



Fundamentos metodológicos de la creación subontológica en FunGramKB

Methodological underpinnings for the creation of subontologies in FunGramKB

Ángel Felices Lago

Universidad de Granada
España

Pedro Ureña Gómez-Moreno

Universidad de Granada
España

Resumen

Este artículo describe las bases metodológicas para la construcción de subontologías terminológicas en la base de conocimiento FunGramKB y se parte de la hipótesis de que el modelo multinivel de la Ontología Nuclear (nivel metaconceptual, nivel básico y nivel terminal) puede conectarse a una subontología con el fin de minimizar la redundancia informativa y maximizar el conocimiento. En el marco de FunGramKB, las subontologías se caracterizan, en primer lugar, por contener conceptos pertenecientes a dominios de conocimiento especializado y, en segundo lugar, por orientar esos conceptos a la semántica profunda –a diferencia del enfoque de la gran mayoría de las ontologías terminológicas, que suelen orientarse a la semántica superficial. El desarrollo de las subontologías permitirá la aplicación futura de FunGramKB a tareas de procesamiento de lenguaje especializado y razonamiento experto.

Palabras clave: FunGramKB; ontología; terminología; lingüística de corpus.

Afiliación: Ángel Felices: Departamento de Filologías Inglesa y Alemana, Facultad de Ciencias Económicas y Empresariales, Universidad de Granada. – Pedro Ureña: Departamento de Filologías Inglesa y Alemana, Facultad de Ciencias Económicas y Empresariales, Universidad de Granada.

Correos electrónicos: afelices@ugr.es; pedrou@ugr.es

Dirección postal: Ángel Felices y Pedro Ureña: Campus de Cartuja s/n, 18071 Granada, España.

Fecha de recepción: febrero de 2012

Fecha de aceptación: septiembre de 2012

Abstract

This paper describes the methodological aspects of the construction of terminological subontologies within the architecture of FunGramKB. The main hypothesis is that the multilevel model of the Nuclear Ontology in FunGramKB (metaconceptual level, basic level and terminal level) can be connected to a terminological subontology in order to minimize redundancy and maximize information knowledge. One of the features of this subontology is, firstly, its implementation for the creation of a specialized module in the Ontology and, secondly, its focus on deep semantics, in contrast with the vast majority of terminological ontologies, which are based on surface semantics.

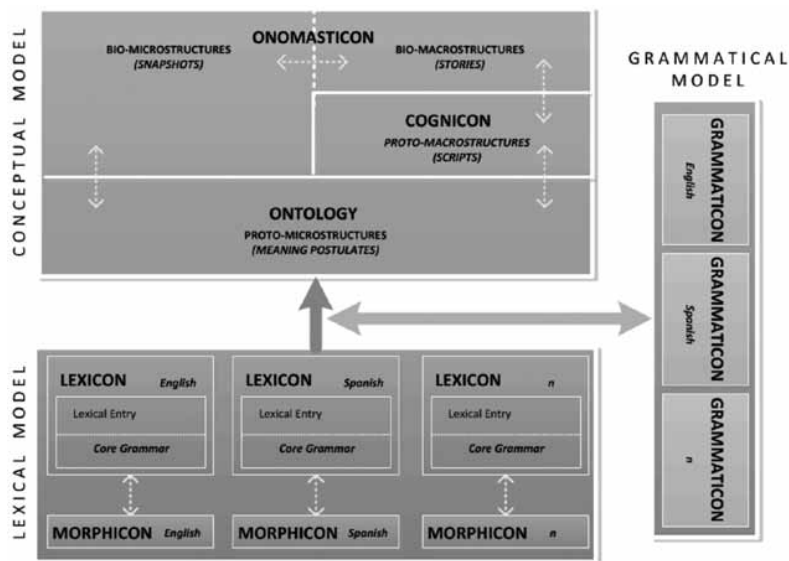
Keywords: FunGramKB; ontology; terminology; corpus linguistics.

1. Introducción¹

FunGramKB es una base de conocimiento que ha sido diseñada para su aplicación en tareas de procesamiento del lenguaje natural (Periñán y Arcas, 2004, 2005, 2006, 2007) y cuya estructura se compone de tres módulos –conceptual, léxico y gramatical– interconectados entre sí. Una de las características distintivas de esta base de conocimiento reside en su módulo conceptual y, en concreto, en su componente esencial denominado “Ontología Nuclear”, que se construye sobre los principios de la semántica profunda y postula la definición de conceptos mediante relaciones complejas de herencia e inferencia (Periñán y Arcas, 2007). Esta perspectiva se contrapone a la semántica superficial, que se fundamenta en asociaciones simples entre unidades léxicas o relaciones de significado generales como la meronimia. En consecuencia, la Ontología Nuclear constituye, junto al Cognición y el Onomástico, el eje principal del módulo conceptual de FunGramKB, ya que permite el funcionamiento del razonador artificial (Periñán y Arcas, 2005). Además, otros componentes como el Lexicón y el Gramaticón añaden el soporte léxico-gramatical de las unidades conceptuales en un variado número de lenguas como el inglés, español, italiano, búlgaro, etc. Los tres módulos que integran FunGramKB aparecen representados en la Figura 1:

¹ Esta contribución forma parte del proyecto de investigación “Elaboración de una subontología terminológica en un contexto multilingüe (español, inglés e italiano) a partir de la base de conocimiento FunGramKB en el ámbito de la cooperación internacional en materia penal: terrorismo y crimen organizado”, financiado por el Ministerio de Ciencia e Innovación de España. Código: FFI2010-15983.

FIGURA 1
La estructura modular de FunGramKB

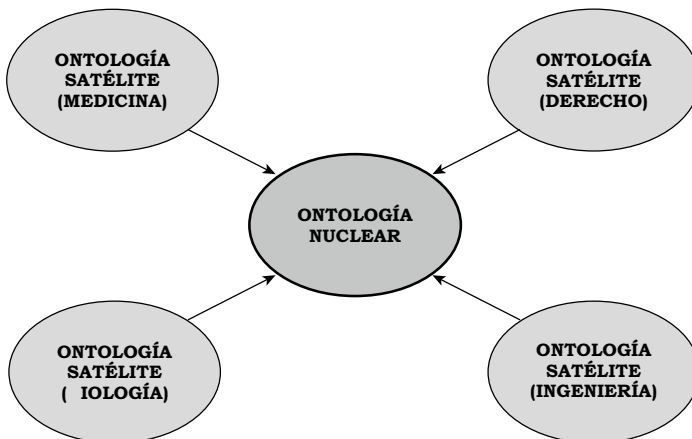


Un aspecto fundamental para el desarrollo de FunGramKB consiste en la progresiva ampliación del módulo conceptual, que, si bien ya se ha aplicado con éxito en labores de razonamiento artificial (Periñán y Arcas, 2005) y traducción automática (Periñán y Mairal, 2010), actualmente incorpora aproximadamente 1.300 conceptos pertenecientes exclusivamente al sentido común o conocimiento humano básico pero carece de conceptos propios de áreas temáticas más especializadas². Precisamente, por esta razón, el objetivo principal de este artículo es proponer una metodología para la construcción de subontologías que doten a FunGramKB de un mayor conocimiento del mundo. Las subontologías, también denominadas “ontologías satélites” (Periñán y Arcas, 2010a), incluyen términos y conceptos pertenecientes a dominios especializados del conocimiento, ya sean del ámbito sociológico-humanístico, como el derecho o la antropología, o bien del científico-técnico, como la medicina o la ingeniería.

² Los conceptos que conforman la Ontología Nuclear han sido obtenidos a partir del vocabulario definitorio del *Longman Dictionary of Contemporary English* (LDOCE).

La metodología de construcción subontológica consta de tres fases: (i) compilación de un corpus especializado; (ii) extracción automática de terminología, esto es, de unidades léxicas especializadas, a partir del corpus, y (iii) modelado del conocimiento especializado mediante acuñación y jerarquización de conceptos. Estas fases aparecerán ilustradas en el presente artículo con ejemplos pertenecientes al área del derecho, uno de los campos de investigación más activos en materia de creación ontológica (Valente, 2003; Breuker y otros, 2005; Breuker y otros, 2008; Breuker y otros, 2009). Los objetivos de la metodología son, por un lado, poblar el módulo conceptual especializado de FunGramKB y, por otro, desarrollar los lexicones terminológicos de distintas lenguas. Para ello, este artículo parte de la hipótesis de que el modelo multinivel de la Ontología Nuclear, que distingue entre metaconceptos, conceptos básicos y terminales (Periñán y Arcas, 2004), puede conectarse a un modelo subontológico, puesto que las ontologías satélites comparten estructura con la Ontología Nuclear, pese a las diferencias de orientación y refinamiento (Figura 2):

FIGURA 2
La integración de las ontologías satélites en la Ontología Nuclear de FunGramKB



La importancia de esta propuesta de desarrollo subontológico es doble. En primer lugar, permite aumentar el repositorio conceptual, así como incrementar la expresividad cognitiva

de la base de conocimiento. En segundo lugar, la inclusión de subontologías en el módulo conceptual contribuye a ampliar el campo de aplicación de FunGramKB hacia ámbitos que requieren formación especializada.

La propuesta metodológica de este artículo hunde sus raíces en varios marcos teóricos. Por un lado, se parte del trabajo fundacional de Perinán y Arcas (2004, 2005, 2006, 2007), en el que se desarrollan los aspectos esenciales de FunGramKB y el lenguaje de representación semántica COREL (*Conceptual Representation Language*) con el que se definen los conceptos de la Ontología (Perinán y Mairal, 2010). Por otro lado, se inspira también en los presupuestos teóricos de los enfoques cognitivo y funcional, principalmente de la Gramática del Papel y la Referencia (Van Valin y LaPolla, 1997; Van Valin, 2005) y el Modelo Léxico Construccional (Ruiz de Mendoza y Mairal, 2008; Mairal y Ruiz de Mendoza, 2008, 2009).

Este estudio está estructurado de la siguiente forma. En primer término ilustra el primero de los pasos de la creación subontológica, en el que se señalan los aspectos más relevantes de la recopilación de corpus especializados. En segundo lugar propone un procedimiento novedoso para la identificación de terminología de un dominio de conocimiento a partir de los corpus. En un tercer apartado expone e ilustra las características principales de la conceptualización y jerarquización de los conceptos encapsulados a nivel léxico en la terminología. En última instancia resume las conclusiones más importantes de este artículo y señala algunas propuestas para la aplicación de las subontologías en etapas futuras.

2. Compilación de un corpus especializado

El primer paso en la metodología subontológica consiste en recopilar un corpus de textos representativo del dominio especializado objeto de estudio. El corpus habrá de contener el mayor número posible de términos, que posteriormente se conceptualizarán y formarán la subontología. Para ello es necesario previamente crear un repositorio de textos en formato electrónico que proceda de fuentes solventes, e instituciones nacionales e internacionales expertas en la materia que vaya someterse a conceptualización (Bowker y Pearson, 2002; Koester, 2010). Identificar un grupo significativo de textos relevantes es

una tarea que se hace más compleja en ámbitos multilingües (cf. Aguado de Cea y otros, 2007; Peters y otros, 2007; Perriñán y Arcas, 2007, 2008; Mairal y Ruiz de Mendoza, 2009; Perriñán y Mairal, 2009). Por ejemplo, para la creación de la subontología de crimen organizado y terrorismo descrita en Felices y otros (2011), se construyó el corpus GCTC (*Global Crime Term Corpus*), formado por documentos de organismos especializados en derecho penal internacional y la lucha contra estos delitos, tales como EUROPOL, EUROJUST, Consejo de Europa o la Corte Penal Internacional, entre otros. Además de la solvencia de las fuentes, otro criterio fundamental para la creación de un corpus especializado es su representatividad. El corpus debe ser una muestra de una amplia variedad de tipos de textos que reflejen de forma global la temática y el léxico propios de un dominio. En la compilación del componente inglés del GCTC, por ejemplo, se incluyeron aproximadamente 45 clases de documentos pertenecientes al ámbito jurídico, legislativo y criminalístico, entre ellos, reglamentos, tratados, acuerdos y ensayos académicos. Finalmente, un tercer criterio de compilación es el balance, es decir, el corpus ha de contener una proporción ponderada de los distintos tipos textuales que lo forman, así como del número de palabras que cada uno aporta al conjunto del corpus. Estos tres criterios son de especial relevancia puesto que permitirán disponer de una base léxica sólida a partir de la cual se obtendría una terminología representativa.

3. Extracción terminológica

La elaboración de subontologías en FunGramKB requiere del análisis de los conceptos que representan el conocimiento experto, así como de la definición de los términos que lexicalizan dichos conceptos. Sin embargo, identificar las unidades léxicas que pertenecen a un dominio especializado es una tarea compleja en la que la consulta bibliográfica, la mera introspección o el análisis de las concordancias que forman un corpus resultan a menudo poco eficientes. En FunGramKB, la recuperación de terminología se realiza mediante la herramienta *FunGramKB Term Extractor* (Figura 3), incluida en el paquete informático de edición de la base de conocimiento (Perriñán y Arcas, 2004, 2010a, 2010b), y cuya función es obtener de forma automática una lista de términos representativos de una rama de conocimiento a partir de un corpus.

FIGURA 3
Menú principal de *FunGramKB Extractor*

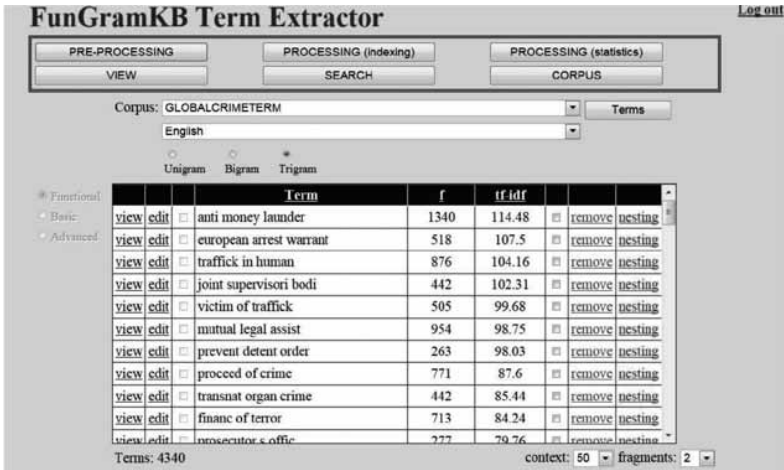


El extractor adopta un acercamiento estadístico basado en el cálculo del índice *tf-idf* (*term frequency inverse document frequency*) de todas las unidades léxicas que componen el corpus. Este valor numérico es el resultado de ponderar la frecuencia absoluta de aparición de las unidades léxicas con respecto a su frecuencia de aparición en cada uno de los textos que conforman el corpus. Éstas reciben un índice *tf-idf* más elevado cuanto mayor sea su relevancia estadística y según sea mayor la probabilidad de postularse como unidades con entidad terminológica. El extractor puede identificar tres tipos de unidades potencialmente “candidatas”: (i) unigramas, esto es, unidades formadas normalmente por una palabra como, por ejemplo, “resolución” o “tratado”; (ii) secuencias bigramáticas o de dos palabras como, por ejemplo, “resolución unilateral” o “acuerdo bilateral”; y (iii), combinaciones trigramáticas o de tres palabras como “recurso de alzada” o “malversación de fondos”. El procesamiento estadístico realiza un primer filtrado de los elementos del corpus eliminando unidades cuya frecuencia de aparición absoluta es menor de 3, así como elementos funcionales, tales como las preposiciones o los pronombres, y caracteres no lingüísticos como las cifras numéricas o los caracteres ortográficos. El extractor dispone además de una herramienta para el filtrado automático de unigramas de uso frecuente, que obtiene una lista más reducida de los términos potencialmente relevantes.

Tanto el filtro estadístico como el filtro de unidades de uso frecuente simplifican drásticamente el proceso de extracción terminológica. No obstante, una vez concluido el procesamiento automático, es necesario realizar una edición manual en la que el terminólogo decidirá qué candidatos son finalmente relevantes para su posterior conceptualización, y cuáles han de considerarse no terminológicos o “candidatos falsos” (Figura 4).

El extractor incluye dos opciones para la eliminación manual de candidatos falsos. El “borrado simple” (*remove*) elimina

FIGURA 4

Menú principal de edición de ngramas en *FunGramKB Term Extractor*

unidades individuales, mientras que el denominado “borrado anidado” (*nesting*) descarta bigramas o trigramas individuales junto con los componentes que lo forman y las distintas combinatorias que tienen lugar entre los mismos. La Tabla 1 muestra ejemplos respectivos de ambas formas de filtrado³.

Para esta labor de edición manual, el terminólogo puede adoptar dos acercamientos metodológicos: el apriorístico o el inductivo. En el acercamiento apriorístico, el filtrado manual se realiza sobre la base de un esquema léxico-conceptual en el que aparecen, de forma predeterminada, una relación de las entidades, eventos y cualidades que tienen lugar de forma prototípica en el dominio. El terminólogo construirá este mapa cognitivo partiendo de la investigación bibliográfica o el asesoramiento de expertos. En el caso, por ejemplo, del derecho penal, el terminólogo puede guiarse por marcos epistemológicos de ontologías precedentes o teorías jurídicas que especifiquen cuáles son el tipo de entidades y eventos que suceden de forma recurrente en la fenomenología jurídica (Breuker y otros, 2005). En el acercamiento inductivo,

³ La interfaz de la Suite de FunGramKB está escrita en inglés. En la tabla aparecen entre paréntesis los nombres originales en inglés de las distintas opciones de filtrado.

TABLA 1
Edición de ngramas en *FunGramKB Term Extractor*

| Tipo de filtro | Término candidato | Resultado tras el filtrado: Elementos eliminados |
|-------------------------------|---|---|
| Simple (<i>remove</i>) | <i>encuesta</i> | <i>Encuesta</i> |
| Anidado (<i>nesting</i>) | <i>elementos disuasorios urgentes</i> | <i>elementos disuasorios urgentes elementos disuasorios disuasorios urgentes elementos disuasorios Urgentes</i> |

por el contrario, la identificación de los términos que representan el dominio objeto de estudio se realiza a partir de los resultados obtenidos en el proceso de extracción exclusivamente.

4. Conceptualización y jerarquización

Una vez identificada la terminología de un dominio especializado, el siguiente paso de la metodología consiste en la población de la subontología mediante la proyección de las unidades terminológicas a unidades cognitivas. Como resultado, se creará un conjunto de conceptos básicos, terminales y subconceptos (“conceptualización”), y se establecerá para cada uno de ellos cuáles son las unidades que los lexicalizan, así como el lugar que ocupan en los lexicones de las distintas lenguas (“jerarquización”). Parte de este procedimiento se basa en la metodología de modelado ontológico nuclear (Periñán y Arcas, 2010b; Jiménez y Luzondo, 2011). No obstante, dadas las características del conocimiento especializado, a continuación se indican aspectos que son propios del modelado de este tipo de conocimiento (Mairal y otros, 2011).

En FunGramKB, la conceptualización se define como el proceso de construcción mediante lenguaje COREL de los correspondientes *aktionsarten* (o “marcos temáticos) y de las proposiciones de significado (o “postulados de significado”). Los marcos temáticos consituyen la definición cognitiva prototípica de los eventos, mientras que los postulados semánticos contienen

información semántica esencial para definir las propiedades de un concepto específico, sea un evento, una entidad o una cualidad. En las subontologías, la construcción de postulados se realiza sobre la base de una o más definiciones léxicas obtenidas de diccionarios especializados. Por ejemplo, (1) muestra el postulado de significado de \$WATERBOARDING_00, concepto característico de los ámbitos del terrorismo y el crimen organizado que hace referencia a un acto en el que un humano inflige una tortura sobre una víctima mediante el uso de técnicas de asfixia con líquido o tejido:

- (1) *Términos*: “waterboarding” (inglés), “submarino” (español), “waterboarding” (italiano).

Concepto: \$WATERBOARDING_00.

Postulado de significado:

+(e1: +BE_00 (x1: \$WATERBOARDING_00)Theme (x2: +TORTURE_00)Referent)

*(e2: +CHANGE_00 (x3: \$TERRORIST_00 ^ +HUMAN_00)Theme (x4: \$VICTIM_00)Referent (f1: (e3: +FEEL_00 (x3)Agent (x4)Theme (x5: +DAMAGE_00 | +PAIN_00 | +FEAR_00)Attribute))Result (f2: (e4: +SUFFER_00 (x4)Theme))Result) (f3: (e5: n +MOVE_00 (x4)Referent))Condition (f4: (e6: +DESCEND_00 (x3)Agent (x5: +LIQUID_00)Theme (x6)Location (x7)Origin (x8: +HEAD_00)Goal))Means (f5: (e7: n +BREATH_00 (x4)Theme))Result) (f6: (e8: +EXPERIENCE_00 (x3)Agent (x4)Theme (x2)Referent (f7: (e9: ing +DROWN_00 (x1)Theme (x2)Referent))Result)

*(e10: +PUT_00 (x3)Agent (x9: +CLOTH_00 ^ +BAG_00)Theme (x10)Origin (x8)Goal)

*Descripción lexicográfica en lenguaje natural*⁴:

A form of torture in which water is poured over the face of a supine, immobilized victim whose head is pulled back so that the victim cannot avoid inhaling water, and thus experiences the sensation of drowning. In some variations, fabric or plastic may be draped over the victim’s face or the victim may be gagged before the water is poured. [*Black’s Law Dictionary* (BLD)]⁵

⁴ Para la conceptualización de términos pueden utilizarse diccionarios de cualquier lengua (Jiménez y Luzondo-Uyón, 2011). En este artículo ilustraremos los ejemplos mediante definiciones tomadas de diccionarios especializados monolingües en inglés.

⁵ “Tipo de tortura en la que se vierte agua sobre la cara de una víctima que yace inmobilizada, en posición supina y con la cabeza hacia atrás, de modo que no puede evitar la inhalación de agua, y por lo tanto experimenta sensación de ahogamiento. En algunos casos la cabeza de la víctima se cubre con tela o plástico antes de verter el agua”.

A diferencia de *waterboarding*, hay otros tipos de unidades léxicas y conceptos que no pertenecen exclusivamente a un dominio de conocimiento especializado, sino que su ámbito de aparición se extiende con frecuencia al habla coloquial. En dominios de conocimiento como el jurídico, el desarrollo de los medios de comunicación y la transmisión de información han jugado un papel importante en la expansión entre hablantes no especializados de términos pertenecientes a la lucha antiterrorista, como es el caso de “bioterrorismo” o “ántrax”, y también al crimen organizado, como “trata de blancas” o “cártel”. Sin embargo, si bien en estos casos la base del concepto y el referente son los mismos tanto para un jurista como para un no jurista, existe, sobre todo a nivel cuantitativo y cualitativo, un enriquecimiento conceptual que sólo el primero de estos posee. Asimismo, términos como “sentencia”, “derecho” o “testigo” pertenecen al lenguaje común de cualquier hablante medio, sin embargo, sus significados se perciben de manera diferente según se trate de un juez o de un lego en la profesión. Esta diferencia conceptual –de conocimiento, al fin y al cabo– queda reflejada en las diferencias de contenido entre los diccionarios generales y los especializados. El ejemplo (2) muestra una definición general de *sentence* (“sentencia”) obtenida del diccionario general *Cambridge Advanced Learner’s Dictionary* (CALD), mientras que el ejemplo (3) muestra una definición especializada del mismo término extraída del diccionario jurídico *Oxford Law Dictionary* (OLD):

- (2) Sentence: a punishment given by a judge in court to a person or organization after they have been found guilty of doing something wrong.⁶
- (3) Sentence: any order made by a court when dealing with an offender in respect of his offence, including imprisonment (which may take the form of a concurrent sentence or suspended sentence), a fine, a community order, or an absolute or conditional discharge. Criminal sentencing is now governed by the Criminal Justice Act 2003, which requires courts to take into account the following purposes of sentence: punishing offenders, reduction of crime (...).⁷

⁶ “Sentencia: un castigo impuesto por un juez a una persona u organización en un tribunal después de haber sido declarada culpable de cometer un delito”.

⁷ “Sentencia: cualquier resolución dictada por un tribunal cuando se trata de un delincuente y con referencia a su delito, incluida la prisión (que puede tomar la forma de una pena concurrente o condena condicional), una multa, prestación de servicios a la comunidad, una libre absolución o absolución

Las diferencias entre una entrada y otra son evidentes, aun tratándose de dos definiciones de una misma unidad léxico-conceptual. La segunda es cualitativamente más precisa que la primera, además de ser cuantitativamente más amplia. Otro ejemplo representativo de las diferencias entre una definición especializada y otra de tipo general es el concepto +CRIME_00, incluido en la Ontología Nuclear, y lexicalizado como *crime* en inglés, *delito* en español y *delitto* en italiano. El postulado de significado de este concepto, tal y como figura actualmente en la Ontología Nuclear, aparece en (4):

- (4) *Términos*: “crime” (inglés), “crimen” (español), “delitto” (italiano).
Concepto: +CRIME_00 (concepto perteneciente a la Ontología Nuclear).

Postulado de significado:

+ (e1: +BE_00 (x1: +CRIME_00)Theme (x2: +OCCURRENCE_00)Referent)

+ (e2: n +BE_01 (x1)Theme (x3: +LEGAL_00)Attribute)

Descripción en lenguaje natural:

An act punishable by law; usually considered an evil act. [Definición incluida en FunGramKB]

Para su inclusión en una subontología jurídica, la definición especializada de +CRIME_00 requiere una mayor riqueza proposicional que la recogida en las dos únicas proposiciones del postulado de significado en (4), puesto que éstas no aportan información necesaria en la práctica del razonamiento jurídico, como, por ejemplo, la descripción de los tipos generales de delito, los aspectos definitorios de su procesamiento o los agentes necesarios involucrados en su comisión, según sean los propósitos a los que queramos dirigir el razonamiento automático. Ante conceptos como +CRIME_00, que ya figuran en la Ontología Nuclear con un significado semántico general, es necesario crear un “término espejo” en la ontología satélite, es decir, una entidad, evento o cualidad, cuya naturaleza semántica coincida con la del concepto nuclear pero que, a diferencia de éste, muestre una mayor granularidad o especificidad en su postulado de significado. Por tanto, la creación de subontologías no consiste únicamente en hallar conceptos especializados exclusivos de un

condicional. La condena penal se rige por la Ley de Justicia Penal de 2003, que exige a los tribunales que tengan en cuenta los siguientes fines de la pena: castigar a los delincuentes, la reducción de la delincuencia (...).”

ámbito de conocimiento, sino que, al mismo tiempo, requiere diseñar definiciones especializadas de conceptos que están integrados en la Ontología Nuclear con un significado general. El modelado especializado del concepto espejo +CRIME_00 se puede observar en (5):

- (5) *Términos*: “crime” (inglés), “delito” (español), “delitto” (italiano).

Concepto: +CRIME_00 (concepto espejo perteneciente a la ontología satélite).

*Postulado de significado*⁸:

+ (e1: +BE_00 (x1: +CRIME_00)Theme (x2: +OCCURRENCE_00)Referent)

+ (e2: n +BE_01 (x1)Theme (x3: +LEGAL_00)Attribute)

* (e3: +PUNISH_00 (x4: +STATE_00)Theme (x5: \$CRIMINAL_00)Referent) (f1: (e4: +WANT_00 | +TRY_00 (x5)Theme (x1)Referent)Condition ((f2: (e5: +DEMONSTRATE_00 (x6: +LAWYER_00)Theme (x1)Referent (e6: cert +DO_00 (x5)Theme (x1)Referent) (f3: +PROOF_00)Means)Condition)) ((f4: (e7: n +DOUBT_00 (x6)Agent) (e8: n +BE_00 (x5)Agent (x7: \$INNOCENT_00)Referent)Condition)))

* ((e9: +COMPRISE_00 (x1)Theme (x8: +HUMAN_00)Referent (e10: +DO_00 (x5)Theme (x1)Referent))

* ((e11: +COMPRISE_00 (x1)Theme (x9: +HUMAN_00) (e12: +THINK_00 (x5 | x8)Agent (x1)Theme))

* (e13: pos +COMPRISE_00 (x1)Theme (x9: +MURDER_00 | \$RAPE_00)Referent)

* ((e14: +REQUEST_01 (x10: +POLICE_00 | +HUMAN_00)Agent (x11: +JUDGE_00)Theme (e15: +JUDGE_01 & +PUNISH_00 (x11)Theme (x1)Referent (f5: \$SENTENCE_00)Means)Goal))

* (e16: \$ARREST_00 (x12: +POLICE_00)Theme (x5)Referent (f6: (e17: +ORDER_00 (x11)Theme)Condition))

* (e18: +DECIDE_00 (x11)Theme (x13: \$PENALTY_00)Referent (f7: +LAW_00)Means)

* ((e19: +BE_00 (x13)Theme (x14)Referent (e20: +STAY_00 (x1)Theme (x15: +PRISON_00)Location (f8: +OFTEN_00 ^ s +PERIOD_00)Duration ^ (e21: +KILL_00 (x16)Theme (x5)Referent) ^ (e22: +PAY_00 (x5)Theme (x15: \$FINE_00)Referent))

Descripción lexicográfica en lenguaje natural.

⁸ Las dos primeras predicaciones del postulado de significado del concepto espejo se corresponden con el postulado de significado del mismo concepto en la Ontología Nuclear. En FunGramKB ambas predicaciones aparecerán coindexadas pero aquí se señalan de forma explícita para mayor claridad expositiva.

- (i) An act that the law makes punishable; the breach of a legal duty treated as the subject-matter of a criminal proceeding (...) [BLD]⁹
- (ii) An act (or sometimes a failure to act) that is deemed by statute or by the common law to be a public wrong and is therefore punishable by the state in criminal proceedings. Every crime consists of an *actus reus* accompanied by a specified *mens rea* (unless it is a crime of strict liability), and the prosecution must prove these elements of the crime beyond reasonable doubt. Some crimes are serious wrongs of a moral nature (e.g. murder or rape); others interfere with the smooth running of society (e.g. parking offences). Most prosecutions for crime are brought by the police (although they can also be initiated by private people); some require the consent of the Attorney General. Crimes are customarily divided into indictable offences (for trial by judge and jury) and summary offences (for trial by magistrates); some are hybrid. Crimes are also divided into arrestable offences and nonarrestable offences. The punishments for a crime include death (for treason), life imprisonment (e.g. for murder), imprisonment for a specified period, suspended sentences of imprisonment, conditional discharges, probation, binding over, and fines; in most cases judges have discretion in deciding on the punishment. Some crimes may also be civil wrongs; for example, theft and criminal damage are crimes punishable by imprisonment as well as torts for which the victim may claim damages. [OLD]¹⁰

⁹ “Un acto tipificado como delito. Violación de un deber legal entendida como objeto de un procedimiento penal”.

¹⁰ “Un acto (a veces, en grado de tentativa) que se considera de acuerdo con la ley un daño público y es por lo tanto punible por parte del Estado mediante proceso penal. Todo delito se compone de *actus reus* y *mens rea* (a menos que sea un delito de responsabilidad objetiva). La acusación debe probar los elementos del delito más allá de toda duda razonable. Algunos delitos son faltas graves de carácter moral (por ejemplo, asesinato o violación), mientras que otros interfieren con el buen funcionamiento de la sociedad (por ejemplo, delitos de estacionamiento). La mayoría de los delitos son denunciados por la policía (aunque también pueden hacerlo los particulares), y algunos requieren el consentimiento del Fiscal General. Los delitos se suelen dividir en delitos graves (juzgados por un juez y un jurado), faltas (juzgados por magistrados) y delitos híbridos. Se pueden dividir asimismo en delitos de arresto y los que no requieren arresto. Las penas por un delito incluyen la muerte (por traición a la patria), la cadena perpetua (por ejemplo, homicidio), el encarcelamiento por un período de tiempo determinado, condicional, libertad probatoria, fianza, y multas. En la mayoría de los casos los jueces tienen discreción para decidir la pena. Algunos delitos también pueden ser faltas civiles como, por ejemplo, el robo y el vandalismo, que son delitos punibles con pena de prisión, así como agravios por los que la víctima puede reclamar daños y perjuicios”.

FIGURA 5
Interfaz de edición léxico-conceptual de *FunGramKB Term Extractor*

The screenshot shows the 'FunGramKB Term Extractor' interface. It features a top section with three main areas: 'Senses' (an empty text area), 'CONCEPT:' (containing a 'DESCRIPTION:' field and a 'Save' button), and 'METACONCEPT:' (with radio buttons for 'Entities', 'Events', and 'Qualities'). Below these are language-specific sections for English, Spanish, Italian, French, German, Bulgarian, and Catalan, each with a text input field and 'Y'/'N' buttons. At the bottom left, there are 'Rename', 'Delete', 'Done', and 'Duplicate' buttons, and the word 'absorb' is displayed.

Una vez identificados los conceptos especializados, y definidos sus marcos temáticos y postulados de significado, el siguiente paso en la creación subontológica consiste en la jerarquización, es decir, en determinar para cada concepto su correspondiente hiperordinado, subordinado o conceptos hermanos, según sea el caso. En FunGramKB, este proceso de subsunción puede realizarse bien mediante el editor de la Ontología, o bien mediante el propio extractor terminológico, a través de su apartado de edición léxico-conceptual, que funciona a modo de interfaz entre la subontología y la Ontología Nuclear, así como el Cognición, el Onomasticón y el Lexicón (Figura 5).

Las características definitorias de un concepto terminal o subconcepto están especificadas en su correspondiente ruta conceptual, definida como el conjunto de conceptos de rango superior a los que se subordina. En el caso de la entidad \$WATERBOARDING_00, dicha ruta se expresa de la siguiente forma:

- (6) *Concepto*: \$WATERBOARDING_00
Concepto superordinado: +TORTURE_00
Ruta conceptual: #ENTITY > #PHYSICAL > #PROCESS > +OCCURRENCE_00 > \$TORTURE_00

Por otro lado, la ruta conceptual de +CRIME_00 se muestra en (7):

- (7) *Concepto*: +CRIME_00
Concepto superordinado: +OCCURRENCE_00
Ruta conceptual: #ENTITY > #PHYSICAL > #PROCESS > +OCCURRENCE_00 > +CRIME_00

Además de la organización jerarquizada de los conceptos, la interfaz de edición del extractor permite asignar para cada concepto uno o más términos de los obtenidos en el proceso de extracción. Estos términos representan la materialización léxica de los conceptos especializados.

5. Conclusiones

Este artículo ha analizado los aspectos principales de la creación de subontologías en la base de conocimiento FunGramKB. Las subontologías pueden incluir el conocimiento experto necesario para la representación de eventos, entidades y cualidades propios de un ámbito profesional o técnico, y pueden utilizarse también en la ampliación conceptual de la base de conocimiento y de su potencial de razonamiento. Una posible aportación de este estudio ha sido la descripción de las diferencias y similitudes entre las subontologías terminológicas y la Ontología Nuclear de FunGramKB. Según se ha sugerido, ambas comparten una misma estructura multinivel que distingue entre metaconceptos, conceptos básicos y conceptos terminales, si bien las definiciones de estos conceptos en las subontologías son cualitativa y cuantitativamente más complejas que en la Ontología Nuclear. Una segunda aportación ha procurado desarrollar una metodología de propósito general para el diseño subontológico, que podrá utilizarse en el modelado de dominios especializados, independientemente del área de conocimiento a la que estos se circunscriban y de la(s) lengua(s) sobre la que se construyan. La metodología, tal y como se ha ilustrado mediante ejemplos pertenecientes al ámbito del derecho penal, se desarrolla en tres fases: compilación de un

corpus especializado; extracción terminología a partir del corpus, y, finalmente, conceptualización y jerarquización de los términos. El desarrollo de subontologías permitirá orientar FunGramKB hacia tareas futuras relacionadas con la creación de conocimiento, la recuperación de información, o la resolución de problemas de razonamiento concretos, a modo de agente experto.

6. Bibliografía citada

- AGUADO DE CEA, Guadalupe, Elena MONTIEL y José Á. RAMOS, 2007: "Multilingüedad en una aplicación basada en el conocimiento", *Procesamiento del Lenguaje Natural* 38, 77-98.
- BOWKER, Lynne y Jennifer PEARSON, 2002: *Working with specialized language. A practical guide to using corpora*, Londres: Routledge.
- BREUKER, Joost, André VALENTE y Radboud WINKELS, 2005: "Use and reuse of legal ontologies in knowledge engineering and information management" en Richard BENJAMINS, Pompeu CASANOVAS, Joost BREUKER y Aldo GANGEMI, *Law and the semantic web*, Vol. 3369, Berlín: Springer, 36-64.
- BREUKER Joost, Pompeu CASANOVAS y Michel C. A. KLEIN (eds.), 2008: *Law, ontologies and the semantic web*. Amsterdam: IOS Press.
- BREUKER, Joost, Pompeu CASANOVAS, Michel C. A. KLEIN y Enrico FRANCESCO (eds.), 2009: *Law, ontologies and the semantic web. Channelling the legal information flood*, Frontiers in Artificial Intelligence and Applications, Vol. 188, Amsterdam: IOS Press.
- FELICES, Ángel, Pedro UREÑA y Ángela ALAMEDA, 2011: "FunGramKB y la adquisición terminológica", *Anglogermánica Online* 8, 66-86.
- JIMÉNEZ, Rocío y Alba LUZONDO, 2011: "Building ontological meaning in a lexico-conceptual knowledge base", *Onomázein* 23, 11-40.
- KOESTER, Almut, 2010: "Building small specialised corpora" en Anne O'KEEFFE y Michael MCCARTHY (eds.): *The Routledge handbook of corpus linguistics*, Londres: Routledge, 66-79.
- MAIRAL, Ricardo y Francisco J. RUIZ DE MENDOZA, 2008: "New challenges for lexical representation within the Lexical-Constructional Model", *Revista Canaria de Estudios Ingleses* 57, 137-158.
- , 2009: "Levels of description and explanation in meaning construction" en Christopher BUTLER y Javier MARTÍN (eds.): *Deconstructing constructions*, Amsterdam: John Benjamins, 153-198.
- MAIRAL, Ricardo, Carlos PERIÑAN y Eva SAMANIEGO, 2011: "Using ontologies for terminological knowledge representation. A preliminary discussion" en Noa TALAVAN, Elena MARTÍN y Francisco PALAZÓN (eds.): *Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10*, Madrid: UNED, 267-280.
- PETERS, Wim, Elena MONTIEL, Guadalupe AGUADO DE CEA y Asunción GÓMEZ, 2007: "Localizing ontologies in OWL" en *OntoLex 2007*,

workshop en la ISWC 2007 + ASWC 2007. (The 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference 2007), 11-15.

- PERIÑÁN, Carlos y Francisco ARCAS, 2004: "Meaning postulates in a lexico-conceptual knowledge base", en el *15.º International Workshop on Databases and Expert Systems Applications*, IEEE, Los Alamitos (California), 38-42.
- , 2005: "Microconceptual-Knowledge Spreading in FunGramKB". *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*, Anaheim-Calgary-Zurich: ACTA Press, 239-244.
- , 2006: "Reusing computer-oriented lexica as foreign-language electronic dictionaries", *Anglogermánica Online* 4, 69-93.
- , 2007: "Cognitive modules of an NLP knowledge base for language understanding", *Procesamiento del Lenguaje Natural* 39, 197-204.
- , 2008: "A cognitive approach to qualities for NLP", *Procesamiento del Lenguaje Natural* 41, 137-144.
- , 2010a: "The architecture of FunGramKB", en *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA), 2667-2674.
- , 2010b: "Ontological commitments in FunGramKB", *Procesamiento del Lenguaje Natural* 44, 27-34.
- PERIÑÁN, Carlos y Ricardo MAIRAL, 2009: "Bringing Role and Reference Grammar to natural language understanding", *Procesamiento del Lenguaje Natural* 43, 265-273.
- , 2010: "Enhancing UniArab with FunGramKB", *Procesamiento del Lenguaje Natural* 44, 19-26.
- RUIZ DE MENDOZA, Francisco J. y Ricardo MAIRAL, 2008: "Levels of description and constraining factors in meaning construction. An introduction to the Lexical Constructional Model", *Folia Linguistica* Vol. 42, Nº 2, 355-400.
- VALENTE, Andre, 2003: "Types and roles of legal ontologies", *Law and the semantic web*, 65-76.
- VAN VALIN, Robert D. Jr., 2005: *Exploring the syntax-semantics interface*, Cambridge: Cambridge University Press.
- VAN VALIN, Robert D. Jr. y Randy J. LAPOLLA, 1997: *Syntax, structure, meaning and function*, Cambridge: Cambridge University Press.

6.1. Diccionarios y corpus

- BLD: GARNER, Bryan (ed.), 2010: *Black's Law Dictionary*. 9ª edición. Minnesota: West.
- CALD: *Cambridge Advanced Learner's Dictionary* [disponible en línea en <http://dictionary.cambridge.org/dictionary/learner-english>].

GCTC: *Global Crime Term Corpus* [véase FELICES, Ángel, Pedro UREÑA y Ángela Alameda, 2011: “FunGramKB y la adquisición terminológica”, *Anglogermánica Online* 8, 66-86].

LDOCE: *Longman Dictionary of Contemporary English* [disponible en línea en <http://www.ldoceonline.com/>, Noviembre-Diciembre 2011].

OLD: LAW, Jonathan y Elizabeth MARTIN, 2009: *Oxford Law Dictionary*. Oxford: Oxford University Press.

