antesala de la aplicación de las ontologías, y de las denominadas “bases de conocimiento” en las cuales se insertan, en tareas de razonamiento artificial. Cabe plantear en este contexto de “ingeniería del conocimiento” una reflexión sobre qué tipo de unidades deben conformar una

To cite this article: Felices, A., Ureña, P. (2020). "Conceptualización de entidades terminológicas en una subontología de derecho penal: análisis del concepto superordinado +DRUG_00 en FunGramKB". *Revista de Lingüística y Lenguas Aplicadas*, 15, 15-25. <https://doi.org/10.4995/rlyla.2020.12772>

Correspondence author: afelices@ugr.es



ontología especializada como la referida al derecho penal, y cómo deben aparecer definidas dentro de ella. En este punto concreto cobra especial relevancia el papel de la terminología como rama lingüística encargada del análisis de unidades léxicas especializadas (esto es, de “términos”). De hecho, los terminólogos constituyen una pieza fundamental en el proceso de creación ontológica, ya que son los encargados de establecer una conexión entre los conceptos de una rama de conocimiento especializada y las unidades léxicas especializadas que aportan sustancia lingüística a los mismos. El área del derecho, no obstante, plantea no pocas dificultades en este objetivo de enlace terminológico-conceptual, pues en no pocas ocasiones las unidades léxicas del ámbito jurídico, sobre todo las más popularizadas, podrían tener cabida tanto en una ontología especializada como en una ontología de propósito general o de sentido común.

En este artículo presentamos la base de conocimiento FunGramKB (Periñán y Arcas, 2004, 2010a, 2010b; Jiménez y Luzondo, 2011; Periñán, 2013) y una implementación de la metodología de modelado conceptual conocida como COHERENT (Periñán y Mairal, 2011), que ofrecen ambas un marco adecuado para la construcción de ontologías jurídicas basadas en semántica profunda (Velardi et al., 1991). Para ello, ilustraremos la representación conceptual de conceptos a partir de unidades pertenecientes a una subontología jurídica actualmente en desarrollo. En concreto, definiremos el concepto DRUG, así como los conceptos subordinados METHAMPHETAMINE, CANNABIS y COCAINE. El proceso de creación de ontologías jurídicas que proponemos en este artículo se sustenta en varios pilares consolidados: a) la experiencia previa en la conceptualización de términos jurídicos (Felices, 2015, 2016); b) los progresos en la elaboración de varias herramientas que potencian el motor de razonamiento de FunGramKB (Periñán y Arcas, 2014, 2019; Cortés y Mairal, 2016; Periñán, 2017, 2018); c) el desarrollo completo de la herramienta DEXTER para la extracción terminológica (Periñán, 2015, 2018); d) una interfaz de conexión entre la ontología de alto nivel de FunGramKB (u Ontología Nuclear) y las ontologías terminológicas (u Ontologías Satélite) (Felices y Ureña, 2012, 2014; San Martín y Faber, 2012); y e) el lenguaje de representación COREL (Periñán y Mairal, 2010), cuyas ventajas sobre otros formalismos existentes ha quedado demostrada en aplicaciones concretas (Periñán y Mairal, 2010; Periñán, 2012; Periñán y Arcas, 2014)¹.

La estructura del artículo es la siguiente. En primer lugar, haremos una revisión de las características fundamentales de la base de conocimiento FunGramKB y analizaremos sucintamente el estado de la cuestión sobre ontologías jurídicas (Apartado 2). En segundo lugar, describiremos las fases para el diseño de una ontología enfocada al derecho penal y, fundamentalmente, al crimen organizado (Apartado 3). En este punto mostraremos la herramienta DEXTER, cuya función consiste en extraer a partir de un corpus de entrada las unidades terminológicas que posteriormente servirán de base léxica para la construcción ontológica. Finalmente, a modo de ilustración, expondremos los componentes más importantes de la definición semántica de los conceptos seleccionados (DRUG, METHAMPHETAMINE, CANNABIS y COCAINE), y mostraremos las fases de conceptualización y jerarquización (Apartado 4).

2. FUNDAMENTOS TEÓRICOS

2.1. FunGramKB y el conocimiento cognitivo

FunGramKB es una base de conocimiento que engloba tres niveles de información (el nivel léxico, el nivel gramatical y el nivel conceptual), todos ellos a su vez constituidos por varios módulos independientes pero interrelacionados (Figura 1). La finalidad de esta herramienta de procesamiento es la de contribuir a que los sistemas informáticos puedan producir, así como entender, lenguaje humano y aplicar este potencial en distintas tareas computacionales como el tratamiento de información, el razonamiento artificial o el descubrimiento de patrones significativos:

1. Nivel léxico:
 - El *lexicón* almacena información morfosintáctica, pragmática y colocacional de unidades léxicas.
 - El *morfocón* ayuda al sistema a manejar casos de morfología flexiva.
2. Nivel gramatical:
 - El *gramaticón* se estructura siguiendo las directrices del Modelo Léxico Construccional (MLC) (Ruiz de Mendoza y Mairal, 2008). Su función consiste en almacenar esquemas constructivos.
3. Nivel conceptual:
 - La *ontología* se presenta como una estructura jerárquica de los conceptos fundamentales que toda persona posee y que utiliza para referirse a situaciones cotidianas. Este módulo permite la herencia múltiple no monotónica. La ontología consiste en un módulo de propósito general (Ontología Nuclear)

¹ Téngase en cuenta que el sistema notacional de COREL está influido de manera muy apreciable por el potencial computacional que poseen tanto el modelo gramatical del Papel y la Referencia (GPR) (Van Valin y Lapolla, 1997; Van Valin, 2005) como el modelo Léxico-Construccional (Ruiz de Mendoza y Mairal, 2008), los cuales facilitan el desarrollo de PLN desde la perspectiva de la semántica profunda.

y varios módulos terminológicos específicos de dominios especializados (Ontologías Satélite).² En nuestro caso, la conceptualización de una subontología basada en el derecho penal constituiría una Ontología Satélite dentro de FunGramKB.

- El *cognición* almacena conocimiento procedimental, es decir, guiones o secuencias sobre acciones estereotipadas, como por ejemplo “comer en un restaurante”, “ir de compras” o “pagar con tarjeta de crédito”.
- El *onomasticón* almacena información enciclopédica sobre instancias de entidades, como por ejemplo personas (Albert Einstein), ciudades (Nueva York), o eventos (Batalla de Stalingrado), etc.

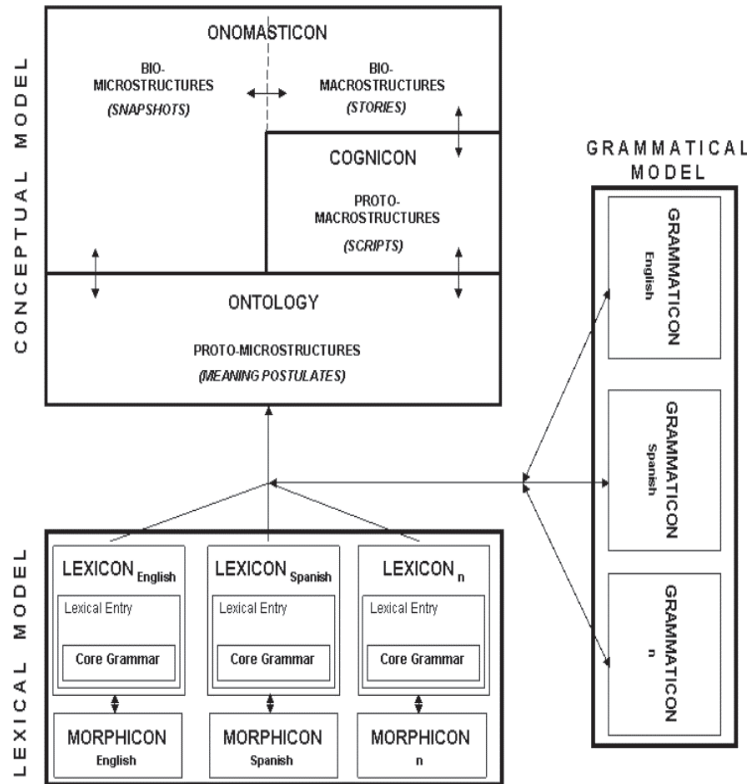


Figura 1: Estructura de FunGramKB.³

La Ontología Nuclear es una taxonomía conceptual de uso general que permite incluir conocimiento especializado en la base de conocimiento mediante un proceso de vinculación jerárquica entre ésta y otras potenciales Ontologías Satélite en campos como la medicina, el derecho, la química o la contabilidad. Respecto a la adecuación tipológica, mientras que los módulos léxico y gramatical son específicos para cada idioma, el nivel conceptual es compartido por todos los idiomas, de tal forma que FunGramKB requiere un único nivel conceptual para procesar una entrada léxica pero requiere un lexicón y un morficon específicos para el inglés, el español, el italiano, etc. Dicho de forma más concreta, la unidad léxica “perro”, por ejemplo, y sus equivalentes en otros idiomas (*dog*, *Hund*, *chien*, etc.) quedan recogidos en sus correspondientes lexicones dentro de FunGramKB, mientras que, por otro lado, todos ellos tienen un único reflejo conceptual dentro de la Ontología Nuclear bajo la unidad +DOG_00. Esta forma de organización en dos planos diferenciados pero conectados tiene claras ventajas computacionales, ya que hace que el procesamiento dentro de la base de conocimiento sea más eficiente y, por otro lado, que la información conceptual puede utilizarse para cualquier idioma independientemente de sus características léxicas o gramaticales. Los conceptos representados en FunGramKB son de tres tipos. En primer lugar, los metaconceptos (por ejemplo, #EMOTION o #COGNITION) constituyen el nivel superior de la taxonomía y son el resultado del conocimiento recogido en ontologías ya existentes, como DOLCE (Gangemi et al., 2005), SIMPLE (Lenci, 2008) o SUMO (Niles y Pease, 2001). A rasgos generales, estas ontologías contienen la mayor parte de los conceptos necesarios para abordar tareas generales de comprensión y razonamiento ontológico y para las que se requiere conocimiento del mundo real o de “sentido común”. La decisión de establecer estas ontologías

² En general, las ontologías se consideran independientes de su implementación y por tanto operan en un nivel superior de abstracción. En el caso de la ontología nuclear y las ontologías satélites de FunGramKB, el modelo ontológico se implementa computacionalmente a través de una base de datos relacional por dos razones. Por una parte, con respecto a la representación de los datos, la base de datos relacional permite realizar una clara distinción entre esquema y datos, algo que no ocurre con la verbosidad de lenguajes orientados a la web semántica como RDF y OWL. Por otra parte, con respecto a la eficiencia, las bases de datos relacionales gestionan mejor la información, especialmente cuando el número de instancias es muy alto.

³ Esta estructura modular de FunGramKB puede verse en www.fungramkb.com (Acceso 21/10/2019).

como punto de partida para la creación del nivel metaconceptual de FunGramKB no solo ha permitido centrar el proceso de ingeniería del conocimiento en la conceptualización de unidades más específicas de los niveles básico y terminal de la Ontología, sino que hace posible una potencial integración e intercambio de información entre FunGramKB y estas ontologías u otras similares (Perián y Arcas, 2010a:2669)

Los metaconceptos se agrupan a su vez en tres grandes grupos: #ENTITY, #EVENT, y #QUALITY. En segundo lugar, los conceptos básicos (por ejemplo, +LAW_00 o +HUMAN_00) dependen de los metaconceptos y tienen como finalidad su utilización en las definiciones de otros conceptos básicos de la ontología. En tercer lugar, los conceptos terminales (por ejemplo, \$SUBLIMINAL_00 o \$MOP_00) constituyen el nodo más básico y específico de la jerarquía conceptual. Éstos, a diferencia de los básicos, sirven para definir conceptos, mientras que los segundos no pueden utilizarse como *definiens* en la definición semántica de otros conceptos.

2.2. Ontologías legales y derecho penal: sus antecedentes

Los antecedentes sobre las ontologías son muy amplios (cf. Musen, 1992 o Gruber, 1993), si bien habría que retrotraerse en primera instancia a los antecedentes más remotos en lexicografía onomasiológica, recogidos por Martín Mingorance (1994), los cuales, aunque procedan de la filosofía y no tengan relación con el concepto de ontología en el PLN, sí dan cabida a la preocupación humana por aprehender la estructura del conocimiento y la realidad. Ahí habría que citar los trabajos de figuras universales como Aristóteles y su filosofía de las *esencias*, el *esquema* de Porfirio, la *Historia Naturalis* de Plinio el Viejo, las *Etimologas* de S. Isidoro de Sevilla o, en época moderna, el *Instauratio Magna* o el *Novum Organum* de Francis Bacon o el *Ianua Linguarum Reserata* de Comenius. En el periodo contemporáneo hay que destacar los *Thesauri* de Roget o la obra del escocés Wilkins. Estos son algunos de los antepasados de diccionarios ideológicos más actuales y útiles, como es el caso del *Longman Lexicon of Contemporary English* de McArthur y, para el español, el histórico diccionario ideológico de Julio Casares, publicado en 1942.

Con respecto al desarrollo contemporáneo de ontologías especializadas en conexión con su significado en Inteligencia Artificial (IA) o PLN merecen destacarse algunas contribuciones entre muchas otras: el proyecto COGNITERM, dirigido por el profesor Skuce de la Universidad de Ottawa y su sistema gestor de bases de datos de conocimientos denominado CODE (*Conceptually Oriented Design Environment*); o bien *MikroKosmos*, que fue diseñada en los noventa por Manesh y Nirenburg para traducir textos concretos sobre fusiones y adquisiciones empresariales.

La creación de ontologías dentro del dominio legal surge en 1994 (cf. Valente y Breuker, 1994) por la necesidad de formalizar información jurídica en un formato legible computacionalmente y que a su vez permitiera el intercambio de datos entre distintos sistemas informáticos para la toma de decisiones o de recuperación de información. El objetivo era entonces, y sigue siendo a día de hoy, proporcionar mecanismos informatizados para la gestión de un volumen siempre creciente de información legal en formato electrónico (Breuker et al., 2008). En 2005 André Valente catalogó veinticuatro ontologías en el ámbito legal (Valente, 2005) y seis años después Casanovas y otros autores incrementaron la lista a sesenta (Sartor et al., 2011). Entre las recogidas por Sartor et al., sólo dos se diseñaron para el derecho penal o estaban referidas a delitos recogidos en el sistema jurídico penal, a saber: la ontología de delitos italiana descrita en Asaro et al. (2003), y la ontología del derecho holandés CRIME.NL (Breuker y Winkels, 2003). La primera ofrece una representación muy esquemática de todos los agentes, procesos, factores, etc. que intervienen en la comisión del delito; la segunda, que se diseñó con el objetivo de reutilizarse para el derecho penal italiano y polaco en el marco del proyecto europeo denominado e-COURT, representa un ambicioso plan que trataba de englobar la estructura fundamental del derecho penal holandés como factor para la construcción de una ontología de derecho penal europea. Posteriormente han surgido otras propuestas igualmente interesantes como, por ejemplo, la de Bezzazi (2007), quien se refiere a la creación de una ontología para la identificación de artículos de derecho penal en materia de cibercriminología, o la propuesta de Bak, Jedrzejek y Falkowski (2010), que aplican un modelo de base ontológica para delitos de naturaleza económica. En esta línea, y de forma más reciente, Chiseung Soh et al. (2015) han presentado el modelado de una ontología que incluye características comunes al derecho penal, y han propuesto un método para diseñar normas de derecho penal a partir de la ley contra el soborno en Corea del Sur. Finalmente, Pongpanut Osathitporn et al. (2017) proponen un sistema de adquisición del código penal tailandés utilizando un diseño ontológico.

El modelado ontológico en el ámbito del derecho penal, por tanto, ofrece un gran potencial. Creemos que la conceptualización de entidades, eventos y cualidades para la creación de una subontología de derecho penal basada en la semántica profunda, como es el caso de *GlobalCrimeterm* (Felices, 2015, 2016), recientemente concluida, y *GroupCrimeTerm* (en desarrollo), y que detallamos en el Apartado 3, pueden contribuir a ampliar el repertorio existente.

3. METODOLOGÍA Y CRITERIOS PARA LA CREACIÓN ONTOLÓGICA: CORPUS Y EXTRACCIÓN DE TÉRMINOS

GroupCrimeTerm es una subontología de derecho penal en fase de construcción que está diseñada sobre la base del corpus especializado *Globalcrimeterm Corpus* (GCTC)⁴. El objetivo último de esta ontología, como hemos mencionado anteriormente, consiste en ofrecer una forma fácil de recuperación de información en contextos profesionales y la consecución de otras tareas subsidiarias como la resolución de problemas en situaciones reales del ámbito jurídico. A continuación, explicaremos sucintamente tanto la construcción del corpus, como la extracción de las unidades terminológicas a partir del mismo. Estas unidades terminológicas constituirán la estructura principal de la Ontología Satélite.

Uno de los principales cometidos a la hora de recopilar un corpus representativo en cualquier rama de conocimiento consiste en la selección de las fuentes y repositorios de textos, o, como en el caso que no ocupa, de textos de tipo jurídico (Bowker y Pearson, 2002; Koester, 2010). Para la construcción del GCTC se ha recurrido a algunas instituciones interesadas en la lucha contra la delincuencia organizada a nivel internacional y así contar con el libre acceso a documentación jurídica en formato digital. En concreto, se han seleccionado los siguientes organismos: Naciones Unidas, el Tribunal de Justicia Penal Internacional, *Europol*, *Eurojust* y la OSCE, entre otros. Se han incluido asimismo recursos textuales de tipo académico debido a la concentración de términos especializados que contienen. El volumen de textos recopilados (621) ha requerido la creación de una base de datos en la que poder organizar la información más relevante de los textos y así facilitar su identificación (Figura 2). En concreto, la base de datos muestra los siguientes apartados: a) el campo "ID" asigna un código numérico único a cada texto; b) el segundo campo contiene información sobre el idioma en el que está escrito el texto; c) "Breve descripción" ofrece una explicación breve sobre el contenido del texto; d) el campo "Título" proporciona una etiqueta que resume el tema específico del documento; e) el campo "Tema" registra el dominio al que pertenece el texto; en el caso del GCTC, se especifica "Delincuencia organizada"; f) "Tipo de documento" contiene información sobre el tipo de texto (por ejemplo, acción conjunta, convenio, acta, procedimientos, etc.); y g) "Fuente" añade una referencia a la fuente de la que se extrajo el documento original.

ID	Language	Brief description	Title	Topic	Type of document	Source
1	English	Fight against organised crime	EOAct (joint) law enforcement cooperation	Organised Crime	Joint Action	Eur-Lex
2	Spanish	Fight against organised crime	SOAct (joint) law enforcement cooperation	Organised Crime	Joint Action	Eur-Lex

Figura 2. Base de datos del corpus.

En lo que se refiere a la fase de extracción, *FunGramKB Suite*, es decir, la plataforma en línea para la edición de la base de conocimiento FunGramKB, incluye el extractor DEXTER (*Discovering and EXtracting TERminology*), una herramienta para la recuperación automática de términos a partir de un corpus representativo de un dominio especializado dado. La Figura 3 muestra el menú principal del DEXTER en la que aparecen las funciones más importantes de la herramienta:

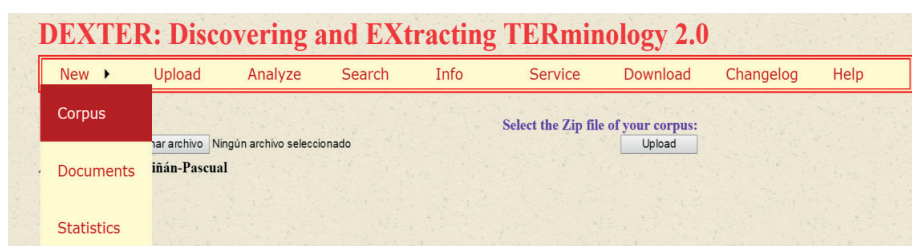


Figura 3. Menú principal de DEXTER (Extractor de términos de FunGramKB).

Para la extracción de estas unidades especializadas, DEXTER actúa en dos fases diferenciadas. En la primera de ellas, la herramienta aplica distintos filtros lingüísticos al corpus de entrada, entre los que se encuentra la eliminación de caracteres no textuales, así como de números y signos de puntuación; igualmente, en el filtrado léxico se lleva a cabo la aplicación de las conocidas como *stop words*, es decir, unidades léxicas que, de forma apriorística e independientemente de la selección de los textos, se eliminan del corpus antes de proceder a las tareas de procesamiento lingüístico propiamente dichas. Para nuestro objetivo de conformar una lista de expresiones terminológicas pertenecientes específicamente al ámbito del derecho, hemos aplicado una lista de *stop words* en la que incluye palabras de tipo funcional, tales como pronombres, adverbios, conjunciones, etc., ya que éstas no pueden formar unidades especializadas de forma aislada, aunque sí pueden aparecer en formaciones multipalabra. en cuyo caso la lista de *stop words* no actúa. Igualmente, forman parte del filtrado léxico del corpus

⁴ La robustez del GCTC reside tanto en la diversidad de textos recogidos como en el elevado número de tokens (5,698,754 tokens en inglés). Entre los 45 tipos de textos que aparecen en el corpus se incluyen informes, acuerdos, declaraciones, reglamentos, leyes, tratados, resoluciones y artículos de revistas científicas especializadas.

inicial un gran número de expresiones pertenecientes al ámbito del lenguaje coloquial o del conocimiento genérico tales como *silla*, *sol*, *manzana*, etc., y que no poseen en ningún caso un significado terminológico en el discurso especializado. Es sobre la base de este corpus depurado donde, en una segunda fase, DEXTER aplica la métrica SRC (*Saliency, Relevance and Cohesion*)⁵, que será la que marque en qué medida una unidad léxica se postula como específica de la rama de conocimiento que trata de representar el corpus – en cuyo caso, el índice SRC será alto – o si por el contrario pertenece al lenguaje común. Como resultado de aplicar la métrica SRC, finalmente, DEXTER ofrece al terminólogo una lista de términos candidatos clasificados según su peso semántico, de modo que los candidatos que aparecen en la parte más alta de la lista se consideran términos estadísticamente más relevantes, mientras que los elementos que muestran un índice inferior a 3 no se consideran especializados. La utilidad de la herramienta en el proceso de creación ontológica resulta evidente, pues permite obtener y gestionar información terminológica actualizada de una forma rápida, evitando así consumir demasiados recursos humanos y de tiempo. En el caso que nos ocupa, una vez concluido el proceso de extracción y filtrado mediante DEXTER según los pasos indicados anteriormente, el número total de n-gramas relevantes ha quedado reducido drásticamente a 57.502 tokens, una cifra muy inferior al corpus entrada.

Pese a las ventajas de una herramienta como DEXTER, es necesario apuntar que, una vez concluida la fase de extracción automática, es necesario llevar a cabo un cribado manual, en el que lo terminólogos decide qué unidades candidatas deben formar parte de la lista final de expresiones propiamente terminológicas y cuáles quedan excluidas. Para ello, en esta fase se aborda el análisis semántico de los candidatos basándose en cuatro criterios básicamente. En primer lugar, la relevancia estadística otorgada por DEXTER a cada unidad candidata, ya que, sobre todo en el caso de las aquellas con un índice SRC visiblemente alto o bajo, servirá para discernir fácilmente entre una unidad léxica no especializada de un término propiamente dicho. En segundo lugar, la idoneidad de un candidato según la evidencia ontológica previa, es decir, el terminólogo, informado a su vez por el ingeniero de conocimiento, valora si la inclusión o exclusión de la lista de términos produce un enriquecimiento de conocimiento en la ontología o, por el contrario, una redundancia. En tercer lugar, el terminólogo deberá realizar un análisis pormenorizado de la semántica de la unidad candidata y dilucidar si las proposiciones que completan su definición son de tipo especializado o más bien de conocimiento común. En cuarto lugar, finalmente, la identificación de una unidad como terminológica está basada en la aparición o no de ésta en un diccionario especializado del dominio meta. Una vez aplicados estos criterios al corpus que nos ocupa el número final de candidatos ganadores (conceptos especializados) ha sido de 406. Entre estas unidades ganadoras cabe destacar DRUG (representado en FunGramKB como +DRUG_00), pues es uno de los conceptos básicos superordinados más relevantes en el mapa conceptual de la ontología sobre crimen organizado debido a su impacto en el ámbito de la actividad criminal.

4. EL CONCEPTO +DRUG_00: CONCEPTUALIZACIÓN Y JERARQUIZACIÓN

La metodología COHERENT (en inglés, Conceptualización, HiErarchization, REmodeling and refineNT), diseñada por Perrián y Mairal (2011) para la construcción de la Ontología Nuclear de FunGramKB, ha servido para el desarrollo posterior de Ontologías Satélite (Carrión Delgado, 2012, Felices 2015, 2016; Felices y Alameda, 2017). Así, por ejemplo, para la creación de la Ontología Nuclear, y siguiendo esta metodología, se identificaron los conceptos básicos incluidos en el *Longman Defining Vocabulary* (LDV) recogidos a su vez en el *Longman Dictionary of Contemporary English* (Procter, 1978). Aunque el LDV ha demostrado ser un referente en la compilación de vocabulario básico para una lengua artificial, fue necesario hacer una revisión en profundidad del mismo para conformar el mapa conceptual para posteriormente incluirlo en la base de conocimiento. Como resultado de la aplicación de COHERENT, se recopiló un catálogo de aproximadamente 1.300 conceptos básicos que constituyó la piedra angular sobre la que posteriormente poblar la Ontología Nuclear con conceptos terminales (proceso aún en curso). De hecho, el nodo final de la jerarquía conceptual se ha enriquecido con la integración de conceptos terminales vinculados a *GroupCrimeTerm*. Si tomamos como ejemplo el concepto básico +DRUG_00 (subordinado al metaconcepto #SELF_CONNECTED_OBJECT) y los conceptos básicos y terminales dependientes de él, se puede configurar un dominio semántico para desarrollar el mapa conceptual de esta Ontología Satélite.

Seguidamente vamos a introducir las dos primeras fases del método COHERENT (conceptualización y jerarquización). Las dos fases restantes (remodelación y refinamiento) se aplicarían en la fase final del diseño ontológico y serán elaboradas por el conjunto de investigadores que forman parte del proyecto denominado “Representación formal de una ontología sobre derecho penal y su integración en una base de conocimiento para la comprensión del lenguaje natural” (véase la sección agradecimientos).

⁵ La métrica SRC se fundamenta en tres nociones: a) Importancia: indica el carácter único o la prevalencia de un término en los datos recopilados; b) Pertinencia: mide el uso de un término mediante la comparación de un corpus específico de un dominio con un corpus de propósito general, y c) Cohesión: cuantifica el grado de estabilidad de los términos multipalabra.

La remodelación afecta a los conceptos verbales dependientes de la subontología #EVENT, pero no las entidades en cuanto conceptos nominales. Respecto al refinamiento, la productividad conceptual de +DRUG_00 está garantizada por un gran porcentaje de conceptos subordinados⁶.

4.1. La conceptualización de +DRUG_00

La primera impresión que tenemos cuando pensamos en el concepto “droga” es que nadie necesita ser un experto para saber cuál es su significado, pues dicho concepto forma parte del conocimiento general del hablante medio de cualquier lengua. La definición de DRUG que mostramos en la Figura 4 es el resultado del análisis de las referencias lexicográficas consultadas y representa el conocimiento medio de un hablante acerca de este concepto⁷; de hecho, DRUG se ubica en la Ontología Nuclear y no en la Ontología Satélite en construcción.

El siguiente paso consiste en transformar la definición en lenguaje natural de DRUG al lenguaje de interfaz COREL, de tal modo que ésta sea comprensible para el sistema computacional. En FunGramKB el conocimiento semántico de los conceptos se expresa mediante lo que se conoce como “Postulados de Significado” (PS), es decir, un conjunto de una o más predicaciones conectadas lógicamente que incluyen los rasgos semánticos del concepto. En los PS cada predicación va precedida de una “e” (véase Figura 4). En este caso, +DRUG_00 va precedido del símbolo “+” como concepto básico y posee cuatro predicaciones (e1, e2, e3, e4)⁸; su definición se presenta sobre la base de otros conceptos básicos (+BE_01, +LEGAL_00, +INGEST_00, etc.), especificándose en cada caso los papeles temáticos que representa cada concepto (*theme, referent, location, result, attribute*). El *genus* de este PS es un caso especial, ya que la droga puede presentarse como un sólido, como líquido o como gas (+GAS_00 ^ LIQUID_00 ^ SOLID_00)⁹. Las predicaciones e2 y e3 van precedidas de un asterisco que implica que la información de las mismas no es estricta, sino rebatible, es decir, que no todas las drogas necesariamente infringen la ley o son perjudiciales, pero normalmente lo son. La “n” que precede a “BE_01” es un operador de negación.

Información conceptual	
CONCEPTO	+DRUG_00
SUPERORDINADO(S):	+GAS_00 ^ +LIQUID_00 ^ +SOLID_00
POSTULADO DE SIGNIFICADO (PS):	+(e1: +BE_00 (x1: +DRUG_00)Theme (x2: +SOLID_00 ^ +LIQUID_00 ^ +GAS_00)Referent) *(e2: n +BE_01 (x1)Theme (x3: +LEGAL_00)Attribute) *(e3: +INGEST_00 (x4)Agent (x1)Theme (x5)Location (x6)Origin (x7)Goal (f1: (e4: +BE_01 (x4)Theme (x8: +HAPPY_00 +NERVOUS_00)Attribute))Result)
DEFINICIÓN EN LENGUAJE NATURAL:	<i>An illegal substance which some people take in order to feel happy, relaxed, or excited</i>

Figura 4: Información conceptual de +DRUG_00 en el editor de FunGramKB.

4.2. La jerarquización de +DRUG_00

La jerarquización es la siguiente etapa en la aplicación de la metodología COHERENT y consiste en atribuir a cada concepto definido en FunGramKB su correspondiente concepto superordinado, así como sus conceptos subordinados y hermanos. Cada concepto subordinado puede a su vez tener uno o más conceptos hermanos, que se caracterizan por compartir rasgos semánticos comunes heredados del superordinado. Esta distribución de conceptos se denomina subsunción “IS-A”, el único tipo de subsunción posible en FunGramKB. Aunque la propia estructura jerárquica de la Ontología Nuclear es exclusivamente de tipo taxonómico, FunGramKB puede albergar otro tipo de relaciones de significado tales como la meronimia (*part_of*) o la subcategorización como, por ejemplo, “subtipo de” (*kind_of*). Todas estas relaciones, no obstante, quedarían definidas en el PS de los propios conceptos (no en la taxonomía nuclear) y solo en aquellos casos en los que éstas fueran parte fundamental de la semántica de estos. La razón para excluir de la Ontología Nuclear este tipo de relaciones más específicas se fundamenta sobre los potenciales problemas en el proceso de herencia de rasgos no necesariamente presentes en todas las unidades que participan de una relación conceptual como la meronimia (Periñán y Arcas, 2010b). Del mismo modo, también se mejora y simplifica la propia organización de la ontología sin perder por ello información (Periñán, 2013). A modo de ejemplo, a continuación se ilustra la definición del concepto “hoja” (+LEAF_00) en FunGramKB y cuya primera proposición (“e1”) recoge su relación partonímica con el concepto “planta” (+PLANT_00):

⁶ Entre esos conceptos subordinados a +DRUG_00, sólo hemos escogido tres ejemplos por razón de representatividad, pero entre otros muchos pueden citarse: PCP, Buprenorphine, Ethylphenidate, Amphetamin, Kratom, Psylocibin, Tramadol, Barbiturate, Desomorphine, Ketamine, Crack, Buprenorphine, Morphine, Ecstasy, Codeine, Heroin, Tramadol, Mescaline, LSD, Crank, etc.

⁷ Cambridge Advanced Learner’s Dictionary; Longman Dictionary of Contemporary English; Oxford Advanced Learner’s Dictionary; Black’s Law Dictionary; Oxford Law Dictionary; Wiktionary.

⁸ La traducción al lenguaje natural de estas cuatro predicaciones en COREL sería el siguiente: 1. La droga es sólo sólido, sólo líquido o sólo gas; 2. Normalmente no es legal; 3. Se introduce a propósito (en un ser vivo); 4. Para excitarse o sentirse feliz.

⁹ El símbolo ^ es un conector lógico que representa la exclusión (a^b^c): sólo a, sólo b o sólo c. En el PS de +DRUG_00 también se aprecia otro conector lógico, el de disyunción | (a | b | c).

+LEAF_00
 MEANING POSTULATE:
 *(e1: +BE_00 (x1: +LEAF_00)Theme (x2: +PLANT_PART_00)Referent)
 *(e2: +BE_01 (x1)Theme (x3: +FLAT_00 & +THIN_00)Attribute)
 *(e3: +BE_01 (x1)Theme (x4: +GREEN_00 ^ +BROWN_00)Attribute)
 +(e4: +ABSORB_00 (x5: +PLANT_00)Agent (x6: +AIR_00)Theme (x7)Location (x8)Origin (x5)Goal (f1: x1)Instrument)

En cuanto a la ruta conceptual y el proceso de jerarquización de +DRUG_00, ambos pueden verse en la Figura 5: +DRUG_00 pertenece a la subontología #ENTITY y a las dimensiones metacognitivas #PHYSICAL, #OBJECT y #SELF_CONNECTED_OBJECT. Su concepto básico superior inmediato es +GAS_00 (o +LIQUID_00, o +SOLID_00) que, a su vez, depende de +SUBSTANCE_00. Esta organización jerárquica implica que +DRUG_00 hereda todas las propiedades conceptuales de los conceptos superordinados mediante un proceso de herencia múltiple no monotónica.

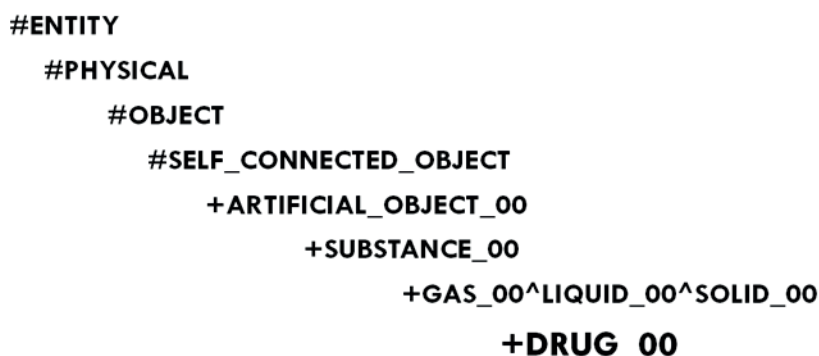


Figura 5: Representación de la ruta conceptual de +DRUG_00 en la Ontología Nuclear.

Asimismo, el concepto +DRUG_00 se convierte en un concepto básico *genus* de una gran cantidad de sustancias narcóticas que se definen a partir de él, convirtiéndose éstas en conceptos terminales (y hermanos entre sí) tanto en la Ontología Nuclear (algunos de ellos) como en las Ontologías Satélite, dependiendo de si la especialización de su definición es mayor o menor. Seguidamente, hemos seleccionado varios conceptos referidos a drogas que pueden ser ampliamente conocidas para el hablante medio, pero que en función de la descripción de sus rasgos técnicos serán conocidas entonces por el hablante experto, y se constituyen en lo que se conoce como conceptos espejo en la ontología especializada¹⁰. En concreto, \$COCAINE_00, \$METHAMPHETAMINE_00 y \$CANNABIS_00 pueden ubicarse también en la subontología *GroupCrimeTerm* en función de la granularidad de sus predicaciones en el PS, fruto del conocimiento experto aportado por fuentes especializadas¹¹. A continuación, se muestran las predicaciones de los tres conceptos subordinados a +DRUG_00 junto a su representación en COREL y una traducción aproximada en lenguaje natural:

\$COCAINE_00
La cocaína es una droga
 +(e1: +BE_00(x1: \$COCAINE_00)Theme (x2: +DRUG_00)Referent)
Se manufactura con hojas de coca
 +((e2: \$MANUFACTURE_00 (x3)Theme (x1)Referent (f1: \$COCA_00)Means) (e3: +BE_00 (x4)Theme (x5: +LEAF_00))
Da placer y energía
 *(e4: +GIVE_00 (x1)Agent (x6: +PLEASURE_00 & m+ENERGY_00)Theme (x1)Origin (x7) Goal)
(Generalmente) Se inhala por la nariz
 *(e5: +TAKE_00 (x1)Theme (x2)Referent (f2:+NOSE_00)Means)
No es legal consumirla o traficar con ella
 *(e6: +SELL_00 | +BUY_00 (x8)Agent (x1)Theme (x6)Origin (x7)Goal (x9: +LEGAL_N_00))

¹⁰ El concepto espejo es un concepto que puede ser utilizado tanto en una Ontología Nuclear como en una Ontología Satélite. La Ontología Nuclear almacena la información compartida por la mayoría de hablantes de un idioma y la satélite expandiría el PS básico incluido en la Ontología Nuclear mediante la adición de las predicaciones necesarias para representar el conocimiento experto.

¹¹ Entre muchas otras fuentes, merece destacarse la información extraída del *National Institute on Drug Abuse* de EE.UU.

\$METHAMPHETAMINE_00

La metanfetamina es una droga

+(e1: +BE_00 (x1: \$METHAMPHETAMINE_00)Theme (x2: +DRUG_00)Referent)

Aumenta el placer y la energía

+(e2: +GIVE_00 (x3)Agent (x4: m+EXCITEMENT_00 & m+ENERGY_00)Theme (x1)Origin (x5: +HUMAN_00)Goal

(Generalmente) Se ingiere por la boca

*(e3: +TAKE_00 (x1)Theme (x2)Referent (f1:+MOUTH_00)Means)

No es legal consumirla o traficar con ella

+(e4: +SELL_00 | +BUY_00 (x6)Agent (x1)Theme (x6)Origin (x7)Goal (x8: +LEGAL_N_00))

\$CANNABIS_00

(Generalmente) El cannabis es una droga

*(e1: +BE_00 (x1: \$CANNABIS_00) Theme (x2: +DRUG_00)Referent

(Generalmente) Se fuma

*(e2: +SMOKE_00 (x4)Theme (x1)Referent)

Da placer y sosiego

*(e3: +GIVE_00 (x1)Agent (x5: +PLEASURE_00 & +CALM_00)Theme (x1)Origin (x4)Goal)

No es legal consumirla o traficar con ella

*(e4: +SELL_00 | +BUY_00 (x6)Agent (x1)Theme (x6)Origin (x7)Goal (x8: +LEGAL_N_00))

Los tres conceptos hermanos comparten una serie de características comunes como muestra la representación de su significado mediante COREL. En los tres casos nos encontramos ante conceptos terminales con una ruta conceptual similar y con un PS que incluye un mínimo de cuatro predicaciones (e1, e2, e3, e4...) y al menos un satélite (f1), excepto \$CANNABIS_00. La predicación *(e: +SELL_00 | +BUY_00 (x)Agent (x)Theme (x)Origin (x)Goal (x: +LEGAL_N_00)) es compartida en los tres casos y se trata de una predicación rebatible, ya que la casuística en torno a la tolerancia o la prohibición del consumo y, en menor medida, del tráfico de estupefacientes (en general) es muy diversa según los países. A partir de conceptos básicos como +ENERGY_00 o +CALM_00 en los PS puede inferirse que \$COCAINE_00 y \$METHAMPHETAMINE_00 son drogas catalogadas como estimulantes en el primer caso, mientras que \$CANNABIS_00 se considera un reactivo depresivo en el segundo. Finalmente, en cada PS se incluye una predicación que establece el medio de consumo más habitual de la sustancia: mediante inhalación nasal (cocaína), consumo oral (metanfetamina) o fumando (cannabis). Evidentemente, la granularidad de estos conceptos podría extenderse más en el caso de ontologías especializadas en química, pero para una subontología vinculada al derecho penal puede ser suficiente.

5. CONCLUSIONES

Este estudio ha procurado describir las fases de conceptualización y jerarquización en el proceso de compilación de una subontología especializada en construcción: *GroupCrimeTerm*, que se sirve del corpus *GlobalCrimeterm* (GCTC) y se enlaza con la arquitectura de la Ontología Nuclear integrada en FunGramKB, una base de conocimiento cuyo propósito es servir como plataforma para el desarrollo de aplicaciones relacionadas con el procesamiento del lenguaje natural y el razonamiento artificial. La formalización de conceptos en la base de conocimiento se lleva a cabo mediante la denominada "metodología COHERENT", así como mediante el lenguaje de interfaz denominado COREL, que nos ha permitido la representación del significado de una forma computacional. Para el presente artículo, y a modo de ejemplo, hemos analizando semánticamente el concepto básico DRUG, así como los conceptos COCAINE, METHAMPHETAMINE y CANNABIS. Estos tres conceptos terminales ampliamente representativos del dominio de las sustancias narcóticas están subordinados a DRUG mediante mecanismos de inferencia y subsunción. El contenido semántico de los conceptos seleccionados sirve de ejemplo para mostrar el inicio del aparato conceptual que en fases posteriores del desarrollo del proyecto de construcción ontológica tiene como objetivo, por una parte, la automatización en el proceso de recuperación de información a partir de unas pautas concretas y, por otro, ayudar a prevenir la actividad delictiva del crimen organizado. Es cierto que ambos objetivos necesitan una fase experimental lo suficientemente amplia para presentarlos como alcanzables a corto plazo. No obstante, el diseño y aplicaciones de algunas herramientas en FunGramKB para el PLN¹², como son ARTEMIS, DAMIEN o PORTER, además de DEXTER, invitan al optimismo.

¹² Las herramientas aludidas pueden conocerse con más detalle consultando el siguiente enlace de Internet <http://www.fungramkb.com/bibliography2.aspx>

AGRADECIMIENTOS

Este artículo forma parte del proyecto de investigación I+D+i titulado “Representación formal de una ontología sobre derecho penal y su integración en una base de conocimiento para la comprensión del lenguaje natural”, financiado en el marco del Programa Operativo FEDER Andalucía 2014-2020 (Junta de Andalucía). Código: B-HUM-177-UGR18

REFERENCIAS

- Asaro, C., Biasiotti, M.A., Guidotti, P., Papini, M., Sagri, M.T., Tiscornia, D. (2003). “A Domain Ontology: Italian Crime Ontology”. *ICAIL 2003 Workshop on Legal Ontologies & Web based legal information management*, 1-7.
- Bak, J., Jedrzejek, C., Falkowski, M. (2010). “Application of an Ontology-Based and Rule-Based Model to Selected Economic Crimes: Fraudulent Disbursement and Money Laundering”, en M. Dean, J. Hall, A. Rotolo, S. & S. Tabet (eds.) *RuleML 2010. LNCS*, vol. 6403. Heidelberg: Springer, 210-22. https://doi.org/10.1007/978-3-642-16289-3_18
- Bezzazi, E. H. (2007). “Building an Ontology That Helps Identify Criminal Law Articles That Apply to a Cybercrime Case”, en *ICSOFT (PL/DPS/KE/MUSE)*, 179-185.
- Bowker, L., Pearson, J. (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*. Londres, Nueva York: Routledge. <https://doi.org/10.4324/9780203469255>
- Breuker, J., Winkels, R. (2003). “Use and reuse of legal ontologies in Knowledge Engineering and Information Management”. *ICAIL03 Wks on Legal Ontologies and Web-based Information Management*, Edimburgo, <http://iri.jur.uva.nl/~winkels/legont/ICAIL2003.html>
- Breuker J., Casanovas, P., Klein, M.A.C., Francesconi, E. (eds.) (2008). *Law, Ontologies and the Semantic Web*. Amsterdam: IOS Press.
- Carrión, M.G. (2012). “Extracción y análisis de unidades léxico-conceptuales del dominio jurídico: un acercamiento metodológico desde FunGramKB”, *RaeL*, 11, 25-39.
- Cortés, F., Mairal, R. (2016). “Building an RRG computational grammar”, *Onomázein*, 34, 86-117. <https://doi.org/10.7764/onomazein.34.22>
- Felices, A. (2015). “Foundational considerations for the development of the Globalcrimeterm subontology: A research project based on FunGramKB”, *Onomázein*, 31, 127-144. <https://doi.org/10.7764/onomazein.31.9>
- Felices, A. (2016). “The Process of Constructing Ontological Meaning Based on Criminal Law Verbs”, *Círculo de Lingüística Aplicada a la Comunicación*, 65, 109-148. https://doi.org/10.5209/rev_CLAC.2016.v65.51983
- Felices, A., Ureña, P. (2012). “Fundamentos metodológicos de la creación subontológica en FunGramKB”, *Onomázein*, 26, 49-67.
- Felices, A., Ureña, P. (2014). “FunGramKB Term Extractor: a key instrument for building a satellite ontology based on a specialized corpus”. En B. Nolan & C. Perrián (eds.) *Language processing and grammars: The role of functionally oriented computational models*. Amsterdam: John Benjamins, 251-269. <https://doi.org/10.1075/slcs.150.10fel>
- Felices, A., Alameda, A. (2017). “The process of building the upper-level hierarchy for the aircraft structure ontology to be integrated in FunGramKB”, *Revista de Lenguas para Fines Específicos*, 23(2), 86-110.
- Gangemi A., Sagri MT., Tiscornia D. (2005) A Constructive Framework for Legal Ontologies. In: Benjamins V.R., Casanovas P., Breuker J., Gangemi A. (eds) *Law and the Semantic Web. Lecture Notes in Computer Science*, vol 3369. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-32253-5_7
- Gruber, T.R. (1993). “A translation approach to portable ontologies”. *Knowledge Acquisition*, 5(2), 199-220. <https://doi.org/10.1006/knac.1993.1008>
- Jiménez, R., Luzondo, A. (2011). “Building Ontological Meaning in a Lexico-conceptual Knowledge Base”, *Onomázein*, 23, 11-40.
- Koester, A. (2010). “Building small specialised corpora”. En A. O’ Keeffe & M. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*. Londres: Routledge, 66-79. <https://doi.org/10.4324/9780203856949.ch6>
- Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Ogonowski, A. Peters, I., Peters, W., Ruimy, N., Villegas, M., Zampolli, A. (2000). “SIMPLE: A general framework for the development of multilingual lexicon”, *International Journal of Lexicography*, 13/4, 249-263. <https://doi.org/10.1093/ijl/13.4.249>
- Martín-Mingorance, L. (1994). “La lexicografía onomasiológica”. En Hernández Hernández, H. (ed.) *Aspectos de lexicografía contemporánea*. Barcelona: Biglograf, 15-28.
- Musen, M.A. (1992). “Dimensions of knowledge sharing and reuse”. *Computers and Biomedical Research* 25, 435-467. [https://doi.org/10.1016/0010-4809\(92\)90003-S](https://doi.org/10.1016/0010-4809(92)90003-S)

- Niles, I., Pease, A. (2001). "Towards a standard Upper Ontology". En *Proceedings of the Second International Conference on Formal Ontology in Information Systems*. Ogunquit. <https://doi.org/10.1145/505168.505170>.
- Osathitporn, P., Soonthornphisaj, N., Vatanawood, W. (2017). "A scheme of criminal law knowledge acquisition using ontology". En *2017 18th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*. IEEE. 29-34. <https://doi.org/10.1109/SNPD.2017.8022696>
- Periñán, C. (2012). "En defensa del procesamiento del lenguaje natural fundamentado en la lingüística teórica", *Onomázein*, 26, 13-48.
- Periñán, C. (2013). "A knowledge-engineering approach to the cognitive categorization of lexical meaning", *VIAL-Vigo International Journal of Applied Linguistics*, 10, 85-104.
- Periñán, C. (2015). "The underpinnings of a composite measure for automatic term extraction: the case of SRC", *Terminology*, 21/2, 151-179. <https://doi.org/10.1075/term.21.2.02per>
- Periñán, C. (2017). "Bridging the gap within text-data analytics: a computer environment for data analysis in linguistic research", *Revista de Lenguas para Fines Específicos*, 23/2, 111-132.
- Periñán, C. (2018). "DEXTER: a workbench for automatic term extraction with specialized corpora", *Natural Language Engineering*, 24/2, 163-198. <https://doi.org/10.1017/S1351324917000365>
- Periñán, C., Arcas, F. (2004). "Meaning postulates in a lexico-conceptual knowledge base". En *15th International Workshop on Databases and Expert Systems Applications, IEEE*, Los Alamitos (California), 38-42.
- Periñán, C., Arcas, F. (2010a). The architecture of FunGramKB. *Proceedings of the 7th International Conference on Language Resources and Evaluation, European Language Resources Association (ELRA)*, 2667-2674.
- Periñán, C., Arcas, F. (2010b). "Ontological commitments in FunGramKB", *Procesamiento del Lenguaje Natural*, 44, 27-34.
- Periñán, C., Mairal, R. (2010). "La gramática de COREL: un lenguaje de representación conceptual", *Onomázein*, 21, 11-45.
- Periñán, C., Arcas, F. (2014). "The implementation of the FunGramKB CLS Constructor". En B. Nolan y C. Periñán (eds.), *Language Processing and Grammars: The Role of Functionally Oriented Computational Models*. Amsterdam: John Benjamins, 165-169. <https://doi.org/10.1075/slcs.150.07per>
- Periñán, C., Arcas, F. (2019). "Detecting environmentally-related problems on Twitter", *Biosystems Engineering*, 177, 31-48. <https://doi.org/10.1016/j.biosystemseng.2018.10.001>
- Periñán, C., Mairal, R. (2010). "Enhancing UniArab with FunGramKB", *Procesamiento del Lenguaje Natural*, 44, 19-26.
- Periñán, C., Mairal, R. (2011). "The COHERENT Methodology in FunGramKB", *Onomázein*, 24, 13-33.
- Procter, P. (Ed.) (1978). *Longman Dictionary of Contemporary English*. Harlow: Longman.
- Ruiz de Mendoza, F., Mairal, R. (2008). "Levels of description and constraining factors in meaning construction: an Introduction to the Lexical Constructional Model", *Folia Linguistica*, 42/2, 355-400. <https://doi.org/10.1515/FLIN.2008.355>
- San Martín, A., Faber, P. (2014). "Deep semantic representation in a domain-specific ontology: Linking EcoLexicon to FunGramKB". En B. Nolan & C. Periñán (eds.) *Language Processing and Grammars: The Role of Functionally Oriented Computational Models*. Amsterdam/Filadelfia: John Benjamins, 271-296. <https://doi.org/10.1075/slcs.150.11mar>
- Sartor, G., Casanovas, P., Biasiotti, M.A., Fernández-Barrera, M. (eds.) (2011) *Approaches to Legal Ontologies, Theories, Domains, Methodologies*. Berlín: Springer. <https://doi.org/10.1007/978-94-007-0120-5>
- Soh, C., Lim, S., Hong, K., Rhim, Y.Y. (2018). Ontology Modeling for Criminal Law. En U. Pagallo, M. Palmirani, P. Casanovas, G. Sartor, S. Villata (eds.) *AI Approaches to the Complexity of Legal Systems AICOL 2015, AICOL 2016, AICOL 2017. Lecture Notes in Computer Science*, vol 10791. Cham: Springer, 365-379. https://doi.org/10.1007/978-3-030-00178-0_25
- Valente, A., Breuker, J. (1994). "A functional view of law", en G. Bargellini & S. Binazzi (eds.) *Towards a Global Expert System in Law*. Padua: CEDAM Publishers.
- Valente, A. (2005). "Types and Roles of Legal Ontologies". En V.R. Benjamins, Casanovas, P., Breuker, j., Gangemi, A. (eds.), *Law and the Semantic Web*. Berlín: Springer, 65-76. https://doi.org/10.1007/978-3-540-32253-5_5
- Van Valin, R.D. (2005). *The Syntax-Semantics-Pragmatics Interface: An Introduction to Role and Reference Grammar*, Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511610578.002>
- Van Valin, R.D., LaPolla, R.J. (1997). *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139166799>
- Velardi, P., Pazienza, M.T., Fasolo, M. (1991). "How to encode semantic Knowledge: a method for meaning representation and computer-aided acquisition", *Computational Linguistics*, 17/2, 153-170.