

**TESIS DOCTORAL**

**DISEÑO Y DESARROLLO DE UN MODELO  
COMPUTACIONAL PARA LA REPRESENTACIÓN DEL  
CONOCIMIENTO EN EL DOMINIO DE LA  
COOPERACIÓN JUDICIAL EN MATERIA PENAL**

**MARÍA DE GRACIA CARRIÓN DELGADO**

Lda. en Filología Inglesa



**DEPARTAMENTO DE FILOLOGÍAS EXTRANJERAS Y SUS  
LINGÜÍSTICAS**

**FACULTAD DE FILOLOGÍA**

**UNIVERSIDAD NACIONAL  
DE EDUCACIÓN A DISTANCIA**

2012

DEPARTAMENTO DE FILOLOGÍAS EXTRANJERAS Y SUS  
LINGÜÍSTICAS

FACULTAD DE FILOLOGÍA

UNIVERSIDAD NACIONAL

DE EDUCACIÓN A DISTANCIA

**DISEÑO Y DESARROLLO DE UN MODELO  
COMPUTACIONAL PARA LA REPRESENTACIÓN  
DEL CONOCIMIENTO EN EL DOMINIO DE LA  
COOPERACIÓN JUDICIAL EN MATERIA PENAL**

Autora

**MARÍA DE GRACIA CARRIÓN DELGADO**

Lda. en Filología Inglesa

Director de la Tesis

**DR. RICARDO MAIRAL USÓN**

Codirector de la Tesis

**DR. CARLOS PERIÑÁN PASCUAL**

*A mis padres y esposo*

*Si alguien te dice que ha buscado el conocimiento y no lo ha encontrado, no le creas. Si te dice que ha alcanzado el conocimiento sin buscarlo, no le creas. Pero si te dice que ha buscado el conocimiento y lo ha alcanzado, puedes creerle. (El Talmud)*



## AGRADECIMIENTOS

En primer lugar, me gustaría dar las gracias al director y al codirector de esta tesis, los profesores Ricardo Mairal Usón y Carlos Perrián Pascual. Al primero, por alimentar en mí las ansias de saber, y al segundo por descubrirme el fascinante mundo de la ingeniería ontológica.

De igual manera, merece una mención especial el director del proyecto\* en el que se enmarca esta tesis, el profesor Ángel Felices Lago, al que agradezco el haberme dado la oportunidad de participar en un proyecto tan apasionante a través de la aportación que supone esta tesis.

Asimismo, quiero dar las gracias a mi querido amigo y profesor Antonio Ballesteros González por sus sabios consejos y por guiarme en los momentos complicados a lo largo de una andadura de años.

Por otra parte, deseo agradecer a mis padres su apoyo y comprensión constantes, porque, si no me hubieran animado a seguir perseverando con mi trabajo, esta tesis no hubiera sido posible. Y hago constar una mención muy especial a mi esposo. Sin su dedicación, sacrificio y entrega incondicional, esta tesis no habría salido adelante.

Por último, y no por ello menos importante, desearía dedicar esta tesis a mi queridísima amiga y mentora Pepa Feu, que, sin estar entre nosotros, no puede estar más presente. Le agradezco el haber despertado en mí la pasión por la Lingüística desde el día en que nos conocimos, cuando yo era estudiante de carrera; y también el haber contado conmigo en su última investigación, mientras realizaba mi Doctorado. Para mí trabajar juntas fue un honor que me ha marcado de manera indeleble. Su amistad, que me transformó como alumna y como persona, me iluminará siempre.

\*Este trabajo forma parte del proyecto de investigación denominado “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. Código: FFI2010-15983.



## ÍNDICE

LISTA DE ABREVIATURAS.....	ix
LISTA DE TABLAS.....	xi
LISTA DE FIGURAS.....	xiii
Capítulo 1: INTRODUCCIÓN.....	1
1.1. La Web Semántica y el derecho.....	1
1.2. Las ontologías legales en la representación de conocimiento legal.....	6
1.3. Contribución de esta tesis.....	7
1.4. Estructura de la tesis.....	8
Capítulo 2: ESTADO DE LA CUESTIÓN.....	11
2.1. Introducción.....	11
2.2. Definición de ontología.....	11
2.3. Tipos de ontologías y criterios de diseño.....	18
2.3.1. Tipos de ontologías basadas en la riqueza de su estructura interna...27	
2.3.2. Tipos de ontologías basadas en el tema de la conceptualización.....28	
2.3.3. Criterios de diseño de ontologías.....33	
2.4. Ontologías y sistemas de librerías ontológicas.....37	
2.5. Las ontologías más significativas.....39	
2.5.1. Ontologías de representación del conocimiento.....39	
2.5.1.1. Frame Ontology (FO) y OKBC.....40	
2.5.1.2. RDF y RDF Schema.....40	
2.5.1.3. OIL.....41	
2.5.1.4. DAML + OIL.....42	
2.5.1.5. OWL.....42	
2.5.2. Ontologías de alto nivel.....43	
2.5.2.1. De universales y particulares.....43	

2.5.2.2. Sowa .....	47
2.5.2.3. CyC .....	48
2.5.2.4. The Standard Upper Ontology (SUO) .....	49
2.5.3. Ontologías lingüísticas .....	51
2.5.3.1. Wordnet .....	52
2.5.3.2. EuroWordNet.....	53
2.5.3.3. The Generalized Upper Model (GUM) .....	54
2.5.3.4. Mikrokosmos .....	55
2.5.3.5. SENSUS .....	57
2.5.3.6. DOLCE.....	58
2.5.4. Ontologías de dominio .....	61
2.5.4.1. De comercio electrónico.....	61
2.5.4.2. Médicas.....	65
2.5.4.3. De empresa .....	74
2.5.4.4. De gestión del conocimiento .....	77
2.5.4.5. Legales.....	81
2.6. Metodologías, lenguajes y herramientas de diseño de ontologías.....	83
2.6.1. Metodologías .....	83
2.6.1.1. El método CyC .....	85
2.6.1.2. El método de Uschold y King .....	86
2.6.1.3. El método de Grüninger y Fox (TOVE).....	87
2.6.1.4. KACTUS .....	89
2.6.1.5. METHONTOLOGY.....	90
2.6.1.6. SENSUS .....	96
2.6.1.7. On-To-Knowledge.....	101
2.6.1.8. COHERENT.....	103
2.6.2. Métodos de evaluación de ontologías.....	110

---

2.6.2.1. Terminología de evaluación de ontologías.....	110
2.6.2.2. Evaluación de la taxonomía.....	111
2.6.2.3. OntoClean.....	112
2.6.3. Lenguajes de implementación de ontologías.....	115
2.6.4. Herramientas de desarrollo de ontologías .....	119
2.7. Estado de la cuestión de las ontologías legales .....	123
2.7.1. Primeras conceptualizaciones legales.....	125
2.7.1.1. NORMA .....	125
2.7.1.2. LLD .....	126
2.7.2. Ontologías legales existentes.....	127
2.7.2.1. FOLaw .....	127
2.7.2.2. FBO .....	131
2.7.2.3. Knowledge Based Model of Law .....	134
2.7.2.4. LRI-Core.....	136
2.7.2.5. LKIF-Core .....	139
2.7.2.6. JurWordnet .....	141
2.7.2.7. OPLK.....	143
2.7.2.8. OPJK.....	147
2.7.2.9. Lame’s French Codes .....	148
2.7.2.10. Italian Crime Ontology.....	149
2.7.2.11. CLIME.....	150
2.7.2.12. IPROnto .....	152
2.7.2.13. Laymen .....	153
2.7.2.14. Ontology of Dutch Criminal Law (OCLN_LN).....	153
2.7.2.15. DALOS .....	154
2.8. Conclusiones.....	162
Capítulo 3: FUNGRAMKB.....	165
3.1. Introducción.....	165

3.2. Las unidades conceptuales.....	170
3.3. La gramática de COREL: el marco temático y el postulado de significado.....	175
3.3.1. Las predicaciones .....	177
3.3.2. Los operadores de razonamiento y del evento.....	179
3.3.3. Los operadores de participante: los cuantificadores y los operadores lógicos.....	182
3.4. Conclusiones.....	184
Capítulo 4: DESARROLLO DE UNA ONTOLOGÍA SATÉLITE EN FUNGRAMKB A PARTIR DEL GLOBAL CRIME TERM CORPUS (GCTC) .....	187
4.1. Introducción.....	187
4.2. La extracción terminológica .....	195
4.3. La conceptualización de los términos.....	211
4.3.1. De los diccionarios a la elaboración de definiciones en lenguaje natural y COREL: Casos de estudio .....	215
4.3.1.1. Conceptos espejo .....	216
4.3.1.2. Conceptos especializados .....	232
4.3.1.3. Conclusiones.....	240
4.4. La jerarquización cognitiva .....	242
4.5. Conclusiones.....	246
Capítulo 5: CONCLUSIONES Y SUGERENCIAS PARA POSTERIORES INVESTIGACIONES .....	249
5.1. Conclusiones.....	249
5.2. Sugerencias para posteriores investigaciones.....	253
Bibliografía.....	255
Apéndices .....	287
Apéndice 1 .....	287
Apéndice 2 .....	302

## **LISTA DE ABREVIATURAS**

BLD	Black's Law Dictionary
CALD	Cambridge Advanced Learner's Dictionary
CED	Collins English Dictionary
COREL	Conceptual Representation Language
DTJ	Diccionario de términos jurídicos de Alcaraz y Hughes
DBTJ	Diccionario bilingüe de términos jurídicos de Bossini
Eurojust	European Union's Judicial Cooperation Unit
Europol	European Police
GCTC	Global Crime Term Corpus
GPR	Gramática del Papel y la Referencia
IA	Inteligencia Artificial
LDCE	Longman Dictionary of Contemporary English
MT	Marco Temático
ODL	Oxford Dictionary of Law
OSCE	Organization for Security and Cooperation in Europe
PLN	Procesamiento del Lenguaje Natural
PS	Postulado de Significado
UN	United Nations



## LISTA DE TABLAS

Tabla 2.1: <i>Knowledge Based Model of Law</i> : Perspectiva general de la ontología del dominio legal basada en el conocimiento .....	135
Tabla 2.2: <i>LKIF-Core</i> : Los diez primeros términos de acuerdo a su importancia, nivel nivel de abstracción y relevancia legal .....	141
Tabla 2.3: Extracto del cuestionario de <i>OPLK</i> .....	144
Tabla 2.4: <i>OPLK</i> : Lista de respuestas con la palabra “guardia” ( <i>on duty</i> ) .....	145
Tabla 2.5: Los 10 conceptos más frecuentes en <i>CLIME</i> y sus relaciones .....	151
Tabla 2.6: Las ontologías legales .....	159
Tabla 2.7: Las ontologías legales (cont.) .....	160
Tabla 2.8: Las ontologías legales (cont.) .....	161
Tabla 3.1: Los operadores de predicación en <i>FunGramKB</i> .....	180
Tabla 3.2. Los operadores de cuantificación en <i>FunGramKB</i> .....	182
Tabla 4.1: Algoritmo para el filtrado terminológico .....	205-206



## LISTA DE FIGURAS

Figura 1.1: La arquitectura de la Web Semántica .....	2
Figura 1.2: El nuevo iPad en la Web actual .....	4
Figura 2.1: Ontologías y compromisos ontológicos.....	20
Figura 2.2: Adaptación de la categorización de Mizouguchi y otros .....	21
Figura 2.3: La representación de la categorización de ontologías.....	24
Figura 2.4: Tipos de ontologías .....	24
Figura 2.5: El espectro de la ontología.....	25
Figura 2.6: Ejemplos de ontologías de alto nivel .....	29-30
Figura 2.7: Tipología de ontologías de Casellas .....	32
Figura 2.8: Los modelos intencionados como reflejo de su compromiso con una conceptualización .....	35
Figura 2.9: El término “médico” en <i>MultiWordNet</i> .....	36
Figura 2.10: El problema de la reutilización y el uso de las ontologías .....	38
Figura 2.11: Captura de pantalla de la librería de ontologías <i>DAML</i> .....	38
Figura 2.12: Las capas de <i>OIL</i> .....	42
Figura 2.13: La relación entre particulares y universales .....	44
Figura 2.14: La taxonomía de clases de la ontología de alto nivel de universales.....	45
Figura 2.15: Vista parcial de la taxonomía de clases de la ontología de alto nivel de particulares .....	46
Figura 2.16: La ontología de alto nivel Sowa.....	47
Figura 2.17: Fragmento de la taxonomía de clases de la ontología de alto nivel <i>Cyc</i> .....	49
Figura 2.18: La estructura de los primeros niveles de <i>SUMO</i> .....	50
Figura 2.19: La estructura modular de <i>SUMO</i> .....	50

Figura 2.20: Extracto de la categoría de sustantivos en <i>Wordnet</i> .....	52
Figura 2.21: Los primeros niveles de las jerarquías de <i>GUM</i> .....	55
Figura 2.22: La taxonomía de clases en <i>Mikrokosmos</i> .....	56
Figura 2.23: La estrategia de fusión de <i>SENSUS</i> .....	57
Figura 2.24: La taxonomía de categorías básicas de particulares en <i>DOLCE</i> .....	60
Figura 2.25: Muestra parcial de la clasificación de <i>RosettaNet</i> .....	65
Figura 2.26: La ontología de alto nivel <i>GALEN CORE</i> .....	66
Figura 2.27: Parte de la red semántica de la ontología <i>UMLS</i> .....	69
Figura 2.28: La red de inclusión de la librería de ontologías <i>ON9</i> .....	70
Figura 2.29: El editor de ontologías de <i>Ontoterm</i> .....	72
Figura 2.30: <i>Ontoterm</i> : Submódulo para la integración de conceptos .....	72
Figura 2.31: <i>Ontoterm</i> : Submódulo de visualización de conceptos: conceptos genéricos.. .....	73
Figura 2.32: <i>Ontoterm</i> : Editor de la base de datos terminológica.....	73
Figura 2.33: <i>Ontoterm</i> : Representación de la información conceptual en HTML.....	74
Figura 2.34: Vista parcial de la taxonomía de clases en la <i>Enterprise Ontology</i> .....	75
Figura 2.35: Estructura de las ontologías <i>TOVE</i> .....	77
Figura 2.36: Las ontologías de información, dominio y empresa en una memoria corporativa.....	79
Figura 2.37: Fragmento de la ontología de documentación de los proyectos de investigación y desarrollo.....	81
Figura 2.38: La representación gráfica de las relaciones terminológicas en las..... metodologías .....	84
Figura 2.39: Los procesos de la metodología de Uschold y King.....	86
Figura 2.40: Los procesos de la metodología de Grüninger y Fox.....	88
Figura 2.41: La metodología <i>KACTUS</i> .....	89
Figura 2.42: Proceso de desarrollo de ontologías en <i>METHONTOLOGY</i> .....	91

Figura 2.43: Ciclo de vida de la ontología en <i>METHONTOLOGY</i> .....	93
Figura 2.44: Tareas de conceptualización en <i>METHONTOLOGY</i> .....	94
Figura 2.45: <i>SENSUS</i> : Proceso 1. Identificar términos clave.....	97
Figura 2.46: <i>SENSUS</i> : Proceso 2. Vincular los términos clave a <i>SENSUS</i> .....	97
Figura 2.47: <i>SENSUS</i> : Proceso 3. Añadir caminos al núcleo.....	98
Figura 2.48: <i>SENSUS</i> : Proceso 4. Añadir nuevos términos del dominio .....	99
Figura 2.49: <i>SENSUS</i> : Proceso 5. Añadir subárboles completos .....	100
Figura 2.50: Los procesos en <i>On-To-Knowledge</i> .....	103
Figura 2.51: La metodología <i>COHERENT</i> .....	104
Figura 2.52: Tipología de propiedades basada en las meta-propiedades de <i>OntoClean</i> .....	108
Figura 2.53: Tipos de errores susceptibles de realizar cuando se desarrollan taxonomías con marcos.....	112
Figura 2.54: Lenguajes de ontologías tradicionales .....	115
Figura 2.55: Lenguajes de marcado ontológico.....	116
Figura 2.56: Papeles funcionales del conocimiento legal en el funcionamiento del sistema legal .....	128
Figura 2.57: Categorías de conocimiento en <i>FOLaw</i> .....	131
Figura 2.58: Distribución de las capas de estatus ontológico y papeles epistémicos en las ontologías legales.....	134
Figura 2.59: Las capas en <i>LRI-Core</i> en Breuker y Hoekstra.....	137
Figura 2.60: Las primeras dos capas de <i>LRI-Core</i> .....	138
Figura 2.61: Ejemplo de clases físicas en <i>LRI-Core</i> .....	138
Figura 2.62: La estructura de <i>JUR-IWN</i> .....	143
Figura 2.63: <i>OPLK</i> : Parte del conocimiento profesional relevante para los jueces noveles.....	146
Figura 2.64: <i>OPLK</i> : La ontología legal en <i>Protégé</i> .....	147
Figura 2.65: Capas de la ontología penal italiana.....	149

Figura 2.66: Ontología penal italiana de Asaro y otros .....	150
Figura 2.67: Las fuentes de conocimiento de <i>IPROnto</i> .....	152
Figura 2.68: El “anclado” de conceptos en <i>OCL.NL/LRI-Core</i> .....	154
Figura 2.69: El sistema de organización del conocimiento ( <i>KOS</i> ) del recurso DALOS..... .....	155
Figura 3.1: El planeta cognitivo en <i>FunGramKB</i> .....	166
Figura 3.2: La arquitectura modular de <i>FunGramKB</i> .....	168
Figura 3.3: La ontología nuclear y las ontologías satélites en <i>FunGramKB</i> .....	169
Figura 3.4: Jerarquía conceptual en <i>FunGramKB</i> .....	170
Figura 3.5: Representación parcial de los metaconceptos de los eventos en <i>FunGramKB</i> .....	171
Figura 3.6: Representación parcial de los metaconceptos de las entidades en <i>FunGramKB</i> .....	171
Figura 3.7: Representación parcial de los metaconceptos de las cualidades en <i>FunGramKB</i> .....	172
Figura 3.8: Representación parcial de los eventos en <i>FunGramKB</i> .....	173
Figura 3.9: Representación parcial de las entidades en <i>FunGramKB</i> .....	173
Figura 3.10: Representación parcial de las cualidades en <i>FunGramKB</i> .....	174
Figura 3.11: Operadores de aspectualidad en <i>FunGramKB</i> .....	180
Figura 3.12: Operadores de temporalidad en <i>FunGramKB</i> .....	181
Figura 3.13: Los conectores lógicos en <i>FunGramKB</i> .....	183
Figura 4.1: Base de datos del GCTC .....	192
Figura 4.2: Proceso de compilación y filtrado en <i>FunGramKB</i> .....	195
Figura 4.3: Pantalla principal del extractor de <i>FunGramKB</i> .....	196
Figura 4.4: Corpus y papelera de reciclaje terminológica del extractor de <i>FunGramKB</i> .....	197
Figura 4.5: Pantalla de visualización del corpus del extractor de <i>FunGramKB</i> .....	197

Figura 4.6: Pantalla de precarga del extractor de *FunGramKB*.....198

Figura 4.7: Pantalla de carga del extractor de *FunGramKB*.....199

Figura 4.8: Componentes de GCTC .....200

Figura 4.9: Tabla de unigramas candidatos en el extractor de *FunGramKB*.....202

Figura 4.10: Tabla de bigramas candidatos en el extractor de *FunGramKB*.....202

Figura 4.11: Tabla de trigramas candidatos en el extractor de *FunGramKB* .....203

Figura 4.12: Papelera de reciclaje terminológica en el extractor de *FunGramKB*. .....204

Figura 4.13: Herramienta de edición de términos ganadores en el extractor de *FunGramKB*. .....210



# Capítulo 1

## INTRODUCCIÓN

### 1.1. La Web Semántica y el derecho

El desarrollo de la Web Semántica ha propiciado que una gran cantidad de información legal esté disponible en la Web. Para ello se hace necesario estructurar dicha información de una forma correcta para su comprensión. Un modo de ordenar y organizar la información es mediante estructuras conceptuales de representación del conocimiento para compartir y gestionar de modo inteligente toda esta información a la vez que se facilita la comunicación entre hombre y máquina (cf. Casellas, 2008: 5).

Tanto la representación del conocimiento legal como su adquisición son cruciales en la gestión del conocimiento del dominio legal. Para ello las ontologías legales sirven de enlace entre el hombre y la máquina ya que, a través de los lenguajes de representación se hace posible la recuperación y la gestión de la información.

La Web Semántica ofrece un marco común que permite que los datos sean compartidos y reutilizados entre los límites de la aplicación, la empresa y la comunidad. Se trata de un esfuerzo colaborativo liderado por el W3C<sup>1</sup> que cuenta con la participación de un gran número de investigadores y socios del ámbito de la industria. Se basa en el *Resource Description Framework (RDF)*.

Según Berners-Lee, Hendler y Lassila (2001), la Web Semántica no es una Web separada de la ya existente, sino una extensión de la actual, también conocida como Web 3.0., en la que la información contiene un significado bien definido que favorece que los ordenadores y los seres humanos trabajen en una mayor cooperación. Por tanto, la clave de la Web Semántica reside en que la web pasa de ser comprensible para el ser

---

<sup>1</sup> <http://www.w3.org/2001/sw/>

humano a ser comprensible para la máquina. Así, se superan las limitaciones a la hora de hacer búsquedas, extraer la información, mantenerla, descubrirla y visualizarla (Antoniou y van Harmelen, 2004: 4).

El desarrollo de la Web Semántica se ha realizado en varias etapas, en las que unas capas se han construido sobre otras. La figura 1.1 muestra las capas principales en el diseño de la Web Semántica.

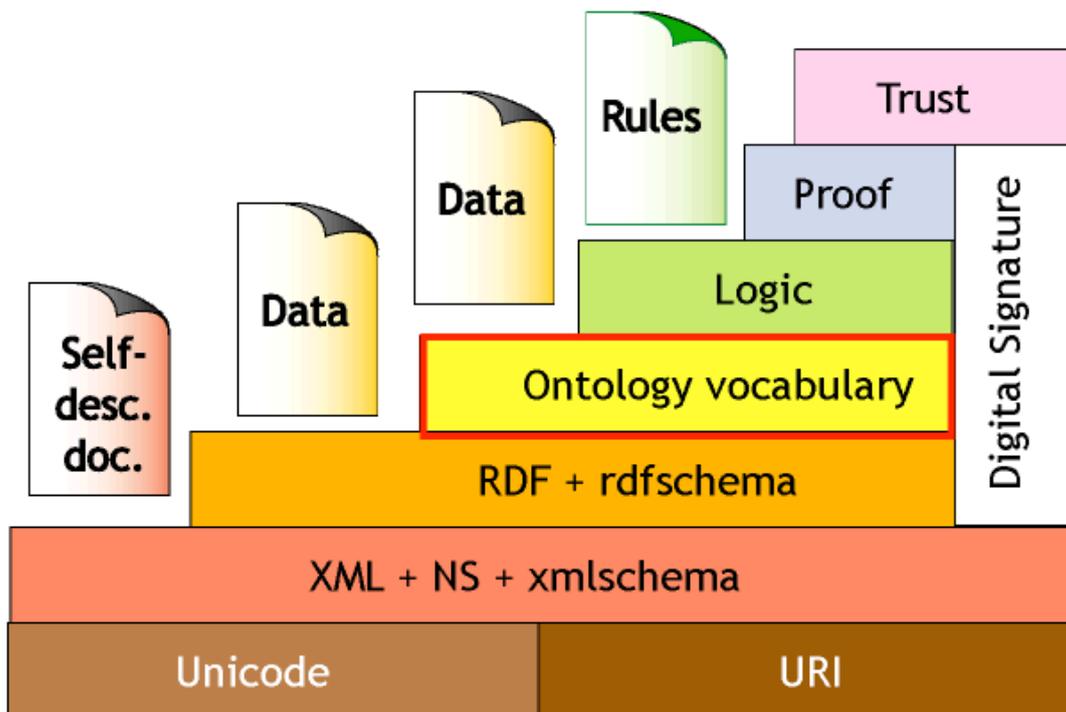


Figura 1.1. La arquitectura de la Web Semántica (charla de Tim Berners-Lee en XML 2000<sup>2</sup>).

En la parte inferior encontramos el lenguaje *XML*, que permite crear documentos web estructurados con un vocabulario definido para el usuario. Este lenguaje es apropiado para enviar documentos a través de la Web. Seguidamente encontramos *RDF*, modelo de datos básico que, como el modelo entidad-relación, permite realizar afirmaciones simples sobre objetos web (fuentes). El modelo de datos *RDF* no utiliza *XML*, pero tiene una sintaxis basada en *XML*, por eso en la figura 1.1 aparece encima de la capa de *XML*. *RDF Schema* ofrece primitivos de modelado para organizar objetos web en

---

<sup>2</sup> <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>

jerarquías. Son primitivos clave las clases y las propiedades, las subclases y las relaciones de subpropiedades y las restricciones de dominio y registro. Como su propio nombre indica *RDF Schema* está basado en *RDF*. Al *RDF Schema* se le puede considerar como lenguaje primitivo para diseñar ontologías, por lo que con el desarrollo de la Web Semántica surge la necesidad de crear lenguajes de ontologías más potentes que amplíen el *RDF Schema* y permitan la representación de relaciones más complejas entre objetos Web. Así, la capa *logic* (“lógica”) se utiliza para mejorar la lengua de la ontología y para permitir formalizar conocimiento específico de aplicaciones declarativas. En cuanto a la capa *proof* (“prueba”), entraña el proceso deductivo real además de la representación de pruebas en lenguajes Web (de los niveles inferiores) y de la validación de pruebas. Por último, la capa *trust* (“confianza”) surge con el uso de firmas digitales y de otro tipo de conocimiento, basado en recomendaciones de agentes fiables o en agencias de calificación o de certificación y entidades de consumo. Como se encuentra en la cima de la pirámide, *trust* es un concepto crucial y de alto nivel: la Web sólo alcanzará su máximo potencial cuando los usuarios tengan total confianza (seguridad) en sus operaciones y en la calidad de la información ofrecida (cf. Antoniou y van Harmelen, 2004: 17-18).

La Web actual se basa en *HTML*, que especifica el modo de diseño de una página web para el lector humano. Así, sucede que el software tiene dificultades en entender el contenido en *HTML* dado que tiene que “adivinar” el significado. Por ejemplo, la figura 1.2 muestra lo que ve el usuario en la página web cuando está buscando información de compra del nuevo iPad, y debajo se muestra cómo aparece parcialmente codificada en *HTML*.



Figura 1.2. El nuevo iPad en la Web actual<sup>3</sup>.

```
<div class="top-hero clearfix" id="ipad-hero-gallery">
  <div class="hero-description">
    <h2></h2>
    <p class="hero tagline">Espectacular pantalla Retina, cámara iSight de
5 Mpx y conexión inalámbrica superveloz. <span class="price">Desde <span
class="no-ws">479,00 €
    </span></span>.</p>
    <p class="hero availability">Pide ahora el nuevo iPad y no pagues
gastos de envío. O cómpralo hoy mismo en tu <a
href="http://www.apple.com/es/retail/ipad">Apple Store</a> más cercano.</p>
    <p class="order shipping">
      <a class="button button rect large more"
href="/es/buy/home/shop_ipad/family/ipad/new_ipad"
onclick="s_objectID='cb4caa60d4ba26f705173a2fcb94f5cd';" data-evar1="AOS:
home/shop_ipad/family/ipad | box-content-1 | AstroButton | select_an_ipad" data-
evar30="home/shop_ipad/family/ipad/box-content-1">
        <span>
          <span class="label">
```

<sup>3</sup> [http://store.apple.com/es/browse/home/shop\\_ipad/family/ipad](http://store.apple.com/es/browse/home/shop_ipad/family/ipad)

```

        Elige un iPad
    </span>
</span>
</a>
    <p>
        <p class="hero"><a href="http://www.apple.com/es/ipad" class="more"
onclick="s_objectID='16560a928d5a742a1ae927e436569203';" data-evar1="AOS:
home/shop_ipad/family/ipad | box-content-1 | Astro Link | box-content-1 | 6" data-
evar30="home/shop_ipad/family/ipad/box-content-1">Visita el sitio web del
iPad</a></p>
        <div class="sharing">
            <div class="questions">
                <p>¿Tienes alguna duda? Pregúntanos.</p>
                <span>Llama al 900 150 503.</span>
            </div>

```

La codificación en *HTML* muestra la complejidad de expresar el significado en este lenguaje. Como señalan Benjamins y otros (2005b: 4), la Web Semántica pretende ofrecer una solución a este problema. Se trata de una extensión de la Web actual en la que la información recibe un significado bien definido, que facilita la interacción entre el hombre y la máquina. Por tanto, la cuestión clave de la Web Semántica radica en cómo dotar de semántica a la Web. Un modo de hacerlo es a través de ontologías, dado que las ontologías representan significado. Hoy en día se utiliza el lenguaje *OWL* (*Ontology Web Language*)<sup>4</sup>. En este lenguaje se describe el significado a través de conceptos y relaciones, cuya semántica se vincula a ontologías disponibles públicamente. Del ejemplo mencionado anteriormente, podríamos imaginarnos algo como lo siguiente (no es un ejemplo realista):

```

<Product> Nuevo iPad</Product>
Desde <Price> <Currency>€</Currency>479 </Price>

```

<sup>4</sup> <http://www.w3.org/2001/sw/WebOnt/>

En el proceso de dotar de semántica a la Web, hacerlo de forma manual es una ardua tarea, con las excepciones de *Wikipedia*<sup>5</sup> y el *Open Directory Project*<sup>6</sup>. Actualmente la Web Semántica está en desarrollo y sólo se pueden ver algunas aplicaciones (cf. Benjamins y otros, 2005b: 5).

En la actualidad, las búsquedas en bases de datos de aplicaciones o en línea se basan en palabras clave, términos que consideramos aparecerán en el documento de búsqueda. Sin embargo, todavía no existe la posibilidad de hacer búsquedas de conocimiento legal en las herramientas de búsqueda de recuperación de información legal, es decir, aún no es posible extraer la información conceptual de las palabras clave de terminología legal por lo que no hay una relación semántica entre las necesidades de información del usuario y la información contenida en los textos, aparte de la coincidencia de patrones en el texto (cf. Peters y otros, 2007: 117).

## **1.2. Las ontologías legales en la representación del conocimiento legal**

Como hemos señalado anteriormente, el uso de ontologías es clave en la implementación de la semántica en la interacción entre el hombre y la máquina. Así, se han construido ontologías diferentes con diferentes propósitos y para dominios de conocimiento distintos: por ejemplo, las ontologías de representación del conocimiento como *Frame Ontology* o *OKBC*; las ontologías de alto nivel como *Cyc* o *SUO*; las ontologías lingüísticas como *EurowordNet* o *GUM* y las ontologías de dominio, como las de comercio electrónico (*e-class*, *RosettaNet*, etc), médicas (*ON9*, *Ontoterm*, etc), o legales (*FoLaw*, *LRI-Core*, etc) entre otras.

Las características propias del dominio legal hacen que su aplicación a la Web Semántica sea de especial interés. Una de las características más significativas del dominio legal es que contiene una gran cantidad de conocimiento del sentido común, por lo que, a diferencia de las ontologías del ámbito de la medicina y la ingeniería, las

---

<sup>5</sup> [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)

<sup>6</sup> <http://www.dmoz.org/>

ontologías del dominio legal han de contener un gran número de conceptos del sentido común que son parte del nivel físico, abstracto, mental o social. Por tanto, el dominio legal no sólo incluye nociones de norma y responsabilidad, sino también incluye conocimiento del mundo y de las actividades sociales.

Benjamins y otros (2005b: 9) proponen el desarrollo de una Web Semántica Legal conformada por el conocimiento de profesionales y usuarios especializados y señalan que, en la construcción de sistemas de información legal, el modelado de conocimiento normativo es crítico en muchas áreas de aplicación típicas de la Web Semántica, tales como el comercio electrónico, la utilización de códigos de práctica, gestión de los derechos de autor, gestión del conocimiento de empresas, etc. De hecho, las ontologías legales existentes hasta el momento versan sobre dichas áreas, como veremos en la sección 2.5.4.5.

### **1.3. Contribución de esta tesis**

Aunque hasta el momento se han realizado muchos esfuerzos en la ingeniería ontológica para la representación del conocimiento legal y la formalización del derecho (como demuestran las ontologías legales existentes presentadas en la sección 2.7), aún queda mucho por hacer para que el conocimiento y la teoría legal se unan por completo a la ingeniería del conocimiento del dominio legal.

Esta tesis se enmarca dentro del dominio de la cooperación internacional en materia penal, ya que describe los procesos de diseño y desarrollo de una ontología satélite del crimen organizado y el terrorismo. Partimos de la hipótesis de que el modelo multinivel de la ontología nuclear de *FunGramKB* se puede exportar a cualquier ontología satélite independientemente de cuál sea su dominio temático, ya sea el derecho, la medicina, etc. En otras palabras, pretendemos demostrar que tanto el modelo teórico de *FunGramKB* como sus herramientas son reutilizables en la elaboración de cualquier ontología satélite de un dominio concreto, como mostraremos en el capítulo 4.

El propósito de esta tesis es, por tanto, describir la metodología para la elaboración de una ontología satélite del crimen organizado y el terrorismo, desde la elaboración de un corpus mediante la recopilación de textos especializados, que posteriormente servirán

para la extracción terminológica, hasta la selección y definición de los términos. En el capítulo 4 explicaremos cada una de las etapas del proceso de adquisición terminológica en el modelado de conocimiento, válido tanto para la elaboración de una ontología satélite del crimen organizado y el terrorismo como para cualquier otra ontología satélite vinculada a la ontología nuclear de *FunGramKB*.

#### **1.4. Estructura de la tesis**

Como hemos señalado anteriormente, el principal propósito de esta tesis es demostrar que tanto la metodología como las herramientas de la base de conocimiento *FunGramKB* son válidas para el desarrollo de ontologías satélite vinculadas a la ontología nuclear. Para ello, presentamos ejemplos obtenidos a partir del Global Crime Term Corpus (GCTC).

Así, la estructura de esta tesis es la siguiente:

Capítulo 2: **Estado de la cuestión.** En este capítulo presentamos el estado de la cuestión de las ontologías. Partimos de la definición de ontología, para describir los tipos de ontologías, criterios de diseño y las ontologías más significativas. A continuación presentamos las metodologías, lenguajes y herramientas de diseño de ontologías. Finalmente analizamos el estado de la cuestión de las ontologías legales.

Capítulo 3: ***FunGramKB*.** En este capítulo hacemos una descripción de la base de conocimiento multipropósito *FunGramKB*, en el que destacamos su estructura multinivel y el enfoque conceptualista de su ontología nuclear, extrapolable a las ontologías satélite vinculadas a la misma.

Capítulo 4: **Desarrollo de una ontología satélite en *FunGramKB* a partir del Global Crime Term Corpus (GCTC).** Aquí describimos el proceso de extracción terminológica, la conceptualización de los términos y hacemos una propuesta de jerarquización cognitiva para la elaboración de una ontología satélite del crimen organizado y el terrorismo.

Capítulo 5: **Conclusiones y sugerencias para posteriores investigaciones.** En este capítulo ofrecemos las conclusiones obtenidas tras el análisis terminológico del Global Crime Term Corpus (GCTC) además de proponer futuras líneas de investigación.

Según las líneas de investigación establecidas en esta tesis, en los apéndices 1 y 2 se incluye un resumen de 20 PS de “conceptos espejo” (+AGGRESSION, +BLACK\_MARKET\_00, +CRIME\_00, \$DIRTY\_MONEY\_00, \$FORGERY\_00, +FREEZE\_00, +LAUNDER\_MONEY\_00, +MASTERMIND\_00, +MONEY\_LAUNDERING\_00, +ORGANIZED\_CRIME\_00, +PHISHING\_00, \$RACKETEERING\_00, \$SAFE\_HAVEN\_00, +SEIZE\_00, \$SMUGGLING\_00, \$SYNDICATE\_00, \$TAX\_EVASION\_00, \$TAX\_HAVEN\_00, +TERRORIST\_00, +TORTURE\_00), y de 10 PS de conceptos especializados del crimen organizado y el terrorismo (\$CAROUSEL\_FRAUD\_00, \$COUNTERFEITING\_00, \$CUCKOO\_SMURFING\_00, \$LAYERING\_00, \$LONE\_SHARK\_00, \$LONE\_WOLF\_00, \$PHARMING\_00, \$SELF\_LAUNDERING\_00, \$VISHING\_00, \$WATERBOARDING\_00) respectivamente.



## Capítulo 2

# ESTADO DE LA CUESTIÓN

### 2.1. Introducción

El uso de ontologías ha sido clave en los últimos años ya que la semántica que utiliza favorece la comunicación entre el hombre y la máquina. Aunque en un principio las ontologías se concebían para compartir conocimiento y para su posterior reutilización, hoy en día son claves para implementar la web semántica. Como consecuencia, la definición de ontología, su propósito y características principales han evolucionado desde que surgieron las primeras ontologías en la década de los noventa hasta ahora.

El propósito de este capítulo es ofrecer una introducción a las ontologías legales y al modelado ontológico. Para ello se presenta primero la problemática suscitada por la definición de ontología en la literatura sobre Inteligencia Artificial (IA). Después se ofrece una visión amplia de los distintos tipos de ontologías, las metodologías, lenguajes y herramientas utilizados. Finalmente se establece una clasificación de los distintos tipos de ontologías legales existentes hasta el momento en el ámbito de la inteligencia artificial y el derecho.

### 2.2. Definición de Ontología

El término “ontología” se ha tomado prestado de la filosofía, como disciplina sobre el ser o la existencia, en el sentido Aristotélico del término (Gómez y otros, 2008: 3). Como señala Guarino (1998: 4) hay que distinguir entre la “Ontología” (con mayúscula) y “ontología” (con minúscula), donde la primera se refiere al significado filosófico del término y la segunda a su significado para la IA y la Lingüística Computacional. Así, para la filosofía, “Ontología” es la ciencia de lo que es, de las clases y estructuras de objetos, propiedades, eventos, procesos y relaciones en cada una de las áreas de la realidad (Smith, 2003: 155). Para un filósofo, una ontología es, por tanto, un sistema particular de categorías sistematizando cierta visión del mundo (Guarino, 1998: 4). Para Guarino y otros (2009: 1-2) la “Ontología” trata del estudio de los atributos de las cosas por su propia naturaleza. Por tanto, este sistema no es dependiente de una lengua en

concreto. Por ejemplo, la ontología de Aristóteles es siempre la misma independientemente de la lengua que se utilice para describirla. Por otro lado, en Inteligencia Artificial se utiliza el término “ontología” para referirse a un artefacto de ingeniería constituido por un vocabulario específico utilizado para describir una realidad concreta, además de un conjunto de supuestos explícitos sobre el significado del vocabulario recogido en la ontología. Este conjunto de supuestos tiene normalmente la forma de una teoría de lógica de primer orden (en este caso la ontología puede recibir el nombre de “ontología formal”), donde las palabras aparecen como conceptos o relaciones. De acuerdo a Gruber (1993: 199; 1995: 908), para los sistemas de Inteligencia Artificial, lo que existe es lo que puede ser representado.

En cuanto a lo que el sistema de representación se refiere, no hay que confundir Ontología y Epistemología, ya que la Epistemología versa sobre el saber y el conocimiento, si bien no se pueden separar, ya que la epistemología se ocupa de los procesos por los que se consigue el conocimiento, y la producción del conocimiento está inextricablemente ligada a los compromisos ontológicos. Para Gruber (1995: 909) los compromisos ontológicos son acuerdos para utilizar el vocabulario definido en la ontología en las afirmaciones y preguntas. Según Boer (2009: 32), conocer es haber adquirido la creencia de algo a través de un proceso por el que se consigue el conocimiento.

Una de las primeras definiciones de “ontología” fue la de Neches y otros (1991: 40): “una ontología define los términos básicos y las relaciones entre el vocabulario de un área determinada además de las reglas de combinación de términos y sus relaciones para definir las extensiones del vocabulario”. Años más tarde, (Gruber 1993: 199) definió “ontología” como la “especificación explícita de una conceptualización”. Esta definición ha sido la más citada por la comunidad científica. Se han propuesto muchas definiciones de “ontología” a partir de la definición de Gruber. Por ejemplo Borst (1997: 12) completa la definición de Gruber por la que define “ontología” como “una especificación formal de una conceptualización compartida”. Así, la definición de Borst destaca el hecho de que debe haber acuerdo en cuanto a la conceptualización que se hace explícita. Más adelante Studer y otros (1998: 185) combinan las definiciones de Gruber y Borst de este modo:

*Una ontología es una especificación formal, explícita de una conceptualización compartida. Conceptualización se refiere a un modelo abstracto de algún fenómeno del mundo identificado por los conceptos más relevantes de tal fenómeno. Explícita significa que el tipo de conceptos utilizados y las limitaciones en su uso se definen de forma explícita. Por ejemplo, en la medicina, los conceptos son las enfermedades y los síntomas, las relaciones entre ambos son causales y una limitación es que la enfermedad no es causada por sí misma. Formal se refiere al hecho de que la ontología debe ser susceptible de ser interpretada por la máquina, lo que excluye el lenguaje natural. Compartida refleja la noción de una ontología que captura el conocimiento consensuado, esto es, no es conocimiento privado de un individuo, sino aceptado por un grupo.*

En el presente trabajo adoptamos esta definición de “ontología” porque la consideramos la más completa, ya que señala que la “ontología” refleja el conocimiento del mundo de un modo interpretable por la máquina y consensuado por un grupo.

Por su parte, Guarino y Giaretta (1995: 25) recogieron y analizaron las siguientes siete definiciones de “ontología”:

1. *Ontología como disciplina filosófica.*
2. *Ontología como un sistema conceptual informal.*
3. *Ontología como una entidad semántica formal.*
4. *Ontología como una especificación de una “conceptualización”.*
5. *Ontología como una representación de un sistema conceptual por vía de una teoría lógica.*
  - 5.1. *caracterizada por propiedades formales específicas.*
  - 5.2. *caracterizada sólo por su propósito específico.*
6. *Ontología como el vocabulario utilizado por una teoría lógica.*
7. *Ontología como una especificación (meta-nivel) de una teoría lógica.*

En el citado trabajo, Guarino y Giaretta proponen la siguiente definición de “ontología” (1995: 32): “una teoría lógica que aporta una explicación parcial y explícita de una conceptualización”. Por tanto, para estos autores “conceptualización” es básicamente la

idea del mundo que tiene una persona o grupo. Aunque aparentemente esta noción de “conceptualización” es muy similar a la de Studer y otros (1998), podría decirse que Guarino y Giaretta van un paso más allá al formalizar la noción de “conceptualización” y establecer que una ontología ha de ser construida a través de una teoría lógica. Por tanto, estrictamente hablando, esta definición sólo sería aplicable a las ontologías desarrolladas en el marco de la lógica. Más adelante, Guarino (1998: 4) perfeccionó la definición de “ontología” presentada por Guarino y Giaretta (1995) al definirla como “un conjunto de axiomas lógicos diseñados para dar cuenta del significado determinado de un vocabulario concreto”.

Por otro lado encontramos un grupo de definiciones clasificadas de acuerdo al proceso seguido en la construcción de ontologías. Estas definiciones incluyen incluso lo más destacado de las relaciones entre ontologías y bases de conocimiento, como por ejemplo la definición de Bernaras y otros (1996: 298) en el marco del proyecto *KACTUS* (Schreiber y otros, 1995): “(una ontología) ofrece el medio para describir de forma explícita la conceptualización que hay tras el conocimiento representado en la base de conocimiento”. Lo más significativo de esta definición es que sugiere extraer la “ontología” de una base de conocimiento dada. Así, la “ontología” se construye siguiendo una metodología de abajo a arriba mediante un proceso de abstracción.

Otra estrategia para construir ontologías es la reutilización de ontologías muy amplias, como *SENSUS* (Swartout y otros, 1997), que tiene más de 50.000 conceptos, para crear ontologías de un dominio específico y bases de conocimiento: “una ontología es un conjunto de términos jerárquicamente estructurados para describir un dominio que se puede utilizar como almacén en el que se asienta la base de conocimiento”. De acuerdo a esta definición, se podría utilizar la misma ontología para construir varias bases de conocimiento que compartirían la misma taxonomía. Así, sería posible ampliar el almacén de la ontología a un nivel inferior al añadir subconceptos específicos del dominio, o a un nivel superior al añadir conceptos de nivel intermedio o alto para cubrir nuevas áreas. Si se construyen sistemas con la misma ontología, éstos compartirán la misma estructura, con lo que se facilitaría enormemente que las bases de conocimiento se fusionaran y pudieran compartir sus mecanismos de inferencia (cf. Gómez y otros 2004: 7-8). De hecho, como mostraremos en el capítulo 4, la ontología nuclear de

*FunGramKB* se reutiliza en la creación de las ontologías satélite de dominios específicos ya que mediante los mecanismos de herencia e inferencia.

Como señalan Studer y otros, puede suceder que meras taxonomías se consideren como ontologías completas (1998: 185-186). Como señalan Lassila y McGuinness (2001: 5-6), *UNSPSC*<sup>7</sup>, *e-cl@ss*<sup>8</sup>, y *RosettaNet*<sup>9</sup>, propuestas de estándar del comercio electrónico, y el directorio de *Yahoo!*, taxonomía de búsqueda en la Web, pueden considerarse como ontologías porque ofrecen una conceptualización consensuada de un dominio concreto. Dentro de la comunidad ontológica se distinguen ontologías que son meras taxonomías de las ontologías que modelan un dominio de un modo profundo y ofrecen más restricciones en la semántica del dominio. A estas ontologías se las llama ontologías superficiales y ontologías profundas respectivamente. Por un lado, las ontologías superficiales incluyen conceptos, taxonomías de conceptos, relaciones entre los conceptos y las propiedades que describen los conceptos. Por otro lado, las ontologías profundas añaden axiomas y limitaciones a las ontologías superficiales. Los axiomas y las limitaciones clarifican por tanto el significado específico de los términos recogidos en la ontología (Gómez y otros 2004: 8). De hecho, en los casos más sencillos, la ontología describe una jerarquía de conceptos relacionados a través de una relación de subsunción; mientras que en casos más complejos se añaden los axiomas apropiados para expresar otro tipo de relaciones entre conceptos y para precisar su interpretación. Así, en Inteligencia Artificial (IA) dos ontologías pueden utilizar vocabulario de lenguas distintas (por ejemplo, inglés y español) pero compartir la misma conceptualización (Guarino, 1998: 7). Efectivamente este es el caso de la ontología de *FunGramKB*, que presentamos en el capítulo 3.

Gruber (1993: 208-210) propuso modelar las ontologías profundas con marcos y lógica de primer orden. Para ello identificó cinco tipos de componentes: clases, relaciones, funciones, axiomas formales e instancias. A continuación presentamos estos componentes con diversos ejemplos del dominio de la medicina.

---

<sup>7</sup> <http://www.unspsc.org>

<sup>8</sup> <http://www.eclass.org>

<sup>9</sup> <http://www.rosettanel.org>

Las **clases** representan conceptos, tomados en su sentido más amplio. En el dominio de la medicina, por ejemplo, los conceptos son las enfermedades (alergia, gripe, sarampión, etc.), los síntomas (estornudo, congestión, dolor corporal, etc.) y las medicinas para su tratamiento (vacuna, analgésico, antibiótico, etc.). En la ontología, las clases están normalmente organizadas en taxonomías a las que se les puede aplicar mecanismos de herencia. Así, podemos representar por ejemplo una taxonomía de enfermedades (gripe, sarampión, varicela, etc.). Por otro lado, también es posible definir metaclases en un paradigma de representación del conocimiento basado en marcos. Las metaclases son clases cuyas instancias son a su vez clases que permiten diferenciar las diferentes capas de clases incluidas en la ontología (Gómez y otros 2008: 173).

Las **relaciones** representan un tipo de asociación entre los conceptos del dominio. Se definen formalmente como cualquier subconjunto de un producto de  $n$  conjuntos, por ejemplo:  $R \subset C_1 \times C_2 \times \dots \times C_n$ . Las ontologías normalmente contienen relaciones binarias, cuyo primer argumento es el dominio de la relación, y el segundo es el rango. Por ejemplo, la relación binaria “subclase de” se utiliza para construir la taxonomía de clases. Ejemplos de clasificaciones son: una “quemadura de tercer grado” sería una subclase de “quemadura” y “quemadura” sería una subclase de “herida”. Las relaciones binarias pueden instanciarse con información del dominio. Por ejemplo, para expresar que la gripe produce malestar corporal, estornudos tenemos que escribir (gripe síntomas malestar corporal). Las relaciones binarias se pueden utilizar para expresar atributos de conceptos conocidos como ranuras (*slots*), que se distinguen de las relaciones porque su rango es un tipo de datos como por ejemplo cadena de caracteres, número, etc., mientras que el rango de relaciones es un concepto.

Las **funciones** son un caso especial de relaciones en las que el elemento  $n$ -th de la relación es único para los elementos  $n-1$  precedentes. Esto se expresa normalmente como:  $F: C_1 \times C_2 \times \dots \times C_{n-1} \rightarrow C_n$ . [No hemos encontrado ejemplos de función en el ámbito de la medicina. Gómez y otros (2004: 13) ofrecen como ejemplo de función en el dominio de los viajes “paga”, que obtiene el precio de una habitación después de aplicar un descuento].

Los **axiomas formales** sirven para modelar afirmaciones que son siempre ciertas. Normalmente se utilizan para representar conocimiento que no puede ser formalmente definido por otros componentes. Además, los axiomas formales se utilizan para comprobar la consistencia de la ontología misma o del conocimiento almacenado en una base de conocimiento. Los axiomas formales son muy útiles para inferir conocimiento nuevo. Un axioma del dominio de la medicina sería que no es posible confirmar la paternidad de un hijo sin una prueba de ADN.

Por último, Gruber (1993: 210) señala que las **instancias** se utilizan para representar elementos o individuos en una ontología. Un ejemplo de instancia del concepto “alergia al polen” sería la “alergia al polen estacional”.

Es bien sabido que en la comunidad ontológica las ontologías se utilizan con diferentes propósitos (procesamiento del lenguaje natural, gestión del conocimiento, comercio electrónico, la Web Semántica, etc.) y en comunidades diferentes (ingeniería del conocimiento, bases de datos, etc.). Así, Uschold y Jasper (1999: 11-12) ofrecieron una nueva definición de “ontología” aplicable a otras disciplinas:

*Una ontología puede adoptar diversas formas, pero necesariamente incluirá un vocabulario de términos y alguna especificación de su significado. Ésta incluye definiciones e indicaciones de cómo están interrelacionados los conceptos, los cuales de forma colectiva imponen una estructura en el dominio y limitan las posibles interpretaciones de los términos.*

Gruber (2008) ofrece una definición de ontología como término técnico para la informática que relaciona su sentido filosófico con la IA. Para él una “ontología” está diseñada por un propósito, que es permitir modelar el conocimiento sobre un dominio concreto, real o imaginario. Desde un punto de vista pragmático, una ontología es una herramienta y producto de la ingeniería del conocimiento y por tanto está definida por su uso. Mientras que la ontología debe estar formulada en algún tipo de lenguaje de representación con un nivel de especificación semántica, es independiente de la estrategia de modelado de datos o de implementación. Las ontologías son parte de los estándares de la Web Semántica del W3C, que se utilizan para especificar vocabularios conceptuales estándares en los que el intercambio de información entre sistemas,

proveen servicios para responder a cuestiones, publicar bases de conocimiento reutilizables, y ofrecen servicios para facilitar la interoperabilidad entre sistemas múltiples, heterogéneos y bases de datos. El papel clave de las ontologías con respecto a los sistemas de bases de datos es especificar los datos de modelado de representación en un nivel de abstracción sobre diseños específicos de bases de datos (lógicos o físicos), por lo que los datos se pueden exportar, traducir, cuestionar y unificar a lo largo de sistemas y servicios independientemente desarrollados.

En informática e IA una ontología es un modelo abstracto consensuado y reutilizable de los conceptos que reflejan una percepción del mundo, formalizable a través de un lenguaje que puede ser procesado por el ordenador (Gruber, 2006; Studer y otros, 1998). Como para la IA lo que existe es lo que puede ser representado, una ontología será, por tanto, una representación de una realidad determinada. Así cada ontología aporta una percepción del mundo diferente, puesto que refleja la visión del mundo particular de los ingenieros del conocimiento encargados de su desarrollo, y, en el caso de las ontologías legales, la visión adoptada por los ingenieros del conocimiento legal caracterizará la ontología que modelen.

### **2.3. Tipos de ontologías y criterios de diseño**

Es difícil clasificar las ontologías porque varían en su concepción. Las hay que son ontologías propiamente dichas, meras taxonomías, representaciones de conocimiento y bases de conocimiento, o incluso una combinación de las dos cosas. En esta sección ofrecemos una síntesis de la clasificación de Gómez y otros (2004: 26-36) y la clasificación presentada por Casellas (2008: 26-45) para elaborar la nuestra.

Inicialmente, Gruber (1993) hizo una primera distinción entre ontologías de representación y de contenido. Entendía que las ontologías de representación ofrecían un marco, pero no una guía de cómo representar el conocimiento. Ésa era la labor de las ontologías de contenido, que establecían cómo el mundo (o una conceptualización del mismo) tenía que ser descrito. Por tanto, concebía las ontologías de contenido como

globales (por ejemplo la ontología CyC), como una “percha conceptual” en la que poder colgar ontologías más específicas y el conocimiento específico de un dominio.

Por su parte, Wielinga y Schreiber (1993: 110-112) distinguen ontologías de acuerdo a su nivel de complejidad y compromiso ontológico. Establecieron cinco tipos de ontologías, del nivel más bajo al más alto:

1. Terminología de aplicación en un dominio. Este tipo de ontología es de menor nivel de complejidad y define el vocabulario de una base de conocimiento.
2. Ontología de dominio. Representa el vocabulario específico de un dominio junto a un conjunto de tipos y relaciones básicas.
3. Ontología orientada al modelo de un dominio. Es una ontología consensuada de un dominio que representa teorías generales de un dominio.
4. Ontología orientada a un tipo de tarea. Es un tipo de ontología que comprende una cierta clase de tareas.
5. Ontología de un método específico, a semejanza del tipo anterior, comprende la aplicación de un método particular de resolución de problemas.

Además, estos autores distinguieron otro tipo de ontologías, las “meta-ontologías”, similares a las ontologías de representación de Gruber (1993), como el conjunto de tipos fijos y de las relaciones básicas con significado estándar – el utilizado en la ontología de dominio (figura 2.1.).

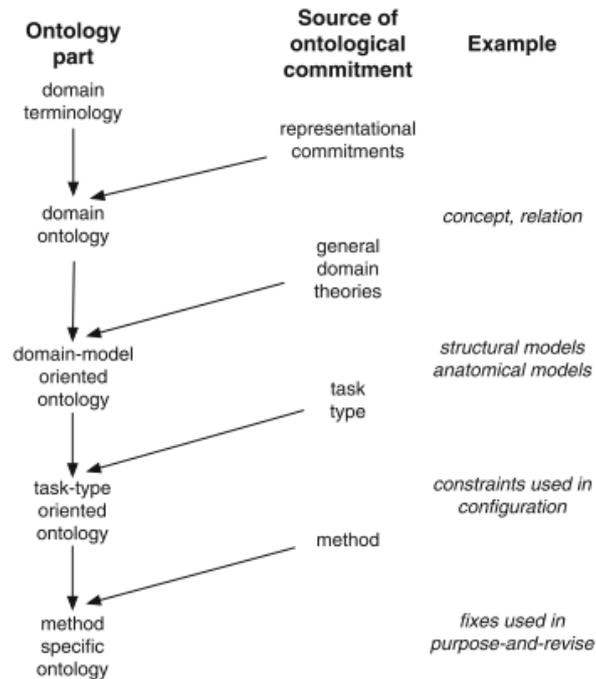


Figura 2.1. Ontologías y compromisos ontológicos. Wielinga y Schreiber (1993: 112).

Más adelante, Mizoguchi y otros (1995: 46) establecieron una tipología de ontologías basada en la idea de que “desde la tecnología de las bases de conocimiento, el conocimiento tiene que ser considerado en algún contexto, esto es, en un contexto de resolución de problemas”. Para estos autores, una base de conocimiento es siempre específica para una aplicación. De hecho, entienden “ontología” desde una perspectiva de base de conocimiento como “sistema de conceptos/vocabulario utilizados como primitivos semánticos para construir sistemas artificiales”, porque estudiaban el conocimiento como forma de resolución de problemas en lugar de conocimiento en general. Desde esta perspectiva clasifican las ontologías en:

1. Ontologías de contenido para reutilización del conocimiento. Estas ontologías incluyen a su vez cinco subcategorías: ontologías de tareas, ontologías de dominio y ontologías generales o comunes.
2. Ontologías de comunicación (*tell & ask*) para compartir conocimiento.
3. Ontologías de indización para recuperación de textos.
4. Meta-ontologías, equivalentes a las ontologías de representación de Gruber (1993).

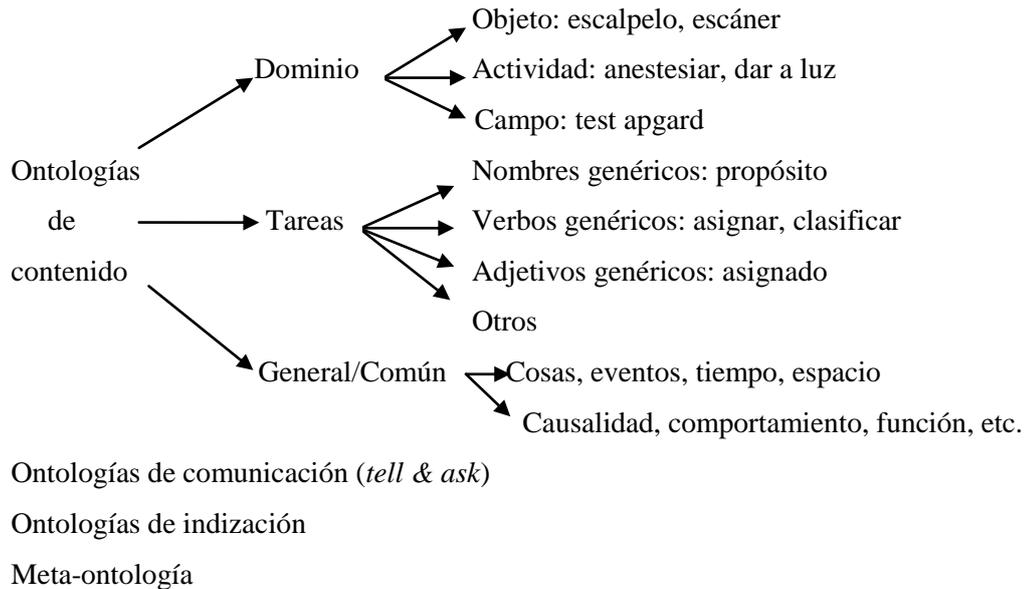


Figura 2.2. Adaptación de la categorización de Mizoguchi y otros (1995: 52)

Valente y Breuker (1996) entendían que diferentes propósitos de uso y de reutilización de las ontologías requerían diferentes requisitos en la ontología. Así, enumeraron las distintas funciones que las ontologías podían realizar:

- a. Ser utilizadas como repositorios para organizar el conocimiento y la información (conocimiento corporativo, terminologías estandarizadas para comunidades profesionales, etc).
- b. Ser utilizadas hacia la adquisición del conocimiento.
- c. Ofrecer documentación (justificación) en relación a los compromisos explícitos de la base de conocimiento.
- d. Facilitar la reutilización del conocimiento en la construcción de nuevas aplicaciones.
- e. Ser utilizadas como base de lenguaje de representación del conocimiento.

Por su parte Van Heijst y otros (1997: 192-193) clasificaron las ontologías de acuerdo a dos dimensiones: la cantidad y tipo de la estructura de la conceptualización, y el tema de la conceptualización. Con respecto a la primera dimensión podemos distinguir tres categorías:

1. Las ontologías terminológicas. Son como los lexicones que especifican los términos utilizados para representar el conocimiento de un dominio determinado. Un ejemplo de este tipo de ontología en el ámbito de la

medicina es la red semántica en *UMLS (Unified Medical Language System; Lindberg y otros, 1993)*.

2. Las ontologías de información, que especifican la estructura de registro de las bases de datos. Esquemas de bases de datos son ejemplos de este tipo de ontologías, como el nivel 1 del modelo PEN y PAD (Rector y otros, 1993), un marco de modelado de registros médicos de pacientes. Éste es un ejemplo típico de esta clase de ontologías en el dominio de la medicina. En este nivel el modelo ofrece un marco para registrar observaciones básicas de los pacientes, pero no hace distinción entre síntomas, signos, tratamientos, etc.
3. Las ontologías que modelan el conocimiento. Especifican conceptualizaciones del conocimiento. Comparadas con las ontologías de información, las ontologías que modelan el conocimiento normalmente tienen una estructura interna más rica. Además, estas ontologías a menudo se ajustan a un uso particular del conocimiento que describen. Dentro del contexto del desarrollo de los sistemas de bases de conocimiento, las ontologías que modelan el conocimiento son las más interesantes. El nivel 2 de descripción del modelo PEN y PAD es un ejemplo de este tipo de ontologías en el dominio de la medicina. En este nivel, las observaciones del nivel 1 se agrupan para describir el proceso de toma de decisiones.

La otra dimensión en la que se pueden clasificar las ontologías es el tema de la conceptualización. En esta dimensión se pueden diferenciar cuatro categorías:

1. Ontologías de aplicación. Contienen todas las definiciones necesarias para modelar el conocimiento requerido por una aplicación particular. Típicamente este tipo de ontologías son una mezcla de conceptos tomados de las ontologías de dominio y de las ontologías generales (que se describen más abajo). Además, este tipo de ontologías pueden contener extensiones específicas de método y de tareas. No son reutilizables por sí mismas. Se pueden obtener al seleccionar teorías de una librería ontológica, que se ajustan a una aplicación particular. Van Heijst y otros (1997) utilizan el término ontología de aplicación de modo similar al que hacen Tu y otros (1995) en *Protégé-II*.

2. Ontologías de dominio. Expresan conceptualizaciones específicas de un dominio particular. Las metodologías de ingeniería del conocimiento actuales hacen una clara distinción entre ontologías de dominio y conocimiento del dominio. Mientras que el conocimiento del dominio describe situaciones objetivas en un cierto dominio, las ontologías de dominio ponen límites a la estructura y contenidos del conocimiento de tal dominio.
3. Ontologías generales. Son similares a las ontologías de dominio, pero los conceptos que definen se consideran generales en varias áreas. Típicamente, las ontologías generales definen conceptos como estado, evento, proceso, acción, componente, etc. Los conceptos de las ontologías de dominio a menudo se definen como especializaciones de los conceptos de las ontologías generales. Ocurre que el límite entre las ontologías generales y las ontologías de dominio es vago, pero la distinción puede ser intuitivamente significativa y útil para construir librerías de ontologías.
4. Ontologías de representación. Explican las conceptualizaciones que subyacen bajo los formalismos de representación del conocimiento (Davis y otros, 1993). Son concebidas para ser neutrales con respecto a las entidades del mundo (Guarino y Boldrin, 1993). Esto es, ofrecen un marco representacional sin hacer del mundo. Las ontologías de dominio y las ontologías generales se describen utilizando los primitivos semánticos que proveen las ontologías de representación. Un ejemplo de este tipo de ontología es la ontología de marcos (*Frame Ontology*) que es la que se utiliza en *Ontolingua* (Gruber, 1992a).

De acuerdo a esta clasificación obtenemos el diagrama ofrecido por Gómez y otros (2004: 27) y que aparece en la figura 2.3.

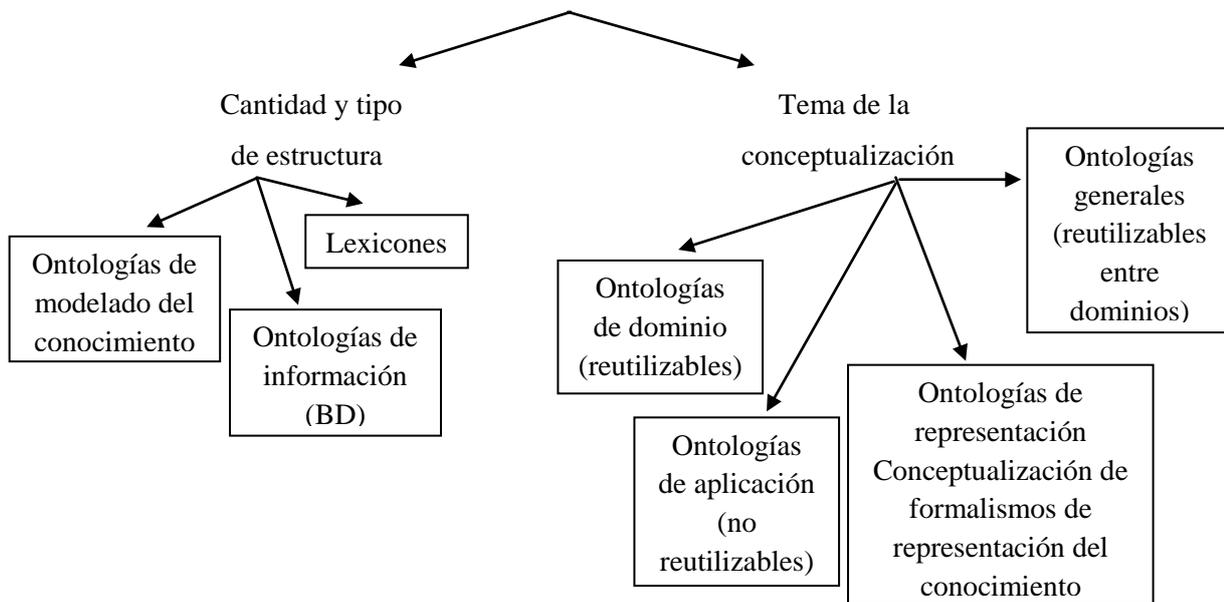


Figura 2.3. La representación de la categorización de ontologías (adaptación de Van Heijst y otros, 1997).

Más adelante, Guarino (1998) clasificó los tipos de ontologías de acuerdo a su nivel de dependencia en una tarea particular o punto de vista. Guarino distinguió ontologías de alto nivel, de dominio, de tarea y de aplicación, como se puede ver en la figura 2.4.

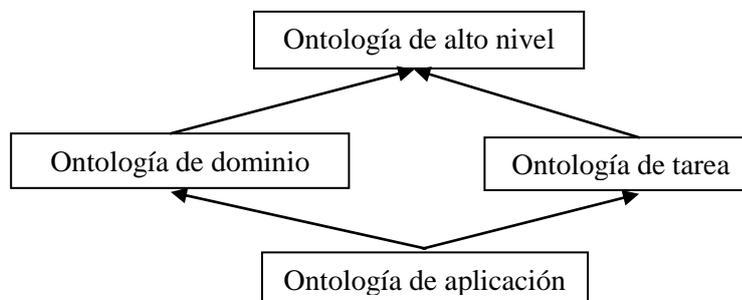


Figura 2.4. Tipos de ontologías, de acuerdo a su nivel de dependencia en una tarea particular o punto de vista. Las flechas representan las relaciones de especialización (adaptación de Guarino, 1998: 9).

Así, para Guarino (1998: 9-10), las ontologías de alto nivel describen varios conceptos muy generales como el espacio, tiempo, materia, objeto, evento, acción, etc., que son independientes de un problema particular o dominio. Por tanto parece razonable, al menos en teoría, tener ontologías de alto nivel unificadas para grandes comunidades de usuarios. Las ontologías de dominio y las ontologías de tareas describen el vocabulario relacionado con un dominio (p. ej. la medicina) o con una tarea o actividad (p. ej. el

diagnóstico) respectivamente, mediante los términos especializados introducidos en la ontología de alto nivel.

Por su parte, las ontologías de aplicación describen los conceptos dependientes tanto de un dominio como de una tarea particular, que a menudo son especializaciones de las ontologías relacionadas. Estos conceptos suelen corresponderse con los papeles realizados por las entidades de dominio mientras que desarrollan una actividad determinada, como unidad reemplazable o componente extra. Guarino (1998) también señala la diferenciación entre las ontologías de aplicación y las bases de conocimiento. El propósito de la ontología es una base de conocimiento determinada que describe los hechos que se presumen siempre ciertos para una comunidad de usuarios, en virtud del significado arbitrario del vocabulario utilizado. Sin embargo, una base de conocimiento genérica puede también describir hechos y afirmaciones relacionadas con un estado de las cosas particular o con un estado epistémico. En una base de conocimiento genérica, podemos distinguir, por tanto, dos componentes: la ontología (que contiene información de estado independiente) y el “núcleo” de la base de conocimiento (que contiene información de estado dependiente).

Finalmente, Lassila y McGuiness (2001) clasificaron las ontologías de acuerdo a la información que contiene la ontología y a la riqueza de su estructura interna. Señalan las siguientes categorías: vocabularios controlados, glosarios, tesauros, jerarquías IS-A informales, jerarquías IS-A formales, instancias formales, marcos, restricciones de valor y límites lógicos generales, como se aprecia en la figura 2.5.

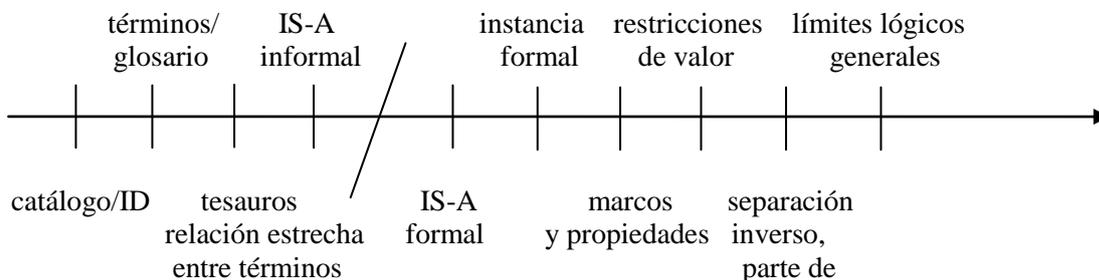


Figura 2.5. El espectro de la ontología (adaptado de Lassila y McGuiness, 2001).

Estos autores sólo consideran ontologías las que contienen más que jerarquías IS-A formales. Por tanto, para ellos la formalidad es requisito sine qua non para la construcción de ontologías. Además, en relación a su estructura interna, consideran las ontologías que contienen al menos jerarquías IS-A como ontologías simples y a las ontologías que incluyen formalismos, marcos, restricciones y límites las llaman ontologías estructuradas (McGuinness, 2003). Según esta clasificación una representación posible sería la siguiente (cf. Casellas 2008: 38):

- **Ontologías simples**
  1. Que incluyen jerarquías formales IS-A.
  2. Que incluyen 1 y relaciones formales de instancias.
  3. Que incluyen 2 y marcos (propiedad de la información).
  4. Que incluyen 3 y restricciones de valor.
- **Ontologías estructuradas**
  5. Que incluyen 4 y una especificación de relaciones de separación de clases, inversas y parte de un todo.
  6. Que incluyen 5 y restricciones de lógica generales.

A continuación mostramos la clasificación de Gómez y otros (2004) que combina las distintas aproximaciones a las ontologías mencionadas más arriba de acuerdo a la riqueza de su estructura interna y al tema central de la conceptualización. Así, el primer grupo se describe de acuerdo a la clasificación de Lassila y McGuinness (2001) en la figura 2.5 y el segundo amplía la clasificación de Van Heijst y otros (1997) mostrada en la figura 2.3, junto a otro tipo de ontologías ya identificadas en la literatura de la IA y el Procesamiento del Lenguaje Natural (PLN) tales como las ontologías de más alto nivel y las ontologías de tareas, dominio-tarea, y método.

### 2.3.1. Tipos de ontologías basadas en la riqueza de su estructura interna

Lassila y McGuinness (2001) clasificaron diferentes tipos de ontologías de semántica superficial y profunda en una línea continua, como se presenta en la figura 2.5. Las principales categorías y sus significados son:

- Vocabularios controlados, por ejemplo una lista cerrada de términos. Un ejemplo típico de esta categoría sería un catálogo.
- Glosarios, esto es, una lista de términos con sus significados especificados en afirmaciones en lenguaje natural.
- Tesauros, que ofrecen semántica adicional entre términos. Aportan información como relaciones de sinonimia, pero no presentan una jerarquía explícita. Por ejemplo médico y doctor pueden considerarse sinónimos en el ámbito de la medicina.
- Jerarquías IS-A informales, tomadas de las especificaciones de las jerarquías de términos como *Yahoo!* Tal jerarquía no es estrictamente una subclase o jerarquía IS-A. Por ejemplo, los términos tratamiento y síntoma no son tipos de enfermedades, pero podrían modelarse en jerarquías IS-A informales bajo el concepto enfermedad porque están íntimamente relacionados con la enfermedad.
- Jerarquías IS-A formales. En estos sistemas, si B es una subclase de A y un objeto es una instancia de B, entonces el objeto es una instancia de A. Las jerarquías de subclases estrictas son necesarias para explotar los mecanismos de herencia. Siguiendo con los ejemplos del dominio de la medicina, pediatra y dentista serían subclases del concepto médico.
- Jerarquías IS-A formales que incluyen instancias del dominio. En este caso incluiríamos por ejemplo instancias de enfermedades (síntomas, tratamiento,...etc.).
- Marcos. La ontología incluye clases y sus propiedades, que las clases de niveles inferiores de la taxonomía IS-A pueden heredar.
- Ontologías que expresan restricciones de valor. Estas ontologías pueden ofrecer restricciones a los valores de una propiedad. Por ejemplo, el tipo de propiedad estornudo es un síntoma.
- Ontologías que expresan restricciones de lógica generales. Estas son las más expresivas. Los ingenieros del conocimiento pueden especificar restricciones de lógica de primer orden entre términos utilizando

lenguajes de representación (sección 2.6.3.). Una restricción lógica sería que no es posible confirmar la paternidad de un hijo sin una prueba de ADN.

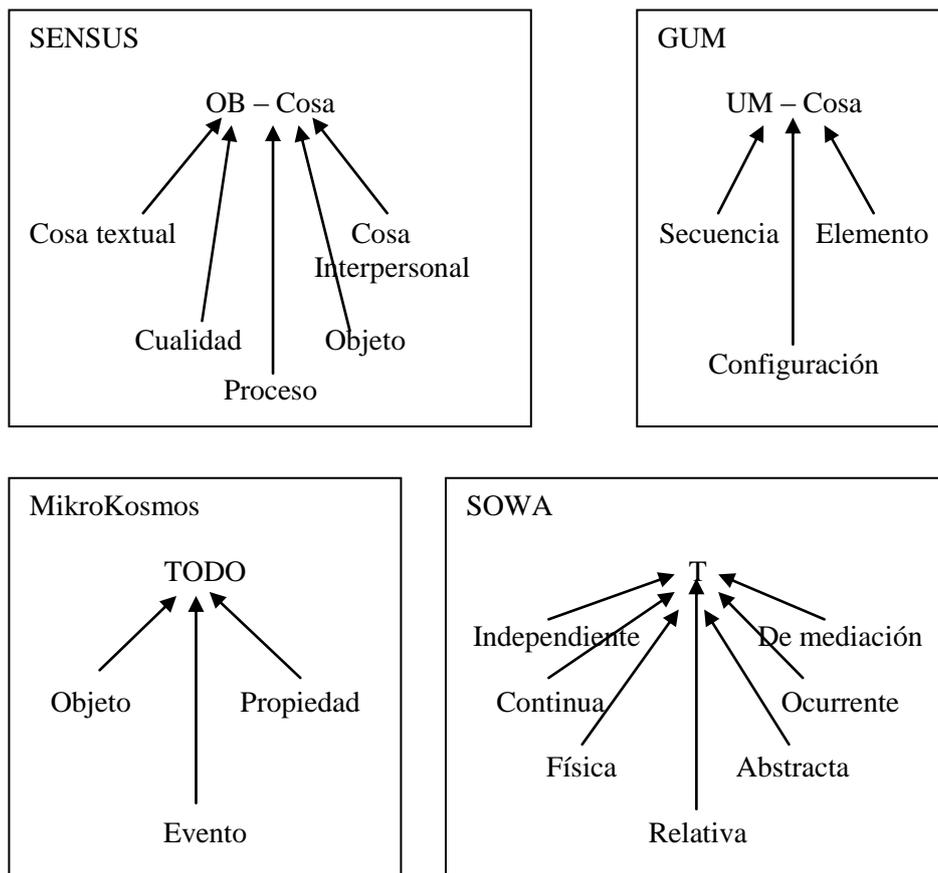
### **2.3.2. Tipos de ontologías basadas en el tema de la conceptualización**

De acuerdo al tema de la conceptualización, podemos clasificar las ontologías en los siguientes grupos:

**Ontologías de representación del conocimiento** (van Heijst y otros, 1997) reflejan los primitivos de representación utilizados para formalizar el conocimiento en un paradigma de representación del conocimiento dado. Como ejemplos más representativos están la ontología de marcos *Frame Ontology* (Gruber 1993) y la ontología *OKBC*, ambas disponibles en el servidor *Ontolingua*. Ofrecen definiciones formales de primitivos de representación utilizados principalmente en lenguajes basados en marcos (por ejemplo, clases, subclases, atributos, valores, relaciones y axiomas). Permiten construir otras ontologías por medio de convenciones basadas en marcos. Ambas ontologías ofrecen definiciones con expresiones lógicas escritas en lenguaje de representación *KIF*. Además de la *Frame Ontology* y la *OKBC Ontology* hay otras ontologías de representación del conocimiento como *RDF*, *RDF Schema*, *OIL*, *DAML+OIL* y *OWL*, que analizamos más abajo.

Las **ontologías generales** (van Heijst y otros, 1997) u **ontologías comunes** (Mizoguchi y otros, 1995) se utilizan para representar conocimiento del sentido común reutilizable en los distintos dominios. Estas ontologías incluyen vocabulario relacionado con las cosas, eventos, tiempo, espacio, causalidad, comportamiento, función, mereología, etc. La *Mereology Ontology* (Borst, 1997) es uno de los ejemplos más clásicos de ontología general. Define la relación “parte-de” y sus propiedades. Con la relación “parte-de” podemos afirmar que los elementos están formados por componentes, cada uno de los cuales también se puede descomponer en subcomponentes. Esta ontología define las propiedades que ha de tener cualquier descomposición. Otro ejemplo de ontología general es la *Standard-Units Ontology*, que incluye definiciones sobre sus unidades de medida (por ejemplo, minuto y segundo-de-tiempo).

Las **ontologías de alto nivel** o de **muy alto nivel** describen conceptos muy generales y ofrecen nociones generales a las que hay que unir todos los términos raíz de las ontologías existentes. El problema principal de estas ontologías es que hay varias ontologías de alto nivel que difieren en su modo de clasificar los conceptos más generales de la taxonomía. La figura 2.6 muestra diferentes ontologías de alto nivel. Para resolver el problema de heterogeneidad en la clasificación (la cual es intrínseca o inherente al problema filosófico), el *IEEE Standard Upper Ontology (SUO) Working Group*<sup>10</sup> pretende especificar una ontología de alto nivel. Los conceptos específicos de un dominio dado no han de ser incluidos en *SUO*. Sin embargo, este estándar aporta una estructura y conjunto de conceptos generales desde los cuales se pueden construir ontologías de dominio (por ejemplo ontologías médicas, financieras, legales, etc).



<sup>10</sup> <http://suo.ieee.org/>

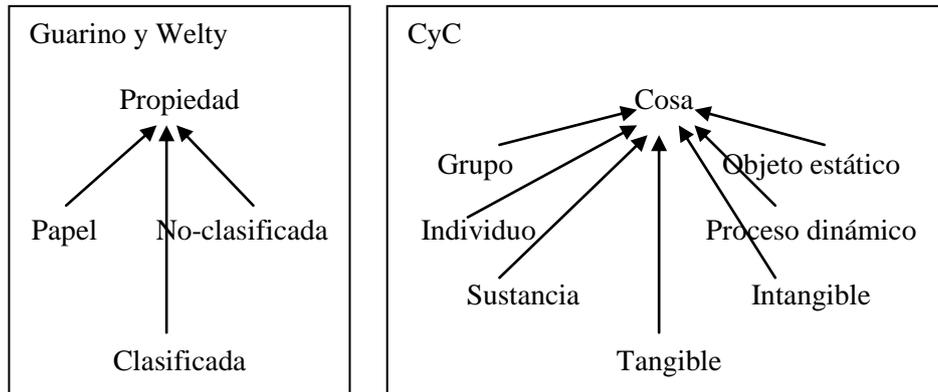


Figura 2.6. Ejemplos de ontologías de alto nivel (adaptados de Gómez y otros, 2004: 33).

Las **ontologías de dominio** (Mizoguchi y otros, 1995; Van Heijst y otros, 1997) son reutilizables en un dominio específico (médico, farmacéutico, ingeniería, derecho, etc.). Estas ontologías ofrecen vocabularios sobre conceptos de un dominio y sus relaciones, sobre las actividades que tienen lugar en tal dominio y sobre las teorías y los principios elementales que gobiernan ese dominio. Hay una línea divisoria clara entre las ontologías de dominio y las de más alto nivel. Los conceptos de las ontologías de dominio son normalmente especializaciones de conceptos ya definidos en las ontologías de alto nivel, y puede suceder algo similar con las relaciones. Por ejemplo, el término “ciudad” en una ontología de dominio es una especialización del concepto más general “ubicación”, que a su vez es una especialización del término “punto espacial” que puede definirse en la ontología de más alto nivel.

Las **ontologías de tareas** (Mizoguchi y otros, 1995; Guarino, 1998) describen el vocabulario relacionado con una determinada tarea o actividad (como el diagnóstico, la organización etc.) mediante la especialización de términos en las ontologías de alto nivel. Las ontologías de tareas ofrecen un vocabulario sistemático de términos que se utilizan para resolver problemas asociados con las tareas que pueden o no pertenecer a ese mismo dominio. Por ejemplo, la *Scheduling Task Ontology* presentada por Mizoguchi y otros (1995) incluye nombres genéricos, verbos genéricos, adjetivos genéricos, etc, como se muestra más arriba en la figura 2.2.

Las **ontologías de dominio-tarea** son ontologías de tareas reutilizables en un dominio determinado. Son independientes de su aplicación. Una ontología de dominio-tarea relativa a la planificación de viajes incluiría los términos “ciudad próxima”, “ciudad anterior”, etc.

Las **ontologías de método** aportan definiciones de conceptos relevantes y de las relaciones aplicadas para especificar el proceso de razonamiento para completar una tarea determinada (Tijerino y Mizoguchi, 1993). A esta categoría pertenecería una ontología sobre la planificación por medio de la descomposición de tareas.

Las **ontologías de aplicación** (Van Heijst y otros, 1997) son dependientes de su aplicación. Contienen todas las definiciones necesarias para modelar el conocimiento requerido para aplicación particular. Las ontologías de aplicación a menudo amplían y especializan el vocabulario del dominio y de las ontologías de tareas de una aplicación dada. Por ejemplo, podríamos crear una ontología de aplicación para las agencias de viajes españolas especializadas en destinos europeos.

A continuación, en la figura 2.7 ofrecemos la tipología de ontologías de Casellas (2011: 53) de acuerdo a los criterios de propósito, tema, generalidad y formalidad.

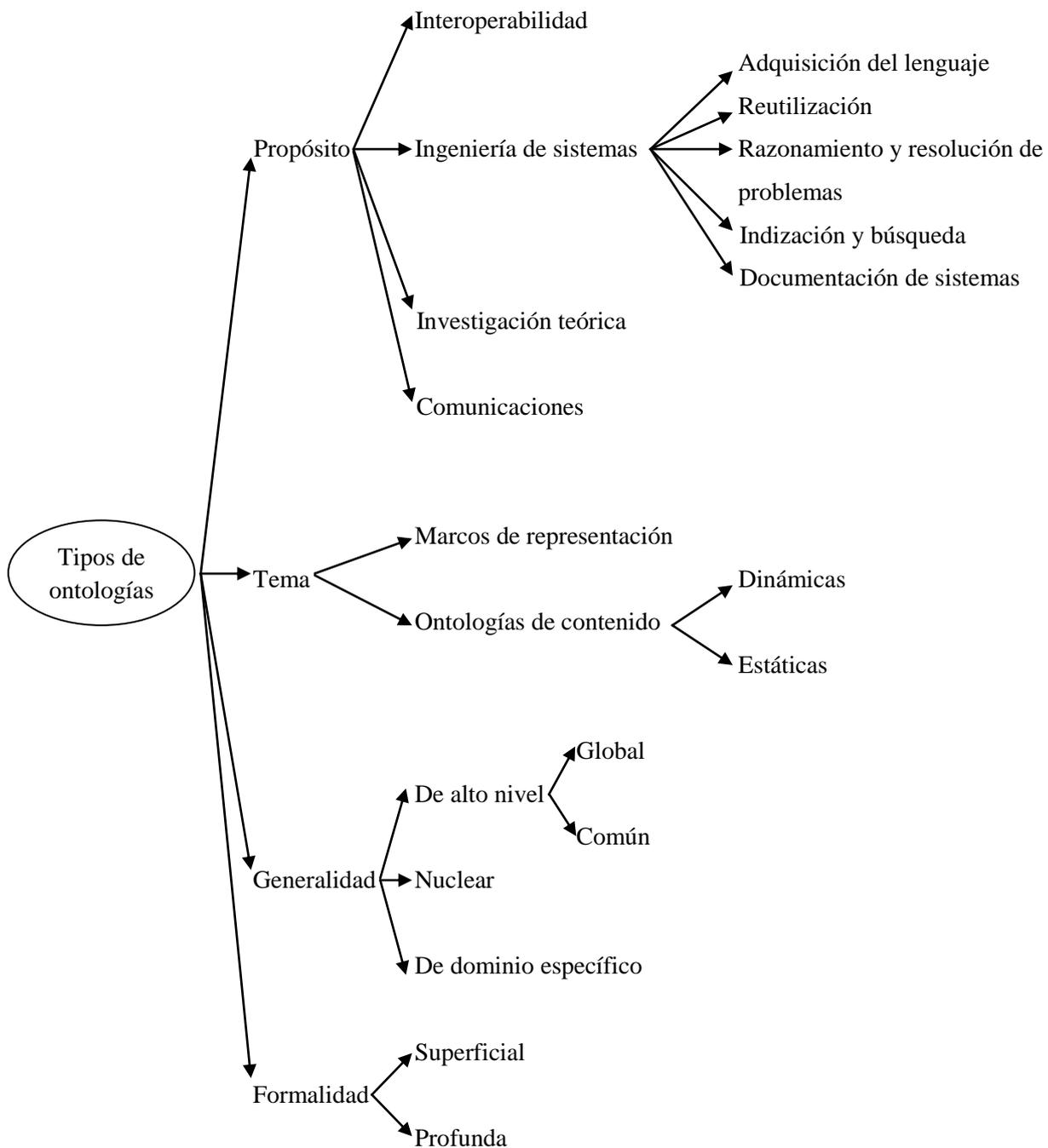


Figura 2.7. La tipología de ontologías (adaptada de Casellas, 2011: 53).

### 2.3.3. Criterios de diseño de ontologías

En cuanto al diseño de las ontologías, para Gruber (1992b) lo importante es el propósito para el que se concibe la ontología. En su trabajo de 1993 señaló que son necesarios criterios objetivos para diseñar una ontología basada en compartir el conocimiento. Estableció los siguientes criterios: a) Claridad: la ontología debe presentar definiciones claras de los términos en lenguaje natural; b) Coherencia; c) Extensibilidad; d) Codificación parcial mínima; e) Compromisos ontológicos mínimos (define sólo los términos básicos del dominio). Con respecto al último criterio, Uschold (1996) y Uschold y Grüninger (1996) señalan además que la ontología ha de ser extensiva y reutilizable mediante una codificación parcial mínima y compromisos ontológicos mínimos, ya que ambos favorecen la consistencia de la ontología.

Por su parte Gómez y Benjamins (1999: 1.2-1.3) van un paso más allá al completar los criterios de Gruber (1993) y establecer una lista de criterios de diseño de ontologías basada en Gruber (1995) y en las contribuciones de Arpírez y otros (1998), Borgo y otros (1996) y Bernaras y otros (1996):

- Claridad y objetividad. (Gruber, 1995). La ontología ha de mostrar términos con definiciones objetivas a través de un lenguaje natural.
- Definiciones completas. (Gruber, 1995). Es preferible una definición completa a una parcial.
- Coherencia. (Gruber, 1995). Permite inferencias consistentes con las definiciones.
- Máxima extensibilidad monotónica. (Gruber, 1995). Los términos nuevos generales o especializados deben estar incluidos en la ontología de modo que no sea necesario revisar las otras definiciones para su inclusión.
- Compromisos ontológicos mínimos. (Gruber, 1995). Realizar las menos afirmaciones posibles sobre el mundo modelado para dotar a la ontología de la libertad para especializarse.
- Principio de distinción ontológica. (Borgo y otros, 1996). Las clases en la ontología deben estar bien diferenciadas.
- Diversificación de las jerarquías. (Arpírez y otros, 1998). Para aumentar el poder de los mecanismos de herencia múltiple.

- Modularidad. (Bernaras y otros, 1996). Para minimizar las combinaciones entre módulos.
- Minimizar la distancia semántica entre conceptos hermanos. (Arpírez y otros, 1998). Implica que los conceptos similares se agrupan y se representan con los mismos primitivos semánticos.
- Estandarización de los nombres cuando sea posible (Arpírez y otros, 1998). Para facilitar el nivel de comprensión de la ontología han de usarse las mismas convenciones para nombrar a términos relacionados.

De todos estos criterios el que ha suscitado más interés ha sido el de los compromisos ontológicos mínimos, ya que restringen la interpretación semántica de los términos de la ontología en lenguaje natural. Tanto el tipo como el propósito de la ontología, y por tanto su reutilización y uso, influyen en el nivel de compromiso ontológico a seguir (cf. Casellas 2011: 46).

Gruber (1993) señala que la ontología ha de cumplir una serie de compromisos ontológicos, que son garantía de consistencia pero no de totalidad, con respecto a las cuestiones y afirmaciones que utilizan el vocabulario definido en la ontología. Más adelante Gruber y Olsen (1994) describieron los compromisos ontológicos como los acuerdos para utilizar un vocabulario compartido de una manera consistente y coherente. Guarino (1998) formalizó la definición de compromiso ontológico basada en la conexión entre el vocabulario de la ontología y el significado de los términos de ese vocabulario. Así, según esta autora, un compromiso ontológico es una función que une los términos del vocabulario de la ontología con una conceptualización.

Los compromisos ontológicos garantizan por tanto la consistencia de la ontología, pero no la totalidad de la misma, lo cual implica que se hace necesario establecer las menos afirmaciones posibles sobre el mundo que se modela. Cuando se construyen ontologías se utilizan términos del lenguaje natural normalmente en inglés para describir las conceptualizaciones de un dominio determinado. Normalmente los términos que se utilizan en lenguaje natural tienen varios significados, por lo que en la ontología se restringe su significado y se aportan definiciones formales. Así, el compromiso ontológico sería una especie de enlace entre los términos de la ontología y su

significado determinado. Por tanto, una característica principal de los compromisos ontológicos es determinar de una forma precisa el significado de un término (cf. Gómez Pérez 2004: 36).

Las relaciones entre el vocabulario, conceptualización, compromiso ontológico y ontologías aparecen en la figura 2.8. Los modelos intencionados del lenguaje de lógica reflejan su compromiso con una conceptualización. Una ontología refleja indirectamente su compromiso (y la conceptualización subyacente) por una aproximación de este conjunto de modelos.

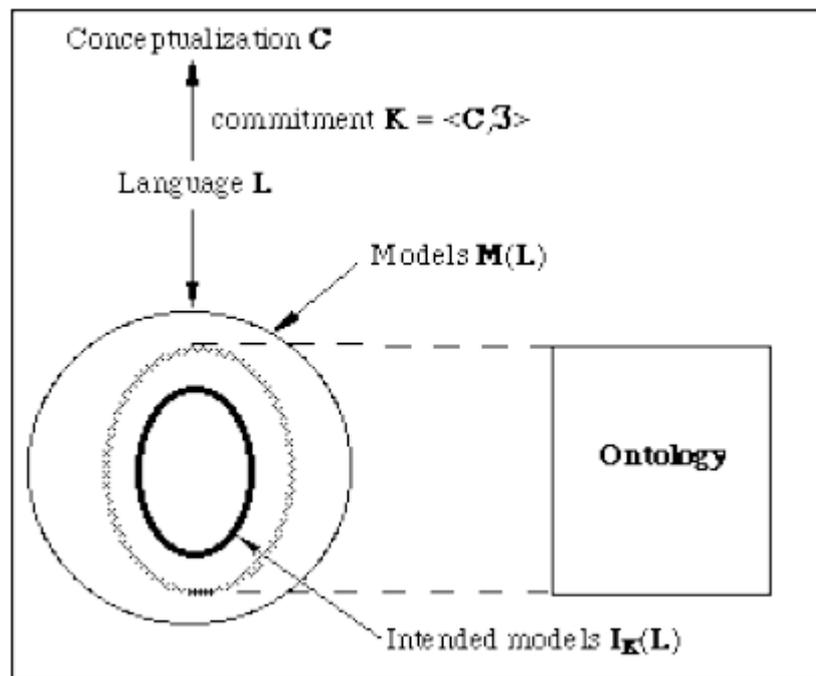
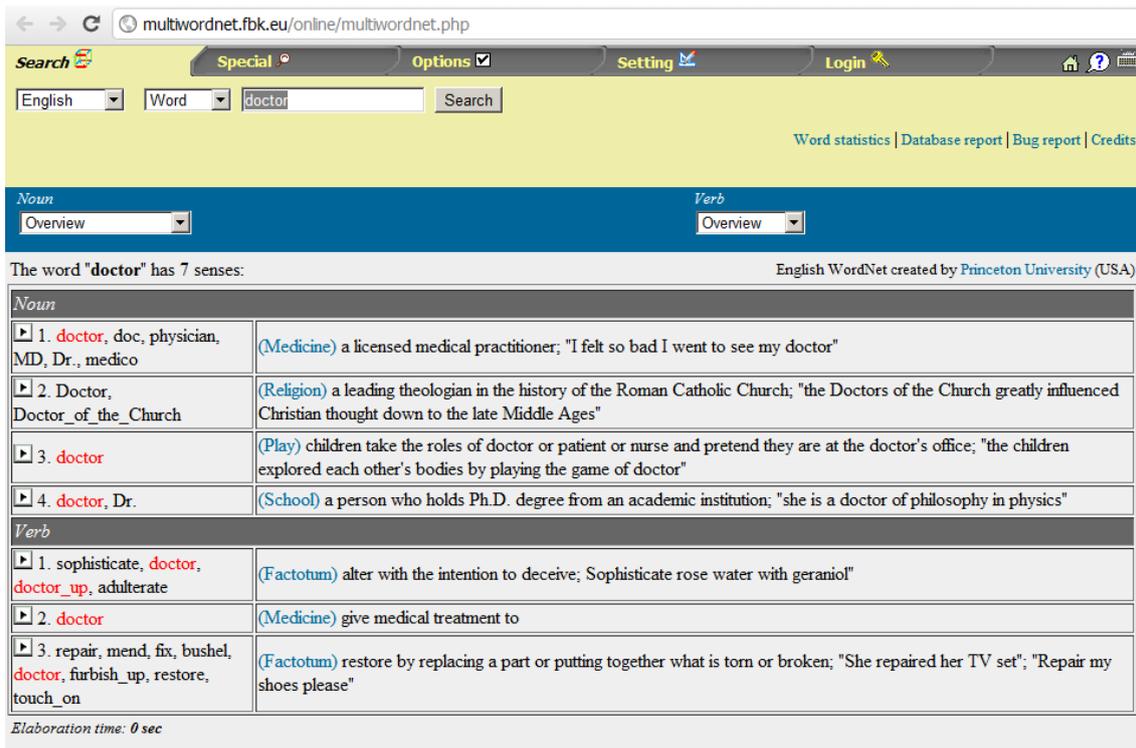


Figura 2.8. Los modelos intencionados como reflejo de su compromiso con una conceptualización (Guarino 1998: 7).

Para ilustrar un ejemplo de compromiso ontológico buscamos el término “médico” en *MultiWordNet*<sup>11</sup>, donde encontramos cuatro significados en la forma del sustantivo *doctor* (figura 2.9). Al analizar los cuatro vemos que sólo el primero es aplicable al ámbito de la medicina. Si elaboráramos una ontología del dominio de la medicina, el compromiso ontológico determinaría la relación entre el término en lenguaje natural y uno de sus significados, que en este caso sería el primero que aparece en *MultiWordNet*. Como la ontología debe restringir la semántica de los términos del dominio de acuerdo

<sup>11</sup> <http://multiwordnet.fbk.eu/online/multiwordnet.php/>

al compromiso ontológico elegido, es necesaria una definición formal para filtrar el significado correcto. Tal definición formal es una definición en la ontología y, como cualquier otra definición, puede ser extendida o especializada, pero las nuevas definiciones siempre han de ser consistentes con el compromiso ontológico.



The screenshot shows the MultiWordNet interface. The search bar contains the word "doctor". The results are categorized into Noun and Verb. The Noun section lists four senses: 1. doctor, doc, physician, MD, Dr., medico (Medicine); 2. Doctor, Doctor\_of\_the\_Church (Religion); 3. doctor (Play); 4. doctor, Dr. (School). The Verb section lists three senses: 1. sophisticate, doctor, doctor\_up, adulterate (Factotum); 2. doctor (Medicine); 3. repair, mend, fix, bushel, doctor, furbish\_up, restore, touch\_on (Factotum).

Noun	
1. doctor, doc, physician, MD, Dr., medico	(Medicine) a licensed medical practitioner; "I felt so bad I went to see my doctor"
2. Doctor, Doctor_of_the_Church	(Religion) a leading theologian in the history of the Roman Catholic Church; "the Doctors of the Church greatly influenced Christian thought down to the late Middle Ages"
3. doctor	(Play) children take the roles of doctor or patient or nurse and pretend they are at the doctor's office; "the children explored each other's bodies by playing the game of doctor"
4. doctor, Dr.	(School) a person who holds Ph.D. degree from an academic institution; "she is a doctor of philosophy in physics"

Verb	
1. sophisticate, doctor, doctor_up, adulterate	(Factotum) alter with the intention to deceive; Sophisticate rose water with geraniol"
2. doctor	(Medicine) give medical treatment to
3. repair, mend, fix, bushel, doctor, furbish_up, restore, touch_on	(Factotum) restore by replacing a part or putting together what is torn or broken; "She repaired her TV set"; "Repair my shoes please"

Figura 2.9. El término “médico” en *MultiWordNet*.

En general, los compromisos ontológicos implican algo más que establecer relaciones entre los términos en lenguaje natural y sus significados, puesto que incluyen también decisiones generales acerca de la estructura de tiempo, espacio, eventos, etc. (cf. Gómez, 2004: 38).

Por su parte, la ontología nuclear de *FunGramKB* cumple los siguientes compromisos ontológicos que garantizan su consistencia: (i) universalidad y motivación lingüística (la ontología nuclear de *FunGramKB* es de propósito general y tiene motivación lingüística); (ii) modelo ontológico en 3 niveles conceptuales (metaconceptos, conceptos básicos y conceptos terminales); (iii) unidades conceptuales no-atómicas (los conceptos básicos y terminales constan de marco temático y postulado de significado como esquemas conceptuales); (iv) los postulados de significado como organizadores de la ontología (todos los conceptos subordinados comparten el postulado de significado

de su superordinado); (v) relación taxonómica IS-A (los conceptos subordinados están ligados a sus superordinados a través de los postulados de significado); (vi) herencia múltiple (dos o más conceptos pueden incluir una misma unidad conceptual, creando así jerarquías complejas en forma de malla); y (vii) herencia no monótonica (a través de predicaciones rebatibles que permiten especificar excepciones a las reglas).

## 2.4. Ontologías y sistemas de librerías ontológicas

Entre las librerías de ontologías más conocidas están *DAML*<sup>12</sup>, *Ontolingua*<sup>13</sup>, *Protégé*<sup>14</sup>, *SHOE*<sup>15</sup>, *WebODE*<sup>16</sup>, *WebONTO*<sup>17</sup> y *(KA<sup>2</sup>)*<sup>18</sup>, de las cuales sólo la primera tiene relación con un lenguaje de representación de ontologías. Las otras librerías de ontologías mencionadas están asociadas a herramientas de creación de ontologías.

El problema de reutilización-uso de las ontologías (Klinker y otros, 1991) radica en que, cuanto más reutilizable es una ontología, menos útil se hace y viceversa (figura 2.10). Las ontologías de más alto nivel, las generales y las de dominio aportan conocimiento de forma independiente a la resolución de problemas, mientras que las ontologías de método, tarea y de dominio-tarea ofrecen conocimiento dependiente de la resolución de problemas. Todas estas ontologías se pueden utilizar partiendo de la misma o de diferentes librerías de ontologías y pueden combinarse para construir una ontología nueva. La figura 2.11 presenta una captura de pantalla de la librería de ontologías *DAML*.

---

<sup>12</sup> <http://www.daml.org/ontologies/>

<sup>13</sup> <http://ontolingua.stanford.edu/>

<sup>14</sup> <http://protege.stanford.edu/ontologies.html/>

<sup>15</sup> <http://www.cs.umd.edu/projects/plus/SHOE/onts/index.html/>

<sup>16</sup> <http://webode.dia.fi.upm.es/>

<sup>17</sup> <http://webonto.open.ac.uk/>

<sup>18</sup> <http://ka2portal.aifb.uni-karlsruhe.de/>

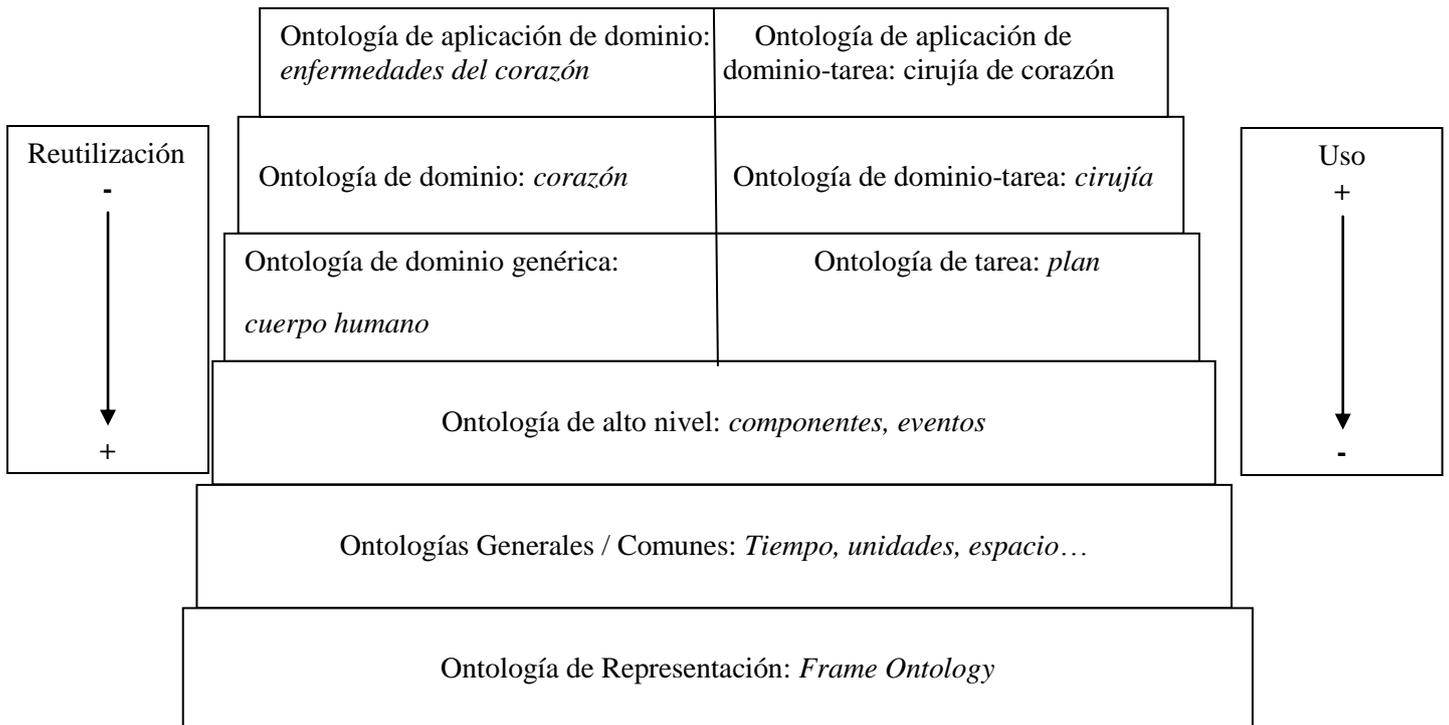


Figura 2.10. El problema de la reutilización y el uso (adaptado de Gómez y otros, 2004: 35).

## DAML Ontology Library

### Summaries

- [Ontologies by URI](#)
- [Ontologies by Submission Date](#)
- [Ontologies by Keyword](#)
- [Ontologies by Open Directory Category](#)
- [Ontologies by Class](#)
- [Ontologies by Property](#)
- [Ontologies by Namespace Used](#)
- [Ontologies by Funding Source](#)
- [Ontologies by Submitting Organization](#)

### Queries

- [Classes by Name](#)
- [Properties by Name](#)

### Statistics

- 282 ontologies
- [Number of Classes, Properties, and Instances Defined](#)
- [DAML Property \(Feature\) Use by Ontology](#)

Send updates/corrections to [ontology-librarian@daml.org](mailto:ontology-librarian@daml.org). Submit additions [here](#).

This catalog of DAML ontologies can also be viewed in [XML](#) and [DAML](#) formats.

---

Fri Apr 30 01:44:56 EDT 2004

Figura 2.11. Captura de pantalla de la librería de ontologías DAML.

## 2.5. Las ontologías más significativas

En esta sección presentamos la clasificación ofrecida por Gómez y otros (2004: 47-106) de las ontologías más significativas realizadas hasta el momento: ontologías de representación del conocimiento, ontologías de alto nivel, ontologías lingüísticas, y ontologías de dominio. Los autores elaboran esta clasificación de acuerdo a que estas ontologías pertenecen a proyectos significativos, por su contribución teórica o por su uso como base experimental para establecer criterios de diseño de ontologías o para elaborar metodologías, etc.

### 2.5.1. Las ontologías de representación del conocimiento

En primer lugar, encontramos las ontologías de representación del conocimiento (Van Heijst, 1997) que recogen los primitivos semánticos que se utilizan para modelar el conocimiento en un paradigma de representación (del conocimiento). Son ejemplos de primitivos semánticos las clases, las relaciones y los atributos.

La ontología de representación del conocimiento más significativa es la *Frame Ontology* (Gruber, 1993), construida para ejemplificar las convenciones establecidas por *Ontolingua* en una aproximación basada en marcos. La *Frame Ontology* (*FO*) fue modificada en 1997, momento en el que algunos de sus primitivos se movieron a la ontología *OKBC* (Chaudhri y otros, 1998). *OKBC* es un protocolo basado en marcos para acceder a bases de conocimiento formalizadas en lenguajes distintos como *Ontolingua* (Farquhar y otros, 1997), *LOOM* (MacGregor, 1991), *CycL* (Lenat y Guha, 1990) entre otros. Otros lenguajes como *CycL* (Lenat y Guha, 1990) y *OCML* (Motta, 1999) tienen sus propias ontologías de representación del conocimiento. Estas ontologías también están basadas en marcos y en lógica de primer orden, al igual que la ontología de marcos *Frame Ontology*. En los últimos años han aparecido lenguajes de marcado dentro del contexto de la Web Semántica como *RDF* (Beckett y McBride, 2004) y *RDF Schema* (Brickley y Guha, 2004), *OIL* (Horrocks y otros, 2000), *DAML+OIL* (Horrocks y van Harmelen, 2001) y *OWL* (Dean y Schreiber, 2004). Todos estos lenguajes tienen sus correspondientes ontologías de representación del conocimiento, las cuales exponemos a continuación (cf. Gómez, 2004: 47-48).

#### **2.5.1.1. Frame Ontology (FO) y OKBC**

La ontología de marcos (*FO*) (Gruber 1993) fue elaborada en *KIF* (Genesereth y Fikes, 1992) por el Laboratorio de sistemas del conocimiento de la Universidad de Stanford. Recoge convenciones comunes sobre organización del conocimiento utilizadas en representaciones del conocimiento basadas en marcos. Su propósito es unir la semántica de los primitivos semánticos más utilizados en el paradigma de marcos y permitir a los desarrolladores de ontologías construir ontologías con una aproximación basada en marcos. Como hemos señalado más arriba, fue modificada en 1997, y como resultado en la ontología *OKCB* se reemplazaron algunas definiciones de la *FO* original. Actualmente la *FO* incluye a la ontología *OKCB* y sólo ofrece definiciones formales de los primitivos semánticos no incluidos en esta última. Esta inclusión refleja que las ontologías construidas con los primitivos semánticos de la *FO* son más expresivas que las construidas con los primitivos semánticos de la ontología *OKCB*, que está principalmente relacionada con marcos, clases y ranuras (*slots*), mientras que la ontología de marcos (*FO*) incluye primitivos semánticos más complejos para representar funciones, relaciones y axiomas. Ambas ontologías están disponibles en la librería de ontologías del servidor de *Ontolingua*<sup>19</sup>(cf. Gómez, 2004: 48).

#### **2.5.1.2. RDF y RDF Schema**

Otra ontología de representación del conocimiento es *RDF* (Beckett, 2004). *RDF* significa *Resource Description Framework*. Es una recomendación del *W3C (World Wide Web Consortium)*, desarrollada para describir los recursos de la Web con metadatos. El modelo de datos de *RDF* es equivalente al paradigma de representación del conocimiento de red semántica, como explican Staab y otros (2000) y Conen y Klapsing (2001). El paradigma de representación del conocimiento de red semántica es menos expresivo que el paradigma de representación del conocimiento basado en marcos porque no permite representar valores predeterminados y límites cardinales de los atributos. El modelo de datos *RDF* tiene tres componentes: recursos, propiedades y afirmaciones. Los recursos son cualquier tipo de datos descritos en *RDF*. Se describen con expresiones *RDF* y se hace referencia a ellos a través de *URIs (Uniform Resource*

---

<sup>19</sup> <http://www.ksl.stanford.edu/software/ontolingua/>

*Identifiers*) y de identificadores de anclado. Las propiedades (predicados) definen los atributos y relaciones utilizadas para describir un recurso. Las afirmaciones asignan un valor a una propiedad en un recurso específico. Al igual que una frase en inglés normalmente consta de sujeto, predicado, verbo y objetos, las afirmaciones en *RDF* constan de sujetos, propiedades y objetos. Por ejemplo, en la frase “Laura dijo que María compró un vestido”, “Laura” es el sujeto, “dijo” es la propiedad y “María compró un vestido” es el objeto, que también a su vez se puede descomponer. A esto en *RDF* se le llama cosificación. Cabe señalar que el modelo de datos *RDF* no realiza suposiciones sobre la estructura de un documento con información en *RDF*. Es decir, las afirmaciones pueden aparecer en cualquier orden en la ontología *RDF* (cf. Gómez, 2004: 52).

La ontología *RDF* está escrita en *RDFS* y contiene 7 clases, 7 propiedades y una instancia como primitivos semánticos. No tiene primitivos semánticos para definir las relaciones entre propiedades y recursos. Por ejemplo, en *RDF* no podemos definir que la relación “lugar de llegada” sólo se puede establecer entre instancias de las clases “viajar” y “lugar”. Esta limitación se resuelve con el lenguaje de descripción de vocabulario *RDF* (Brickley y Guha, 2004), también conocido como *RDF Schema* o *RDFS*. *RDFS* es un documento de trabajo del *W3C* que amplía *RDF* con primitivos basados en marcos. La combinación de *RDF* y *RDFS* se conoce como *RDF(S)*. La ontología *RDFS* está escrita en *RDFS*. Contiene 16 primitivos semánticos nuevos (6 clases y 9 propiedades) añadidos a los primitivos semánticos de *RDF* (cf. Gómez, 2004: 53).

### 2.5.1.3. *OIL*

*OIL* (Horrocks y otros, 2001) significa *Ontology Inference Layer*. Este lenguaje es una extensión de *RDF(S)* con más primitivos semánticos basados en marcos y que evita el mecanismo de cosificación. Utiliza la lógica descriptiva para dotar de una semántica clara a sus primitivos semánticos. *OIL* se desarrolló con una aproximación por capas (figura 12). Cada nueva capa se construye encima de las ya existentes y añade funcionalidad y complejidad a la capa inferior. La capa *OIL nuclear* agrupa los primitivos semánticos de *OIL* que tienen un vínculo directo con los primitivos *RDF(S)*, aunque no permite la cosificación de *RDF(S)* como se ve en la figura 2.12. El *OIL estándar* añade primitivos semánticos basados en marcos. El *OIL de instancias* permite

definir instancias de conceptos y roles e incluye una capacidad de base de datos. El *OIL pesado* se reserva para extensiones futuras como reglas y metaclasses. El *OIL estándar* y el *OIL de instancias* comparten la misma ontología de representación del conocimiento.

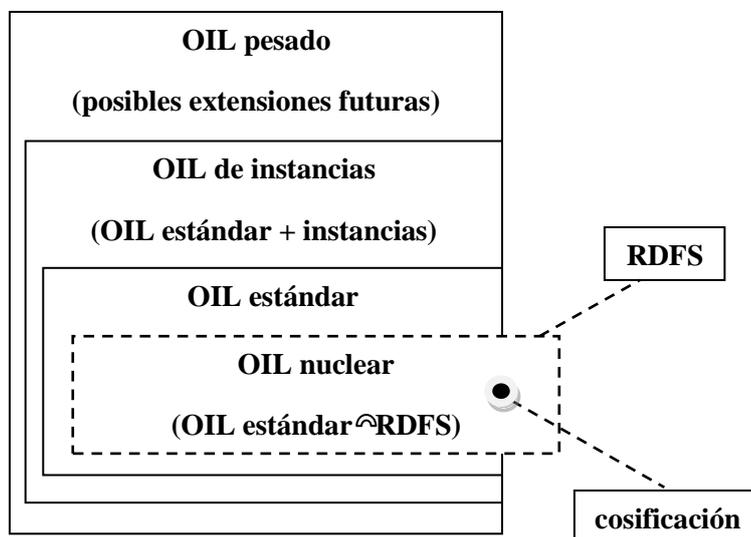


Figura 2.12. Las capas de *OIL* (adaptadas de Gómez y otros, 2004: 56).

#### 2.5.1.4. DAML + OIL

Al igual que *OIL*, la ontología *DAML+OIL* (Van Harmelen y otros, 2001) se desarrolló como extensión de *RDF(S)*. Sin embargo, este lenguaje no está dividido en capas sino que ofrece extensiones *DL* de *RDF(S)*. La ontología *DAML+OIL* está escrita en lenguaje *DAML+OIL* y contiene 53 primitivos semánticos (14 clases, 38 propiedades y una instancia). Dos de las clases y 10 de las propiedades son equivalentes a sus correspondientes clases y propiedades en *RDF(S)*. Las expresiones de clases de *DAML+OIL* se construyen con primitivos semánticos que son propiedades que permiten expresar: conjunción, disyunción, negación, grupos de personas y restricciones de propiedad.

#### 2.5.1.5. OWL

El lenguaje *OWL* (Dean y Schreiber, 2004) ha sido creado por el grupo de trabajo *W3C Web Ontology (WebOnt)*. Se deriva del lenguaje *DAML+OIL* y se construye sobre *RDF(S)*. Al igual que *OIL*, *OWL* se divide en capas: *OWL Lite*, *OWL DL* y *OWL Full*. *OWL Lite* amplía *RDF(S)* y recoge las características más comunes de *OWL*, por lo que está dirigida a usuarios que sólo necesitan crear taxonomías de clases y restricciones simples. *OWL DL* incluye el vocabulario de *OWL* completo. Finalmente *OWL Full*

ofrece más flexibilidad que *OWL DL* para representar ontologías. Hay 40 primitivos semánticos en la ontología *OWL DL* (16 clases y 24 propiedades). Algunos primitivos semánticos de *RDF(S)* pueden utilizarse en todas las versiones de *OWL*, y los primitivos semánticos de *OWL Lite* se pueden utilizar en *OWL DL*. Los primitivos semánticos de *OWL Full* son los mismos que los de *OWL DL*.

### 2.5.2. Ontologías de alto nivel

De acuerdo a la clasificación de Gómez y otros (2004: 71), en segundo lugar están las ontologías de alto nivel (también conocidas como de muy alto nivel). Estas ontologías describen conceptos muy generales que son comunes en los distintos dominios y aportan nociones generales a las que todos los términos en ontologías existentes deben vincularse. Las ontologías de alto nivel suelen utilizarse para construir ontologías de dominio, pero a veces las ontologías de dominio se construyen primero y luego se unen a las ontologías de alto nivel.

Son ontologías de alto nivel las ontologías de alto nivel de universales y particulares de Guarino y otros, la ontología *Sowa*, la ontología *CyC* y la del grupo de trabajo *Standard Upper Ontology (SUO)*. Las ontologías de alto nivel de universales y particulares están disponibles en *WebODE*. La *CyC Upper Ontology* está disponible en *CycL* y *DAML+OIL* y la *Standard Upper Ontology* está disponible en *KIF* y *DAML+OIL*.

#### 2.5.2.1 De universales y particulares

Guarino y sus colegas han construido dos ontologías de alto nivel, como se observa en la figura 2.13., una de universales y otra de particulares. Para estos autores un universal es un concepto como “coche” o “viajero”, mientras que un particular es un individuo como “mi coche” o “Jesús Sánchez”. Por consiguiente, los términos “coche” y “viajero” en una ontología de dominio pueden unirse a los particulares de alto nivel a través de la relación “subclase de”, y pueden unirse a los universales de alto nivel a través de la relación “instancia de”. A continuación presentamos estas ontologías con más detalle.

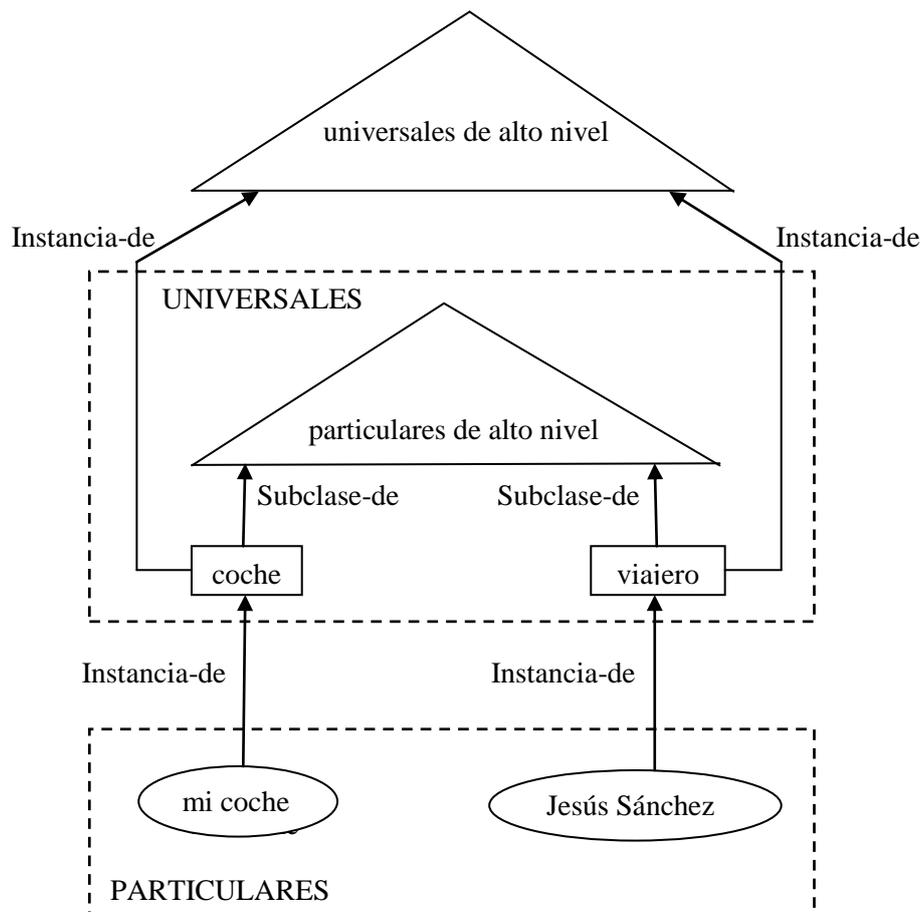


Figura 2.13. La relación entre particulares y universales (adaptada de Gómez y otros, 2004: 72).

La **ontología de alto nivel de universales** (Guarino y Welty, 2000) contiene conceptos cuyas instancias son universales. Esta ontología se ha modelado de acuerdo a las nociones filosóficas de rigidez, identidad y dependencia. Gangemi y otros (2001) definen la rigidez como la idea de “esencia”. Una propiedad es esencial para una instancia si y sólo si es necesaria para esta instancia. Así, una propiedad es rígida (+R) si y sólo si es necesariamente esencial para todas sus instancias; una propiedad es anti-rígida ( $\square R$ ) si y solo si no es esencial para todas las instancias; y una propiedad es no rígida (-R) si y sólo si no es esencial para alguna de sus instancias. Por ejemplo, el concepto “persona” es normalmente considerado rígido puesto que cada persona esencialmente lo es. El concepto “viajero” es considerado anti-rígido porque cada viajero puede posiblemente ser no viajero una vez que el viaje ha finalizado. Finalmente, el concepto “naranja” es no-rígido ya que hay instancias que son esencialmente naranjas (por ejemplo una naranja), y otras que no son esencialmente naranjas (mis calcetines).

Una propiedad implica un criterio de identidad (+I) si y sólo si todas sus instancias pueden (re)identificarse a través de una relación de “igualdad”. Una propiedad aporta un criterio de identidad (+O) si y sólo si tal criterio no lo hereda por ninguna propiedad de subsunción. Por ejemplo, si tomamos el ADN como criterio de identidad podemos decir que la persona no solo lleva consigo el criterio de identidad sino que también lo aporta. Además, si “viajero” es una subclase de “persona”, entonces el “viajero” sólo hereda el criterio de identidad de la persona sin ningún otro criterio más de identidad.

Un individuo  $x$  es dependiente de un individuo  $y$  si y sólo si, en cualquier momento  $x$  no puede estar presente a no ser que  $y$  esté totalmente presente, e  $y$  no es parte de  $x$ . Por ejemplo, un agujero en la pared es constantemente dependiente de la pared. El agujero no puede estar presente si la pared no está presente. La dependencia constante de una propiedad se define de acuerdo a la dependencia constante de los individuos.

Cada concepto de la **ontología de alto nivel de universales** tiene cuatro atributos: *rigidez*, *aporta identidad*, *lleva identidad* y *dependencia*. Esta ontología se ha construido considerando varias combinaciones de valores de estos atributos (Welty y Guarino, 2001). Cada concepto de un dominio específico será instancia de al menos una de las hojas del nivel superior de los universales (cf. Gómez y otros, 2004: 73). En la figura 2.14 se observa la taxonomía de clases de esta ontología.

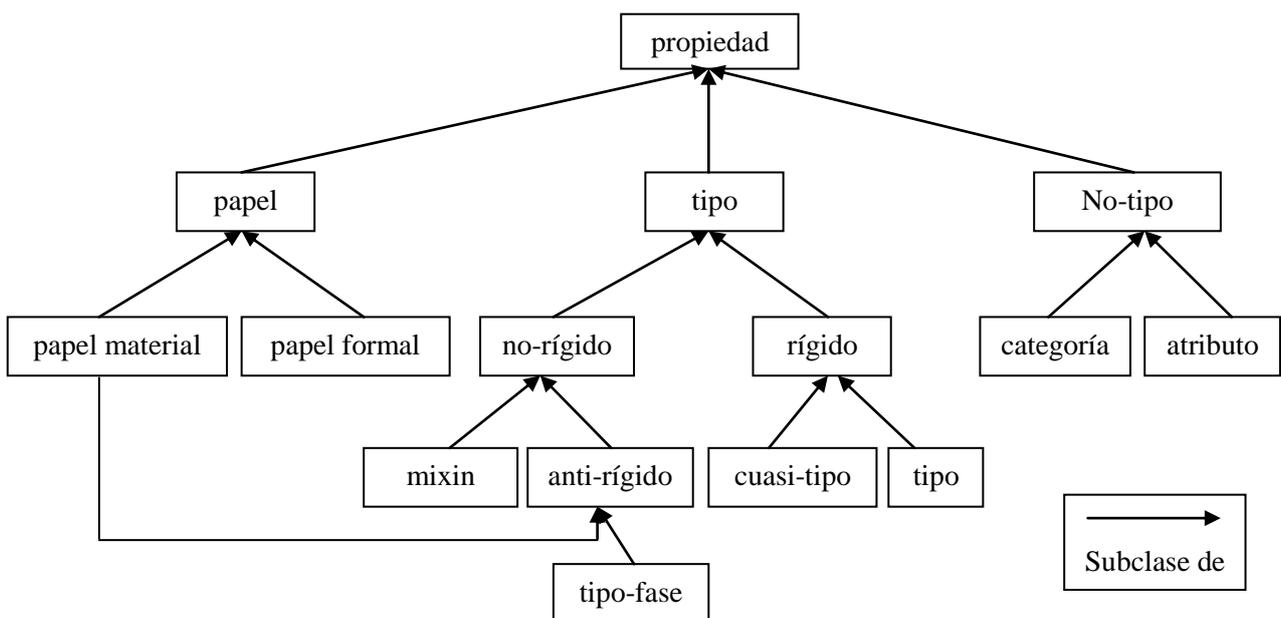


Figura 2.14. La taxonomía de clases de la ontología de alto nivel de universales (adaptada de Gómez y otros, 2004: 74).

La **ontología de alto nivel de particulares** (Gangemi y otros, 2001) tiene conceptos generales, como por ejemplo “objeto” a los que se pueden vincular los conceptos de un dominio específico mediante la relación “subclase de”. La figura 2.15. ofrece una representación parcial de la taxonomía de clases de esta ontología. Podemos observar que la ontología tiene tres raíces (abstracto, concreto y relación). Se ha desarrollado de acuerdo a los principios establecidos en el método *OntoClean* para “limpiar” ontologías (Welty y Guarino, 2001).

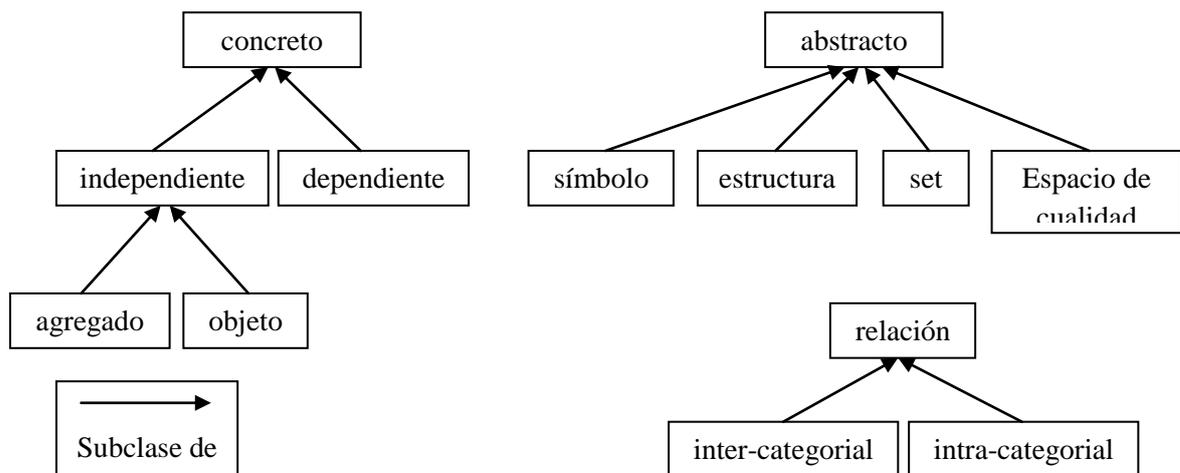


Figura 2.15. Vista parcial de la taxonomía de clases de la ontología de alto nivel de particulares (adaptada de Gómez y otros, 2004: 74).

Si quisiéramos unir conceptos del dominio de los viajes a la ontología de alto nivel de universales y particulares, los conceptos del dominio (por ejemplo, “coche”, “viajero”) serían subclases de los conceptos de la ontología de alto nivel de particulares, y también serían instancias de conceptos de la ontología de alto nivel de universales. Las instancias como “mi coche” y “Jesús Sánchez” son instancias de clases vinculadas a los particulares de alto nivel. Tanto la **ontología de alto nivel de universales** como la de **particulares** están disponibles en la estación de trabajo *WebODE* (cf. Gómez y otros 2004: 75).

2.5.2.2. Sowa

La ontología de alto nivel SOWA incluye las categorías básicas y distinciones que se han derivado de una variedad de fuentes en lógica, lingüística, filosofía e inteligencia artificial (Sowa, 1999). Esta ontología tiene 27 conceptos. Aparecen todos reflejados en la figura 2.16.

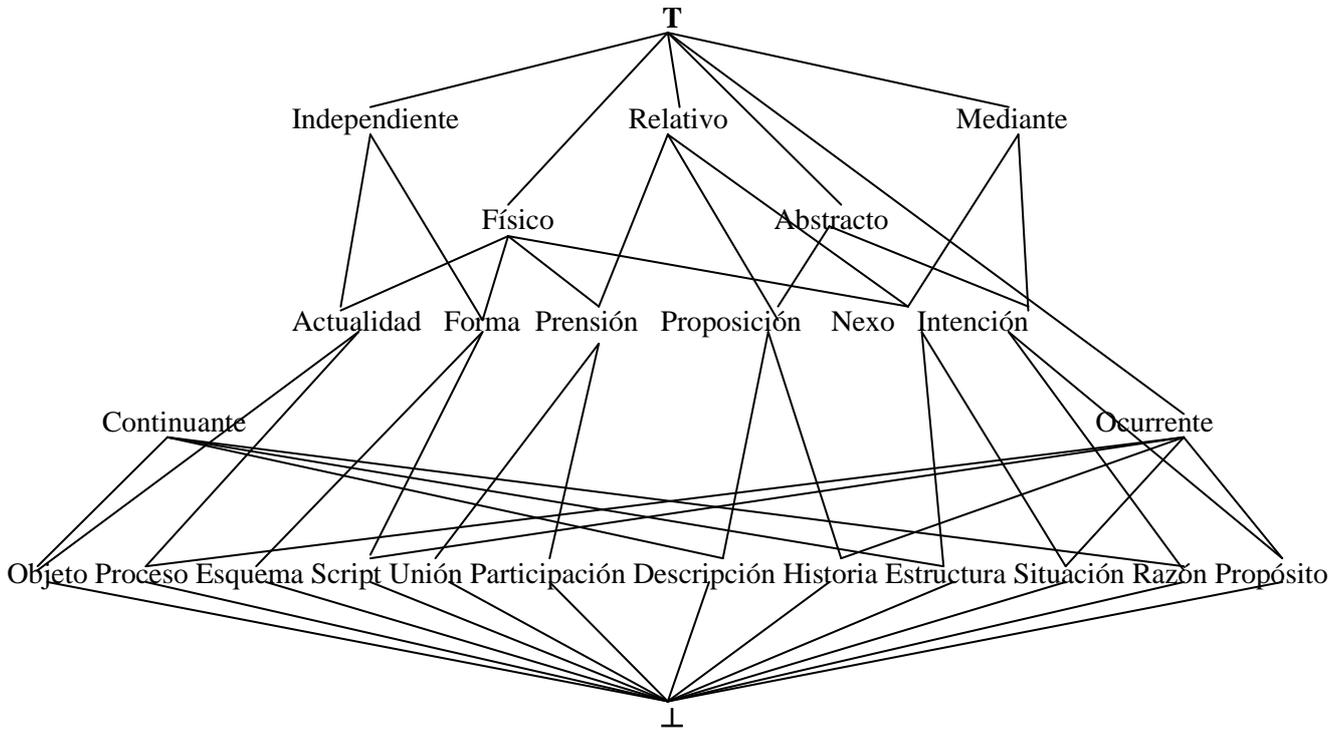


Figura 2.16. La ontología de alto nivel Sowa (adaptada de Gómez y otros, 2004: 75).

Esta ontología tiene una estructura en forma de rejilla porque cada par de conceptos de la taxonomía tiene, al menos, una superclase común directa o indirecta, y cada par de conceptos tiene, al menos, una subclase común directa o indirecta. En esta ontología el concepto superior es el *tipo universal* (T), y el concepto inferior es el *tipo absurdo* ( $\perp$ ). El *tipo universal* contiene todas las instancias posibles de la ontología. El *tipo absurdo* no tiene instancias y es una subclase de cada concepto de la taxonomía.

### **2.5.2.3. CyC**

La ontología de alto nivel *CyC* es parte de la base de conocimiento *CyC* (Lenat y Guha, 1990), que contiene una gran cantidad de conocimiento del sentido común. La base de conocimiento *CyC* se ha construido sobre un núcleo de 1.000.000 afirmaciones y se ha diseñado para recoger una gran parte de lo que se considera conocimiento compartido del mundo. Está dividida en cientos de microteorías (conjuntos de afirmaciones en el mismo dominio) y está implementada en el lenguaje *CycL*.

La ontología de alto nivel *CyC* contiene alrededor de 3.000 términos clasificados en 43 grupos temáticos (fundamentales, tiempo y fechas, relaciones espaciales, etc.). La clase “cosa” es la raíz de la ontología y también es un conjunto universal. La ontología se ha construido siguiendo los siguientes pasos: (1) la división del conjunto universal en tangible e intangible, en cosa estática frente a proceso dinámico, en grupo frente a individual, etc; (2) refinando el resultado cuando se introduce conocimiento nuevo a través de nuevos conceptos, nuevas relaciones de “subclase de”, etc.). Cabe destacar que durante este proceso algunas categorías pueden desaparecer. La figura 2.17 muestra un fragmento de la ontología. *Cycorp*<sup>20</sup> es la empresa que gestiona la ontología, que ofrece varias herramientas para ayudar al usuario en el manejo de la ontología. Entre estas herramientas están enlaces de hipertexto que permiten hojear directamente las taxonomías y navegar por las referencias (un listado temático de la ontología de alto nivel dividido en áreas para facilitar un estudio sistemático, etc) (cf. Gómez 2004: 76-77).

---

<sup>20</sup> <http://www.cyc.com/>

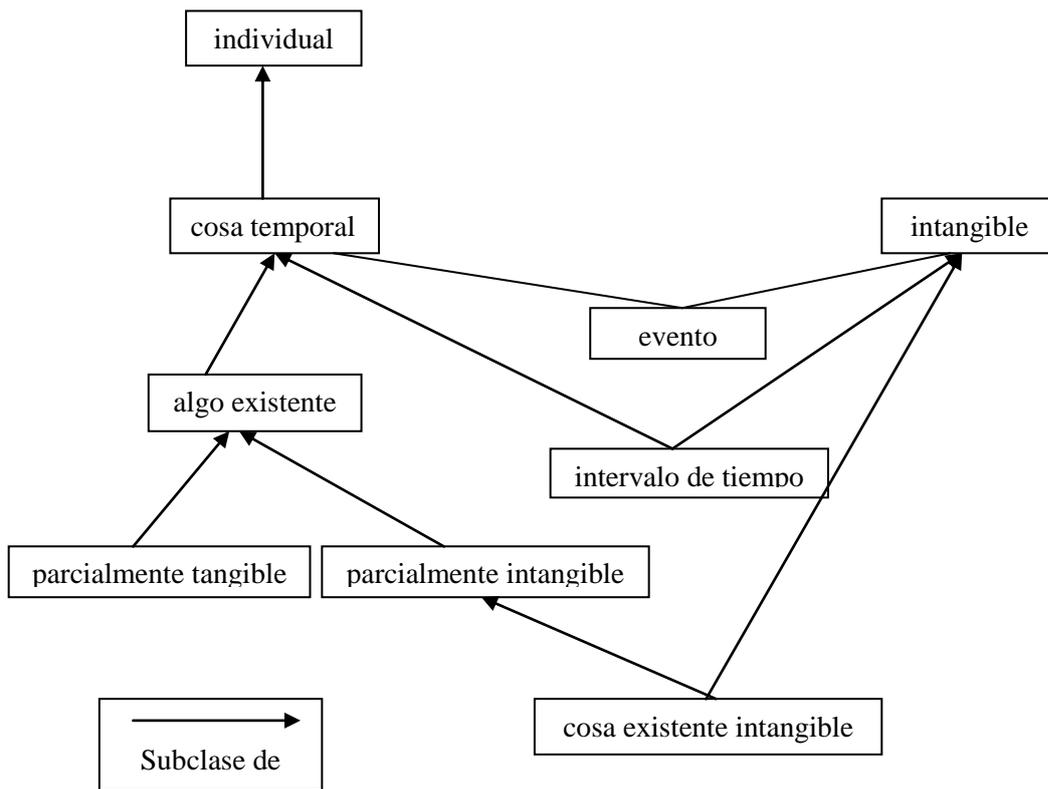


Figura 2.17. Fragmento de la taxonomía de clases de la ontología de alto nivel Cyc (adaptado de Gómez y otros, 2004: 76).

#### 2.5.2.4. The Standard Upper Ontology (SUO)

La ontología *SUO* (*Standard Upper Ontology*)<sup>21</sup> es el resultado del esfuerzo de unión para crear una gran ontología formal de propósito general (Pease y Niles, 2002). Promovida por el grupo de trabajo *IEEE Standard Upper Ontology* su desarrollo comenzó en mayo de 2000. Uno de los documentos fundacionales es conocido como *SUMO* (*Suggested Upper Merged Ontology*). A continuación ofrecemos parte de su estructura.

<sup>21</sup> <http://suo.ieee.org>

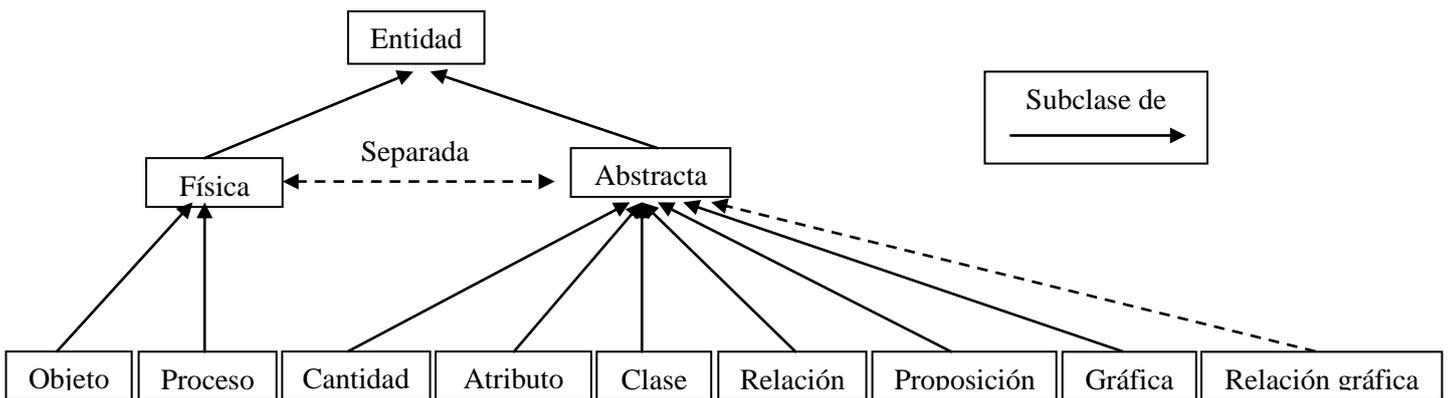


Figura 2.18. La estructura de los primeros niveles de *SUMO* (mayo de 2002) (adaptada de Gómez y otros 2004: 77).

El propósito de *SUMO* era crear una ontología de alto nivel integral y consistente con algunos de los mejores recursos públicos del momento, tales como la mereotopología del grupo *CNR* (Borgo y otros, 1996; Borgo y otros, 1997); las ontologías de alto nivel *Sowa* y la de Russell y Norvig (1995); los axiomas temporales de Allen (1984); y las teorías de planes y procesos de Pease y Carrico (1997).

Por tanto, *SUMO* establece distinciones de alto nivel y contiene conceptos y procesos temporales. Se trata de una ontología modular dividida en subontologías. En la siguiente figura podemos observar las relaciones de dependencia entre varias subontologías.

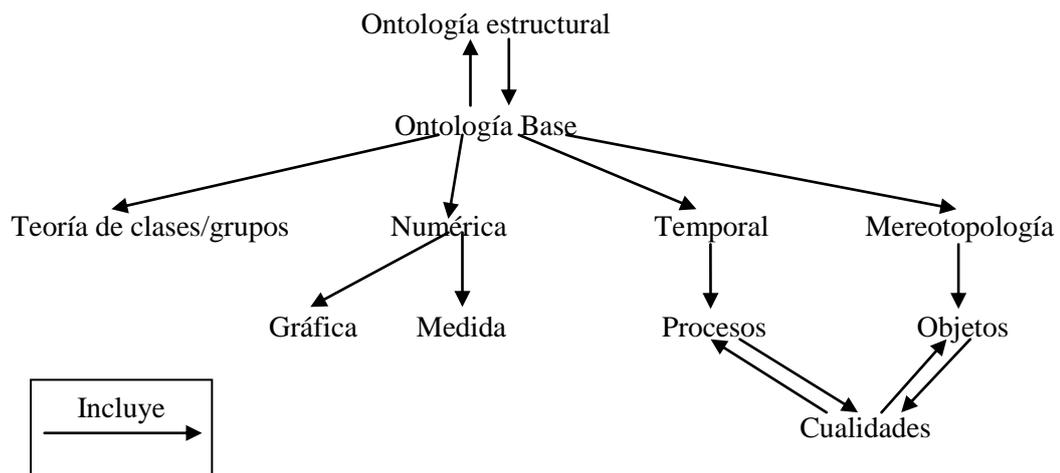


Figura 2.19. La estructura modular de *SUMO* (adaptado de Gómez y otros, 2004: 78).

En el proceso de evolución de esta ontología de alto nivel, para decidir qué conceptos había que eliminar, añadir o mantener, Niles y Pease (2001) establecieron los pasos a seguir:

1. Utilizar la versión actual de la ontología de alto nivel como base y añadirle las ontologías de niveles inferiores de una en una.
2. Eliminar de la ontología de alto nivel los conceptos no relacionados con ningún concepto de las ontologías de dominio y añadir los conceptos de alto nivel susceptibles de modelar las ontologías de nivel inferior.

### 2.5.3. Ontologías lingüísticas

En el tercer grupo de ontologías, Gómez y otros (2004: 79) clasifican las ontologías lingüísticas, que tienen como propósito describir constructos semánticos en lugar de un dominio específico. La característica principal de estas ontologías es que están ligadas a la semántica de sus unidades gramaticales (palabras, grupos nominales, adjetivos, verbos, etc.).

La mayoría de las ontologías lingüísticas utilizan palabras como unidades gramaticales. De las analizadas a continuación, sólo la ontología *Generalised Upper Model (GUM)* y *SENSUS* recogen información sobre las unidades gramaticales superiores a las palabras. Otras ontologías como *WordNet* se centran en el significado de las palabras. Además, en ontologías como *EuroWordNet* hay una relación unívoca entre el concepto y la palabra en lenguaje natural, mientras que en otras muchos conceptos no están vinculados a ninguna palabra en una lengua o están vinculados a varias palabras en la misma lengua, como sucede en la ontología *Mikrokosmos*.

En cuanto a su nivel de dependencia lingüística, hay que señalar que algunas ontologías lingüísticas son monolingües, como *WordNet*, y otras multilingües, como *GUM*; las hay también que contienen una parte dependiente de la lengua y otra que no (*EuroWordNet*); y otras que son totalmente independientes de la lengua (*Mikrokosmos*).

En su origen la motivación para su creación también varía. *WordNet* surgió como una base de datos léxica en línea; *SENSUS* se creó para labores de traducción automática; y *GUM* para generar lenguaje natural. En cuanto a estas dos últimas, también se las ha considerado ontologías de alto nivel porque contienen conceptos muy abstractos. A continuación analizamos en más detalle las ontologías anteriormente citadas.

### 2.5.3.1. WordNet

*WordNet*<sup>22</sup> (Miller y otros, 1990; Miller, 1995) es una gran base de datos léxica en inglés creada por la Universidad de Princeton. Los sustantivos, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos (*synsets*), donde cada uno expresa conceptos distintos. Los *synsets* se entrelazan por medio de relaciones conceptuales-semánticas y léxicas. En total hay 117.000 *synsets* que a su vez están ligados a otros *synsets* a través de un número limitado de relaciones conceptuales. Adicionalmente, cada *synset* contiene una breve definición (glosa) y, en muchos casos, se ofrece una o varias frases para ilustrar el uso de los componentes del *synset*. Las palabras con varios significados se representan en varios *synsets*, por lo que cada par de forma-significado es único. Los *synsets* se interrelacionan a través de relaciones de sinonimia y antonimia, hiperonimia e hiponimia (“subclase de”, y “superclase de”), meronimia y holonimia (“parte de” y “Has-a”). *WordNet* divide el lexicón en cinco categorías: sustantivos, adjetivos, adverbios y palabras funcionales. Los sustantivos se organizan en jerarquías temáticas. En la figura 2.20 se ofrece parte de la jerarquía de los sustantivos donde aparecen relacionados los términos de la persona, sus componentes, sus sustancias y su organización familiar. En este ejemplo podemos ver relaciones de meronimia, antonimia e hiponimia.

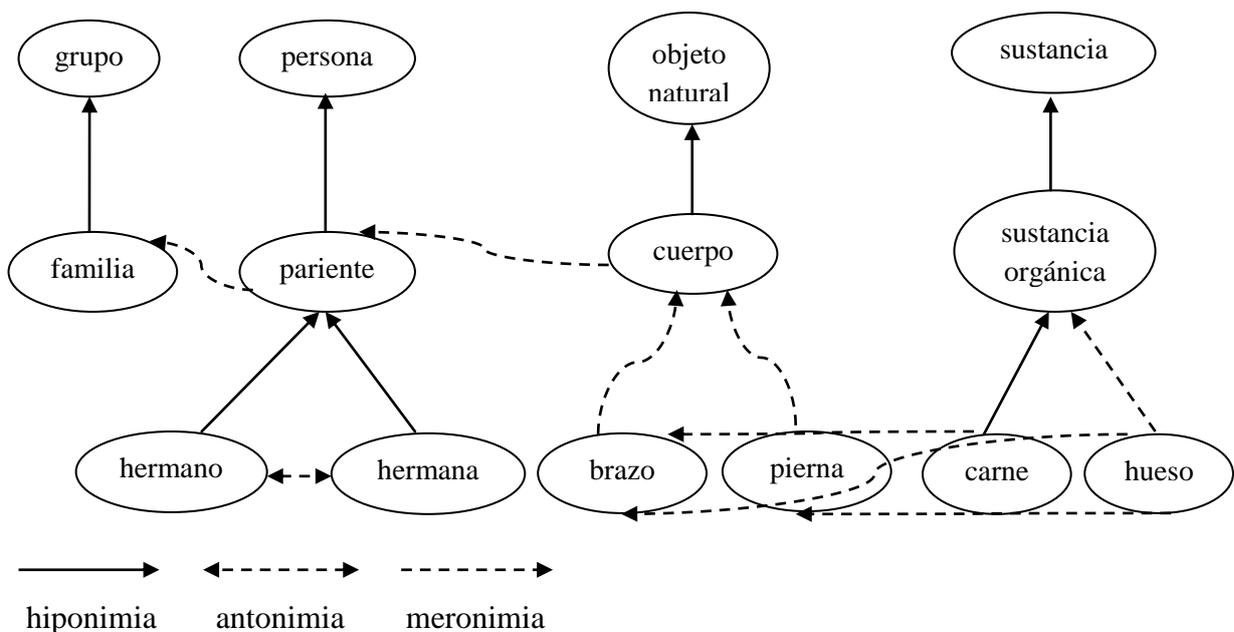


Figura 2.20. Extracto de la categoría de sustantivos en *Wordnet*. (adaptada de Gómez y otros, 2004: 80).

<sup>22</sup> <http://wordnet.princeton.edu/>

Los verbos se organizan de acuerdo a una variedad de relaciones de vínculos. Por ejemplo, los verbos *succeed* y *try* se relacionan por una implicación hacia atrás, y *buy* y *pay* se relacionan por una inclusión temporal. Con los adjetivos y los verbos las relaciones de sinonimia y antonimia ocupan un importante papel.

La red resultante de palabras y conceptos relacionados significativamente se puede consultar en el navegador<sup>23</sup>. *WordNet* es de libre acceso para su descarga. Su estructura hace que sea una herramienta útil para la lingüística computacional y el procesamiento del lenguaje natural.

Superficialmente tiene el aspecto de un Tesauro porque agrupa las palabras en base a su significado, aunque realmente entrelaza los distintos significados de las palabras y no simplemente la forma de la palabra (cadena de letras).

### **2.5.3.2. EuroWordnet**

La segunda ontología lingüística de la clasificación de Gómez y otros (2004: 80) es *EuroWordnet*<sup>24</sup> (Vossen, 1998; 1999). Se trata de una base de datos multilingüe con conjuntos de palabras (*wordnets*) para varias lenguas europeas (holandés, italiano, español, alemán, francés, checo y estonio). Algunas de las instituciones que participaron en el proyecto fueron la Universidad de Amsterdam (Holanda), la UNED (España) y la Universidad de Sheffield (Gran Bretaña). Los *wordnets* se estructuran en *synsets* (conjuntos de palabras sinónimas) al igual que el *WordNet* del inglés de la Universidad de Princeton, de Miller y otros (1990). Cada *wordnet* representa un sistema interno único de lexicalizaciones de la lengua. Además, los *wordnets* están vinculados a un índice entre lenguas basado en *WordNet*. A través de este índice, las lenguas se interconectan por lo que es posible navegar entre las palabras de una lengua y las palabras similares en otra lengua. El índice también da acceso a la ontología superior de 63 distinciones semánticas compartidas. Esta ontología superior ofrece un marco semántico común para todas las lenguas, mientras que las propiedades específicas de cada lengua se mantienen en los *wordnets* individuales. La base de datos se puede

---

<sup>23</sup> <http://wordnetweb.princeton.edu/perl/webwn>

<sup>24</sup> <http://www.illc.uva.nl/EuroWordNet/>

utilizar, entre otros, para recuperar información tanto monolingüe como multilingüe, como demostraron los usuarios del proyecto.

El proyecto finalizó en el verano de 1999. El diseño de la base de datos de *EuroWordNet*, las relaciones definidas, la ontología de nivel superior y el índice entre lenguas están ahora parados. Sin embargo, muchos otros institutos y grupos de trabajo siguen desarrollando *wordnets* similares en otras lenguas (tanto europeas como no europeas). Si se muestran compatibles, estos *wordnets* podrán añadirse a la base de datos de *EuroWordNet* y, a través del índice, podrán conectarse con cualquier otro *wordnet*. Actualmente se están desarrollando *wordnets* en sueco, noruego, danés, griego, portugués, vasco, catalán, rumano, lituano, ruso, búlgaro y eslovaco.

La asociación *Global WordNet Association*<sup>25</sup> continúa el marco cooperativo de *EuroWordNet*, una asociación pública que trabaja tanto en *EuroWordnet* y *WordNet*. Su propósito es estimular la construcción de más *WordNets*, una mayor estandarización e interrelación y el desarrollo de herramientas y diseminación de información.

### **2.5.3.3. The Generalized Upper Model (GUM)**

La *Generalized Upper Model (GUM)*<sup>26</sup> (Bateman y otros, 1995) es la tercera ontología lingüística en la clasificación de Gómez y otros (2004: 82). Es una ontología lingüística general que modela la semántica de las expresiones del lenguaje natural. Tiene un propósito multilingüe. Es el resultado de la evolución continua que comenzó con el *Penman Upper Model*, utilizado en el sistema *Penman* de generación de textos (Bateman y otros, 1990). Tres organizaciones tomaron parte del desarrollo de *GUM*: el Instituto de Ciencias de la Información (*ISI*, EEUU), *GMD/IPSI* (Alemania) y el Instituto de Investigaciones Científicas y Tecnológicas (*CNR*, Italia). Tiene dos jerarquías, una de conceptos y otra de relaciones. En la figura 2.21 ofrecemos los primeros niveles de estas jerarquías.

---

<sup>25</sup> <http://www.globalwordnet.org/>

<sup>26</sup> <http://www.fb10.uni-bremen.de/anglistik/langpro/webSPACE/jb/gum/index.htm>

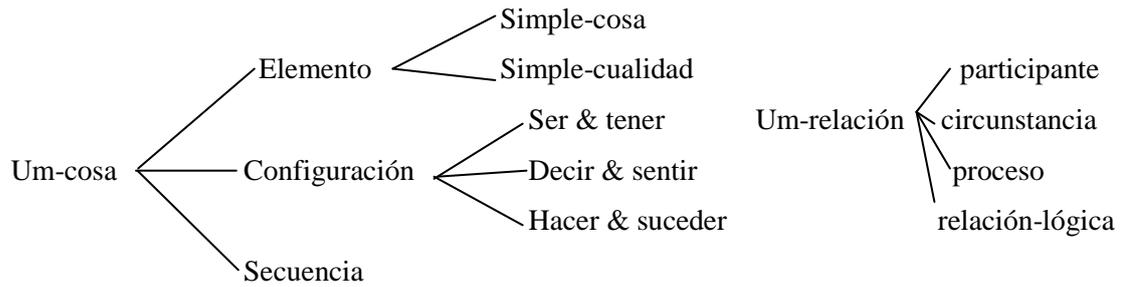


Figura 2.21. Los primeros niveles de las jerarquías de *GUM* (adaptado de Gómez y otros 2004: 83).

Está disponible en varias formas incluyendo una versión para la Web Semántica, compatible con *OWL-DL* y emplea principios actualizados de ingeniería ontológica. Las taxonomías empleadas en *GUM* tienen su origen en la gramática funcional de Halliday (1985), pero se pueden aplicar a cualquier teoría. El concepto jerarquía representa las entidades semánticas básicas e incluye configuraciones de los procesos, y los diferentes tipos de objetos y cualidades. Una configuración es un conjunto de objetos que participan en alguna actividad o que están en algún estado. Un ejemplo de configuración es “ser & tener”, que indica la existencia de algo o una relación de identidad, posesión o atributo. La jerarquía de relaciones representa a los participantes y las circunstancias que forman parte de los procesos, y las combinaciones lógicas entre ellos.

#### 2.5.3.4. *Mikrokosmos*

La ontología *Mikrokosmos* (Mahesh y Nirenburg, 1995; Mahesh, 1996) es una ontología independiente de la lengua que es parte del proyecto *Mikrokosmos* de traducción automática en el dominio de las fusiones y la adquisición de compañías. En este proyecto participaron la Universidad del estado de Nuevo Méjico, la Universidad Carnegie Mellon y otras organizaciones del gobierno de EEUU.

Como señala Mahesh (1996), *Mikrokosmos* no se construyó de acuerdo a ninguna teoría ontológica particular, sino que está construida en base a consideraciones más prácticas. Su principal criterio de diseño es una cuidadosa distinción entre el conocimiento específico de una lengua representado en el lexicón, y el conocimiento neutral de la lengua representado en la ontología. Las entradas del lexicón representan el significado de las palabras o frases por su vinculación con los conceptos de la ontología.

La figura 2.22 muestra los primeros niveles de la ontología. Esta ontología tiene varios miles de conceptos. Cada concepto se representa con un marco, que tiene el nombre en inglés. Los términos que están también en inglés se utilizan como referencia de cada concepto en la ontología.

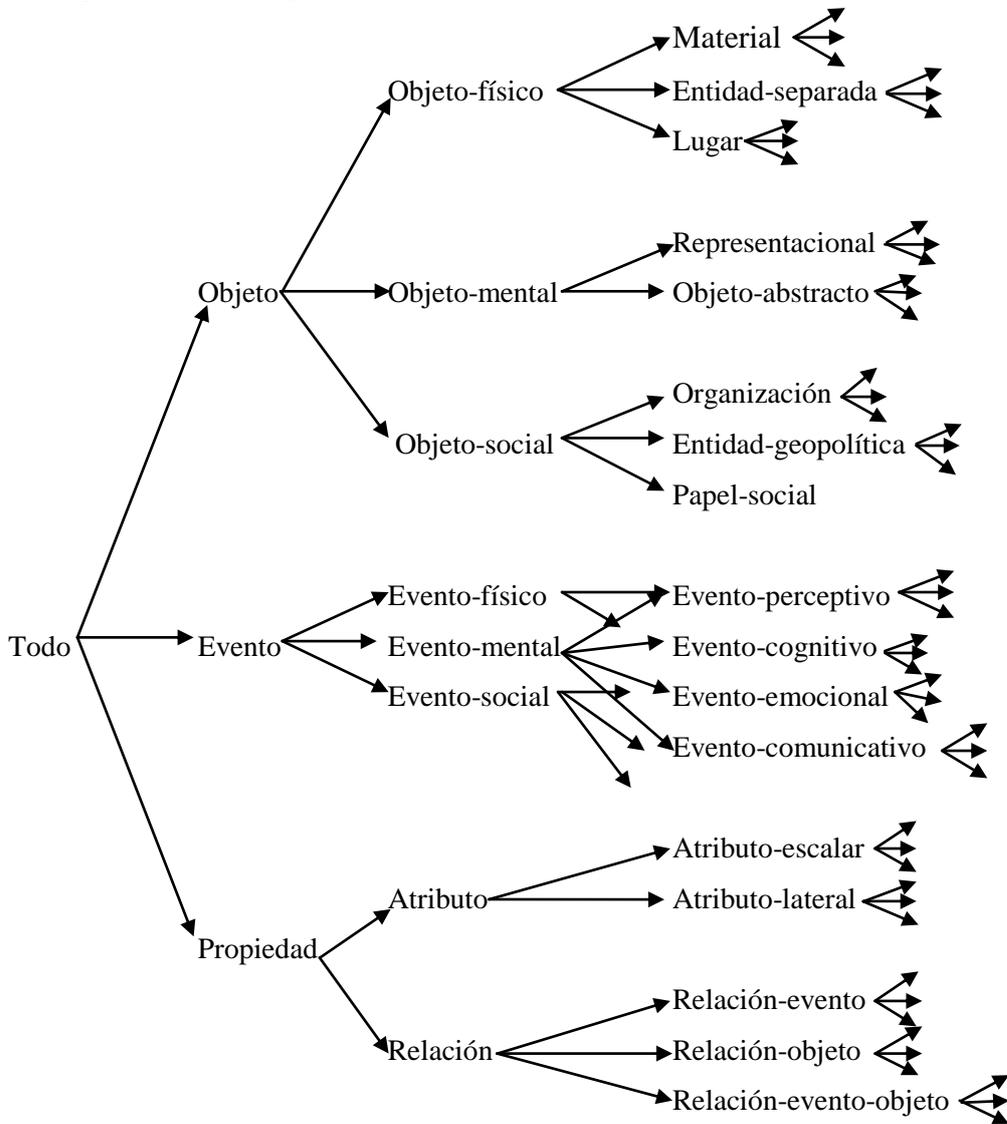


Figura 2.22. La taxonomía de clases en *Mikrokosmos* (adaptada de Mahesh y Nirenburg, 1995: 84).

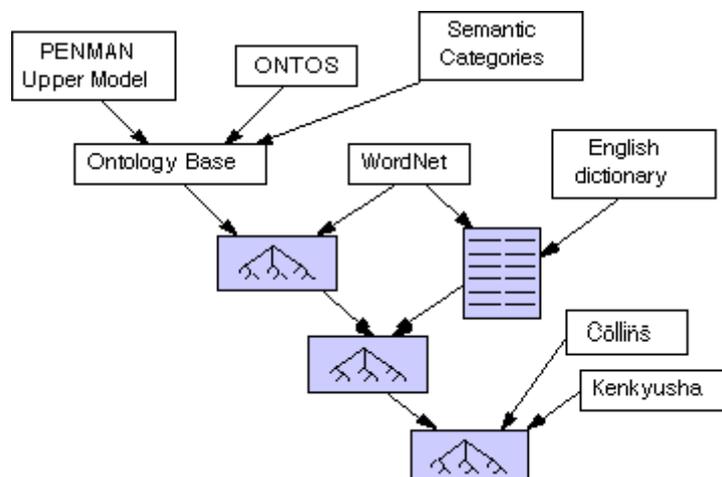
En paralelo al desarrollo de la ontología *Mikrokosmos* también se ha construido un lexicón en español de varios miles de palabras. Estas palabras cubren una amplia variedad de categorías, aunque la gran mayoría pertenecen al dominio de las fusiones y de la adquisición de compañías.

### 2.5.3.5. SENSUS

La última ontología lingüística mencionada por Gómez y otros (2004: 84) en su clasificación es *SENSUS*<sup>27</sup> (Swartout y otros, 1996), que es una ontología basada en el lenguaje natural desarrollada por el grupo de lenguaje natural *ISI* para ofrecer una amplia estructura conceptual para trabajar en traducción automática.

*SENSUS* contiene 70.000 nodos que representan los objetos, entidades, cualidades y relaciones más comunes. La ontología ofrece una base conceptual jerárquicamente estructurada (Knight y Luck, 1994). La parte superior de la ontología es llamada ontología base y consiste en aproximadamente 400 unidades que representan generalizaciones esenciales para el procesamiento lingüístico que tiene lugar en la traducción. La parte central de la ontología ofrece un marco para un modelo del mundo genérico y contiene unidades que representan muchos sentidos del mundo en inglés. La parte inferior es la más específica y ofrece puntos de unión para distintas lenguas.

*SENSUS* es una extensión y reorganización de *WordNet*. En el nivel alto se añadieron nodos del *Penman Upper Model*, *ONTOS* (una ontología lingüística de alto nivel) y las categorías semánticas se tomaron de un diccionario. Como resultado se obtuvo la ontología base. Después, a la ontología base se le fusionó *WordNet*, también de forma manual, y se utilizó una herramienta de fusión para fusionar *WordNet* con un diccionario inglés. Finalmente, como apoyo a la traducción automática, se amplió el resultado de la fusión con entradas en español y japonés de los diccionarios *Collins Spanish/English dictionary* y *Japanese/English dictionary*.



<sup>27</sup> <http://www.isi.edu/natural-language/projects/ONTOLOGIES.html>

Figura 2.23. La estrategia de fusión de *SENSUS* (Swartout y otros, 1996).

### **2.5.3.6. DOLCE**

A la clasificación de ontologías lingüísticas de Gómez y otros (2004) añadimos la ontología *DOLCE*<sup>28</sup> (Gangemi y otros, 2002). El acrónimo *DOLCE* significa “ontología descriptiva para la ingeniería lingüística y cognitiva” (*descriptive ontology for linguistic and cognitive engineering*). Es una ontología de alto nivel que funciona como el primer módulo de la librería de de ontologías fundacionales desarrollada dentro del proyecto *WonderWeb*<sup>29</sup> (*WFOL*). En contraste con las ontologías de semántica superficial, que tienen una estructura terminológica mínima (a menudo meras taxonomías), el principal propósito de las ontologías fundacionales es la negociación del significado. El propósito fundamental de *Wonderweb* es la creación de una librería de ontologías fundacionales que reflejen las diferentes elecciones ontológicas. Así, la librería conforma una red de módulos semánticos diferentes pero relacionados entre sí aplicables a la Web Semántica de acuerdo a sus supuestos ontológicos.

El propósito fundamental de *DOLCE* era la creación de un punto inicial de comparación de la relación con otros futuros módulos de la librería y también para clarificar los supuestos ontológicos ocultos en las ontologías existentes o en los recursos lingüísticos como *WordNet*. Como refleja su acrónimo, tiene un claro propósito cognitivo ya que pretende reflejar las categorías ontológicas que surgen en el lenguaje natural y el sentido común. Las categorías que ofrece son artefactos cognitivos dependientes de la percepción humana en última instancia, improntas culturales y convenciones sociales (un tipo de metafísica “cognitiva”). Las categorías que presenta describen entidades de forma “post-hoc”, que más o menos reflejan la estructura superficial del lenguaje y la cognición.

*DOLCE* es una ontología de particulares en el sentido de que su dominio de discurso se restringe a éste. La distinción ontológica fundamental entre “universales” y “particulares” puede entenderse como que los “particulares” son entidades que no tienen instancias y los “universales” son entidades que sí tienen instancias. Las propiedades y

---

<sup>28</sup> <http://www.loa.istc.cnr.it/DOLCE.html>

<sup>29</sup> <http://wonderweb.semanticweb.org/deliverables/D17.shtml>

las relaciones, que en lenguaje lógico se corresponden con las predicaciones, se consideran “universales”. Por tanto en *DOLCE* la ontología de universales está claramente diferenciada de la de particulares, aunque los “universales” también aparecen en la ontología de particulares en la medida en que se utilizan para organizarlos y caracterizarlos (simplemente como no están en el dominio de discurso no están sujetos a ser organizados y caracterizados). En la figura 24 aparece la ontología de universales.

*DOLCE* se basa en la distinción fundamental entre entidades “duraderas” y “circunstanciales” que tiene que ver con su duración en el tiempo. Las entidades “duraderas” están presentes en su totalidad, ya que todas sus partes están presentes en cualquier momento. Por otro lado, las entidades “circunstanciales” están parcialmente presentes en el sentido de que algunas de sus partes temporales pueden no estar presentes. Por ejemplo, el periódico que estoy leyendo aquí y ahora está presente en su totalidad, mientras que algunas partes temporales de mi lectura ya no están presentes. Las entidades “duraderas” están presentes a lo largo del tiempo mientras que las entidades “circunstanciales” ocurren en un momento concreto del tiempo. La principal relación entre ambos es la de participación: una entidad duradera coexiste en el tiempo como parte de una entidad circunstancial. Por ejemplo, una persona sería una entidad duradera y una discusión sería circunstancial. La vida de una persona es también circunstancial, ya que una persona participa en ella durante toda su duración (Gangemi y otros, 2002).

La taxonomía de las categorías más básicas de particulares en *DOLCE* aparece en la figura 2.24 a continuación. Se las considera propiedades rígidas de acuerdo a la metodología *OntoClean* que hace énfasis en la importancia de centrarse en estas propiedades primero.

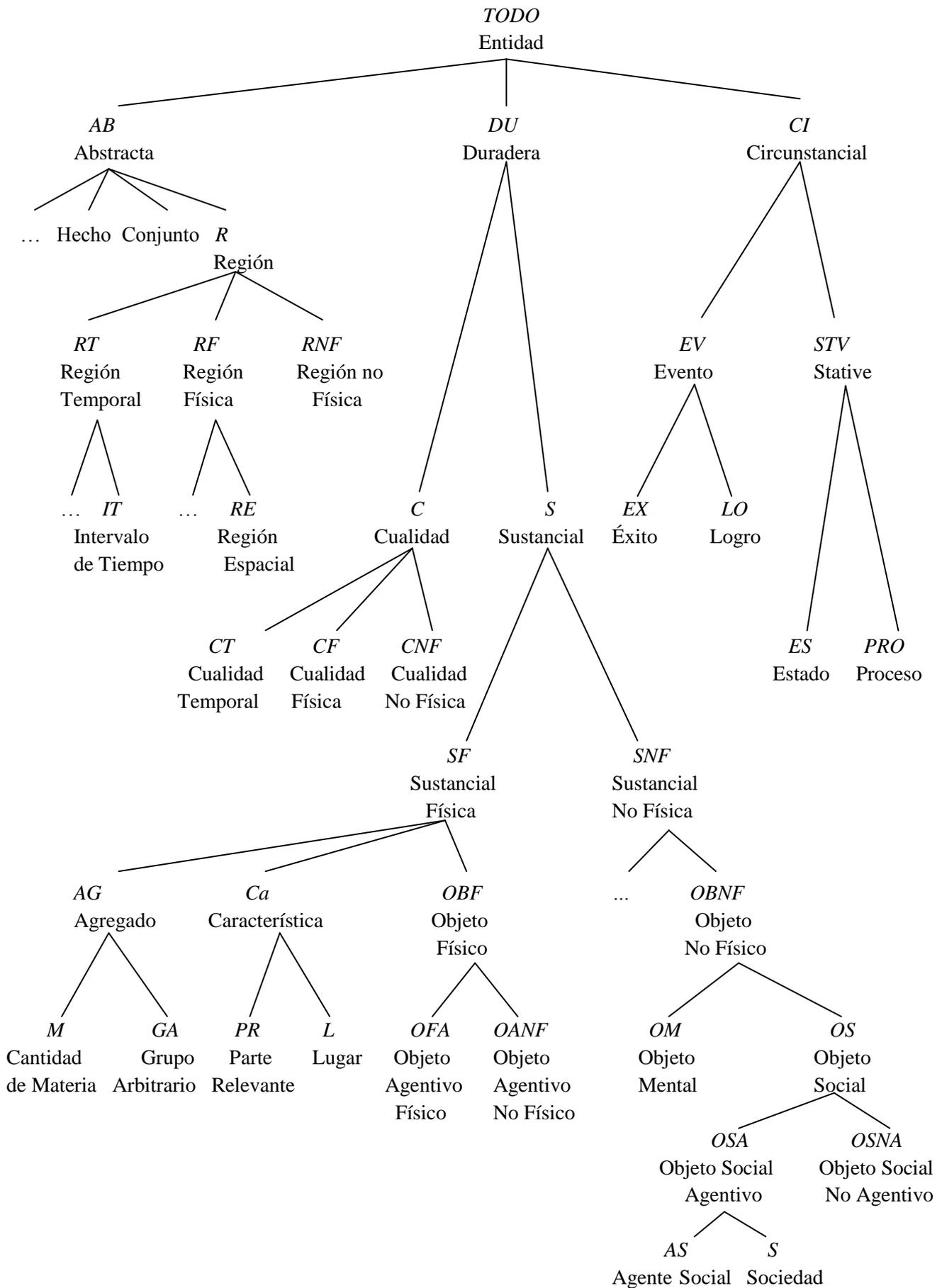


Figura 2.24. La taxonomía de categorías básicas de particulares en *DOLCE* (adaptada de Gangemi y otros, 2002).

#### 2.5.4. Ontologías de dominio

La última tipología de ontologías de la clasificación de Gómez y otros (2004: 85-105) son las ontologías de dominio. De acuerdo a estos autores, las ontologías de dominio son vocabularios reutilizables de los conceptos de un dominio y sus relaciones, de las actividades típicas de ese dominio y de las teorías y principios elementales que gobiernan ese dominio. Entre las ontologías de dominio analizadas por estos autores están las de comercio electrónico, de medicina, de ingeniería, de empresa, de química, y de gestión de conocimiento, a las que añadiremos también las ontologías legales que presentamos a continuación.

El desarrollo de Internet y de las tecnologías relacionadas con la red de redes ha dado lugar a muchas aplicaciones de comercio electrónico (Fensel, 2001; Berners-Lee, 1999). Aunque la tecnología no es el único factor clave en el desarrollo de las actuales aplicaciones electrónicas, ya que el comercio electrónico en general y las actividades entre negocios en particular requieren de una comunicación efectiva entre las máquinas. Como consecuencia se desarrollaron varios estándares e iniciativas para facilitar el intercambio de información entre clientes y empresas y entre varias empresas al ofrecer marcos para identificar los productos y servicios en mercados globales.

##### 2.5.4.1. De comercio electrónico

Dentro del dominio del comercio electrónico se encuentran las ontologías *UNSPSC*<sup>30</sup>, *NAICS*<sup>31</sup>, *SCTG*<sup>32</sup>, *e-class*<sup>33</sup> y *RosettaNet*<sup>34</sup>. Estas propuestas se han realizado por el acuerdo de un grupo amplio de personas y se han codificado utilizando distintos lenguajes y formatos de computación. Se trata de taxonomías de conceptos y de sus relaciones, por lo que se trata ontologías de semántica superficial.

---

<sup>30</sup> <http://www.unspsc.org>

<sup>31</sup> <http://www.naics.com>

<sup>32</sup> [http://www.bts.gov/publications/commodity\\_flow\\_survey/methods\\_and\\_limitations/commodity\\_classification\\_in\\_1997/classification.html](http://www.bts.gov/publications/commodity_flow_survey/methods_and_limitations/commodity_classification_in_1997/classification.html)

<sup>33</sup> <http://www.eclass.de/>

<sup>34</sup> <http://www.rosettanel.org>

La *United Nations Standard Products and Services Codes (UNSPSC)* ha sido creada por el programa de desarrollo de Naciones Unidas y Dun y Bradstreet. Se trata de un estándar multilingüe<sup>35</sup> de códigos de productos que clasifica los productos y servicios generales y está diseñado para facilitar el comercio electrónico a través del intercambio de descripciones de productos.

El sistema de códigos de *UNSPC* se organiza en una taxonomía de cinco niveles de productos. Cada nivel contiene un valor numérico de dos caracteres y una descripción textual. Los niveles se definen así:

- Segmento. Se agregan familias de manera lógica para su análisis.
- Familia. Grupo de categorías de productos interrelacionadas.
- Clase. Grupo de productos que comparten un mismo uso o función.
- Producto. Grupo de productos o servicios que se pueden sustituir.
- Función comercial. La función de una organización en apoyo del producto. Este nivel no se suele utilizar.

La *NAICS (North American Industry Classification System)* fue creada por la Oficina del Censo de EEUU en cooperación con el Comité de clasificación de EEUU, Estadística de Canadá y el Instituto Nacional de Estadística, Geografía e Informática (*INEGI*) de Méjico. Clasifica los productos y servicios en general y se utiliza en EEUU, Canadá y Méjico.

Los productos se clasifican con un código de seis dígitos, en contraste con el código de cuatro dígitos *SIC (Standard Industrial Classification)*. El código *NAICS* incluye a un gran número de sectores y permite más flexibilidad para diseñar subsectores. El acuerdo *NAICS* internacional establece sólo los cinco primeros dígitos del código. El sexto, si utilizado, es para identificar las subdivisiones de las industrias de *NAICS* que consideran las necesidades del usuario en países individuales. Así, los códigos americanos de seis dígitos pueden diferir de sus correspondientes códigos en Canadá o Méjico, pero están estandarizados hasta el nivel de los cinco dígitos. Su estructura general es la siguiente:

XX Sector industrial (hasta 20 sectores más que en *SIC*)

---

<sup>35</sup> *UNSPSC* está disponible en 14 lenguas: inglés, húngaro, francés, alemán, español, italiano, japonés, coreano, holandés, chino, portugués, danés, noruego y sueco.

- XXX Subsector industrial
- XXXX Grupo industrial
- XXXXX Industria
- XXXXXX Específico de EEUU, Canadá o Méjico

La *SCTG (Standard Classification of Transported Goods)* fue patrocinada por la oficina de estadística del transporte, *Bureau of Transportation Statistics (BTS)*. Es una clasificación producto de la recopilación de datos de la encuesta de circulación de productos, *Comodity Flow Survey (CFS)*. Fue desarrollada por el departamento de transporte americano, *Department of Transportation's (DOT)*, el *Volpe National Transportation Systems Center (Volpe Center)*, la *Standard and Transportation Divisions of Statistics* de Canadá, la *US Bureau of the Census (BOC)* y la *US Bureau of Economic Analysis (BEA)*.

Tiene cuatro niveles de clasificación, cada uno de los cuales sigue dos principios fundamentales. Primero, cada nivel cubre el universo de mercancías transportables, y segundo, cada categoría en cada nivel es mutuamente exclusiva. Su estructura general es la siguiente:

- XX Categoría del producto
- XXX Producto o grupos de productos (diferentes en EEUU y Canadá)
- XXXX Análisis del transporte de mercancías nacionales
- XXXXX Datos de movimiento de mercancías

El primer nivel de *SCTG* (dos dígitos) consiste en 43 categorías de productos. Estas categorías se diseñaron para enfatizar en vínculo entre la industrias y sus productos. El segundo nivel (tres dígitos) está diseñado para ofrecer datos para la comparativa entre los productos canadienses y los americanos. El tercer nivel (cuatro dígitos) está diseñado para ofrecer datos para el análisis del transporte de mercancías nacionales. El cuarto nivel (cinco dígitos) está diseñado para ofrecer categorías para recoger (y potencialmente informar) de los datos del movimiento de mercancías. Los códigos de los productos en este nivel han sido diseñados para crear categorías estadísticamente significativas para el análisis de transporte.

La *e-cl@ss* es una iniciativa alemana para crear una clasificación estándar de materiales y servicios para el intercambio de información entre proveedores y clientes, y es utilizada por compañías como *BASF*, *Bayer*, *Volkswagen-Audi* o *SAP*. Consiste en cuatro niveles de conceptos (llamados “clases materiales”) con un código numérico similar al utilizado por *UNSPSC*, donde cada nivel añade dos dígitos al nivel anterior. Los cuatro niveles son: segmento, grupo principal, grupo, y clase de producto. Dentro de la misma clase de producto podemos tener varios productos (en este sentido varios productos pueden compartir el mismo código). A diferencia de *UNSPSC*, ofrece un conjunto de atributos para cada producto que aparece como una hoja en la clasificación. La herramienta de búsqueda de *e-cl@ss*<sup>36</sup>, que está disponible en línea, permite encontrar términos con una interfaz en varias lenguas (alemán, español, inglés y checo). De hecho, los términos encontrados se presentan en cualquiera de estas lenguas. La clasificación puede descargarse desde esa misma dirección web.

La clasificación *RosettaNet* ha sido creada por *RosettaNet*, que es un consorcio autofinanciado sin ánimo de lucro de unas 400 empresas de componentes electrónicos, tecnología de la información, fabricación de semiconductores y compañías proveedoras de soluciones. Comenzó en la industria de las tecnologías de la información y ahora se encuentra en desarrollo en las áreas de la automoción, la electrónica de consumo y las industrias de la telecomunicación.

La clasificación de *RosettaNet* no utiliza un sistema numérico sino que está basada en los nombres de los productos que define. Su clasificación está relacionada con la clasificación de *UNSPSC*, por lo que ofrece el código de esta última para cada producto definido. La clasificación sólo tiene dos niveles en la taxonomía de productos:

- La categoría RN. Un grupo de productos, tales como productos de vídeo.
- El producto RN. Un producto específico, tal como una tarjeta de TV o de radio.

---

<sup>36</sup> <http://www.eclasscontent.com/index.php?language=en>

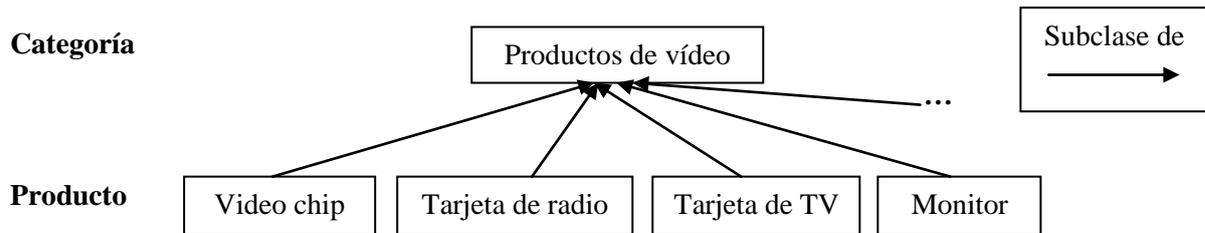


Figura 2.25. Muestra parcial de la clasificación de *RosettaNet* (adaptada de Gómez y otros, 2004: 91).

La clasificación de *RosettaNet* consiste en 14 categorías y en unos 150 productos. Es más específica que la clasificación de *UNSPSC*. La figura 25 ofrece una pequeña sección de la clasificación relacionada con los productos de vídeo para equipos informáticos.

Como señalan Gómez y otros (2004: 91) estas cinco ontologías de comercio electrónico ofrecen un gran nivel de solapamiento puesto que no hay consenso en la clasificación de productos. De hecho, un producto o servicio pueden clasificarse en lugares diferentes en cada ontología. La proliferación de distintas ontologías revela que no se ha alcanzado un consenso en cuanto a los sistemas de codificación, el nivel de detalle o la granularidad, lo cual ofrece un obstáculo para la interoperabilidad de las distintas aplicaciones puesto que siguen estándares diferentes. Por ejemplo, una aplicación que utiliza un código *UNSPSC* no puede interoperar con una aplicación que sigue un sistema de códigos de *e-cl@ss*. Sin embargo, hay contribuciones que pretenden alinear estas iniciativas a través de establecer vínculos ontológicos entre los estándares existentes (Bergamaschi y otros, 2001; Corcho y Gómez, 2001; Gordijn y otros, 2001).

#### 2.5.4.2. Médicas

El siguiente grupo de ontologías de dominio lo conforman las ontologías médicas. Se desarrollan para resolver problemas tales como la demanda de reutilización y para compartir datos de los pacientes, la transmisión de esos datos o la necesidad de criterios basados en la semántica y para propósitos estadísticos. La comunicación no ambigua de conceptos médicos complejos y detallados es una característica crucial en sistemas actuales de información médica. En estos sistemas deben interactuar varios agentes entre sí para compartir sus resultados, y así deben utilizar terminología médica con un significado claro y no confuso.

*GALEN*<sup>37</sup> (Rector y otros, 1995), desarrollada por la organización sin ánimo de lucro *OpenGALEN*, es una terminología clínica representada en lenguaje formal y médico *GRAIL* (Rector y otros, 1997). Este lenguaje fue especialmente desarrollado para especificar restricciones utilizadas en el dominio médico. La intención era que *GALEN* fuera utilizada con diferentes lenguas naturales e integrada con diferentes esquemas de códigos. Se basa en el modelo semántico de terminología clínica conocido como el modelo *GALEN CODING REFERENCE (CORE)*. La figura 2.26 muestra la ontología de alto nivel *GALEN CORE*.

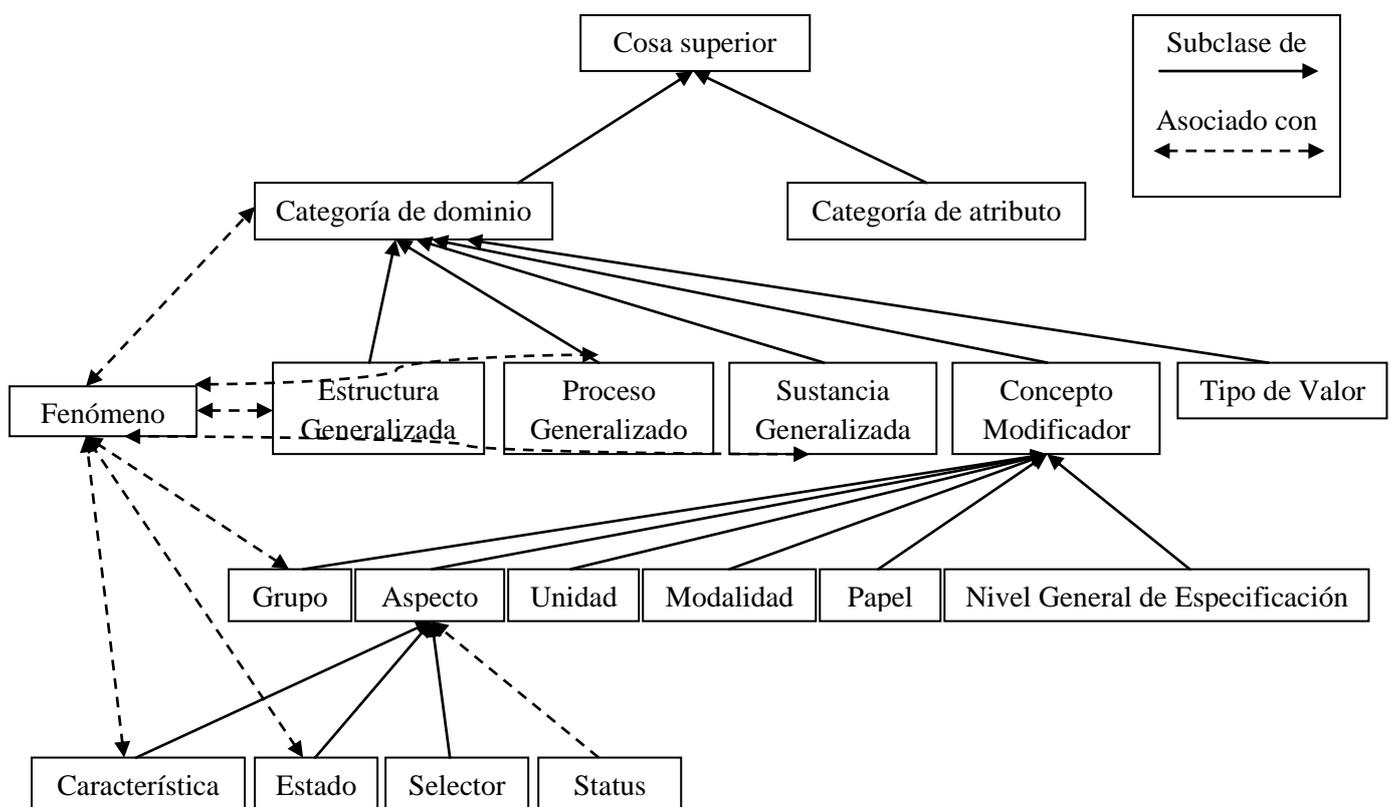


Figura 2.26. La ontología de alto nivel *GALEN CORE* (adaptada de Gómez y otros, 2004: 93).

Esta ontología establece cuatro categorías generales que son subclases de la categoría de dominio:

- Estructuras (Estructura generalizada), que son cosas abstractas o físicas con partes que son independientes del tiempo, tales como los microorganismos, protocolo o corazón.

<sup>37</sup> <http://www.opengalen.org>

- Sustancias (sustancia generalizada), que son cosas continuas abstractas o físicas que son independientes del tiempo, tales como bilis, drogas o radiación.
- Procesos (proceso generalizado), que son cambios que ocurren a lo largo del tiempo, tales como irradiación, acto clínico o respiración.
- Modificadores (concepto modificador), que refina o modifica el significado de las otras tres categorías, tales como diabetes severa. Se consideran los siguientes tipos de modificadores: modificadores de aspecto (clasificados en característica, estado, selector y status), unidad, modalidad, papel, nivel general de especificación y grupo.

Finalmente, la ontología de alto nivel *GALEN CORE* define relaciones entre conceptos que pertenecen a categorías generales. A estas relaciones se las llama “atributos” (“atributo de dominio”) y se dividen en dos tipos: “atributos constructivos”, que unen procesos, estructuras y sustancias conjuntamente; y “atributos modificadores”, que unen procesos, estructuras y sustancias a modificadores.

La ontología *UMLS*<sup>38</sup> (*Unified Medical Language System*) es una ontología desarrollada por la Librería Nacional de Medicina de EEUU, una gran base de datos diseñada para integrar un gran número de datos biomédicos recogidos de varias fuentes tales como vocabularios clínicos o estándares (*MeSH*, *RxNorm*, *SNOMED CT*, etc) para permitir su interoperabilidad con sistemas de computación. Se puede utilizar *UMLS* para mejorar o desarrollar aplicaciones tales como registros sanitarios, herramientas de clasificación, diccionarios o traductores.

Una utilidad de *UMLS* es vincular información sanitaria, términos médicos, nombres de medicamentos y códigos de facturación a través de varios sistemas de computación. Algunos ejemplos son los siguientes:

- Vincular términos y códigos entre el médico, la farmacia y el seguro médico.
- Coordinación del cuidado del paciente entre varios departamentos de dentro del hospital.

---

<sup>38</sup> <http://www.nlm.nih.gov/research/umls/>

*UMLS* tiene otros muchos usos que incluyen recuperación de información a través de un buscador, minería de datos, estadísticas de informes de la sanidad pública americana y búsqueda de terminología.

*UMLS* funciona con tres herramientas, llamadas fuentes de conocimiento:

- *Metathesaurus*: con términos y códigos de diversos vocabularios que incluyen *CPT*®, *ICD-10-CM*, *LOINC*®, *MeSH*®, *RxNorm*, y *SNOMED CT*®. Contiene información biomédica sobre cada uno de los términos incluidos en *UMLS*. Si un término aparece en varias fuentes, lo que es común, se creará un concepto en *UMLS* con un nombre de término preferido asociado a él. La fuente original de información de los términos tales como la definición y fuente se incluyen en el concepto y también se especifican las propiedades semánticas, tales como conceptos sinónimos, hermanos y padres, o las relaciones entre los términos.
- *Semantic Network*: categorías amplias (tipos semánticos) y sus relaciones (relaciones semánticas). Es una ontología de alto nivel de conceptos biomédicos y sus relaciones. La figura 2.27. muestra una vista parcial de esta ontología. La red semántica no se deriva de las fuentes biomédicas integradas en *UMLS* sino que ha sido creada como parte de *UMLS* para ofrecer una estructura consistente o categorización en la que se incluyen los conceptos del *Metathesaurus*. Cada concepto del *Metathesaurus* se incluye a un concepto o conceptos de la red semántica. Así, la red semántica se introduce para resolver la heterogeneidad entre las fuentes de *UMLS*.
- *Specialist Lexicon* y *Lexical Tools*: contienen información semántica sobre términos biomédicos para ser utilizados con aplicaciones de PLN.

*UMLS* utiliza el *Semantic Network* y las *Lexical Tools* para construir el *Metathesaurus*.

La producción del *Metathesaurus* incluye:

- Procesamiento de términos y códigos a través de *Lexical Tools*.
- Agrupación de términos sinónimos y conceptos.
- Categorización de conceptos en tipos semánticos de la red semántica.
- Incorporación de relaciones y atributos de los vocabularios.
- Obtención de datos en un formato común.



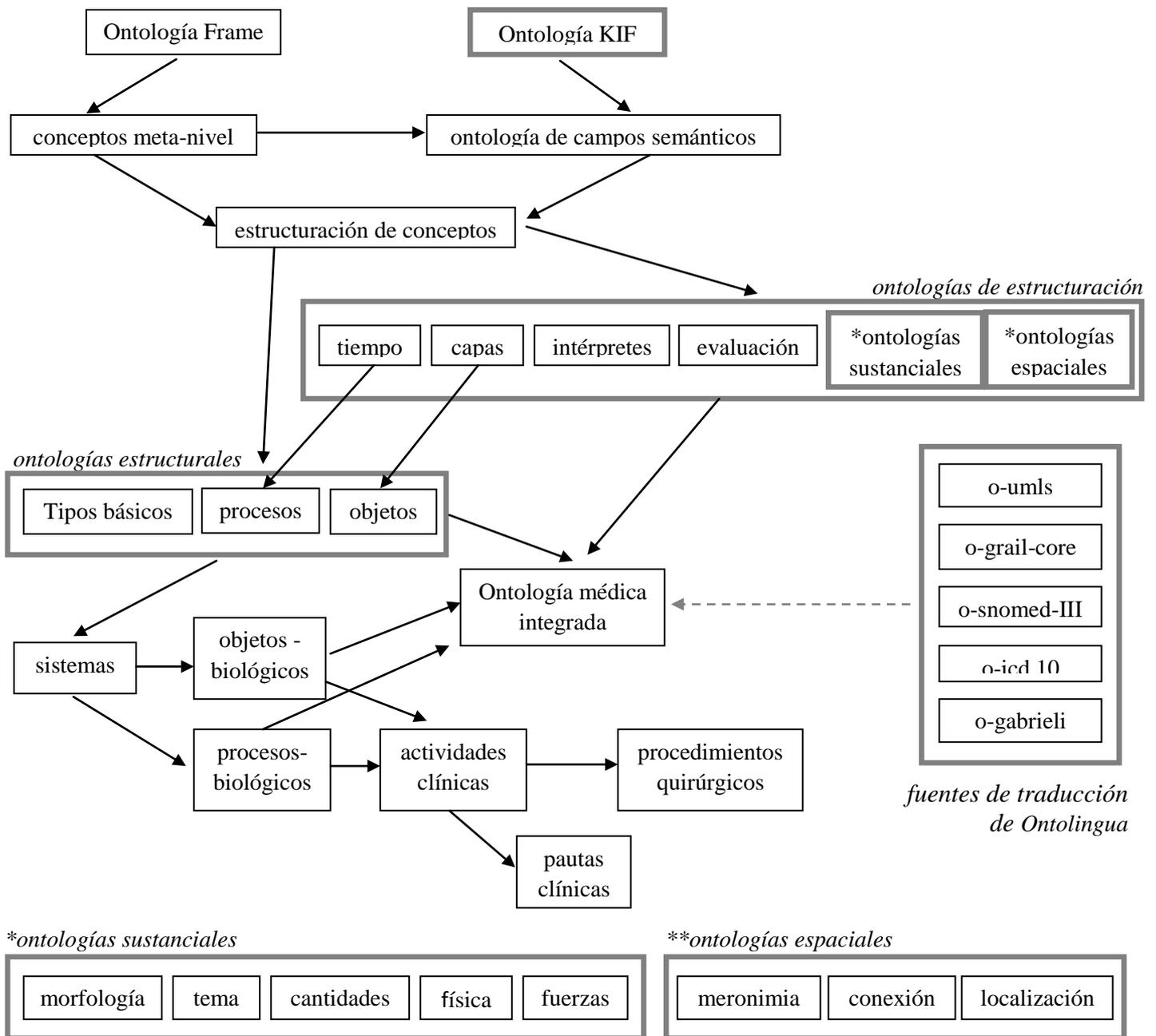


Figura 2.28. La red de inclusión de la librería de ontologías ON9 (adaptación de Gangemi y otros, 1998: 17).

Para vincular estas ontologías con la librería de ontologías general se han definido las ontologías “estructuración de conceptos”, “conceptos meta-nivel” y la “ontología de campos semánticos”. El conjunto de “ontologías estructurales” y de “ontologías de estructuración” contienen ontologías generales. En concreto, la “ontología médica integrada” incluye todas las ontologías generales utilizadas para recoger las ontologías terminológicas de los cinco sistemas terminológicos.

En nuestro país, el proyecto interdisciplinar *Oncoterm*<sup>40</sup> ofrece un completo repositorio de información sobre la terminología asociada al cáncer: enfermedades, medicinas, tratamientos y ámbitos relacionados. Se creó con el propósito de elaborar un sistema de información específico del subdominio biomédico de la oncología.

Como primer paso se elaboró una base de datos terminológica, para lo que se utilizó tanto la información extraída de diccionarios como la de textos especializados. Mediante la extracción de información conceptual de textos especializados y diccionarios médicos se construyó la estructuración conceptual inicial del dominio a través de la elaboración de jerarquías terminográficas, por la que los conceptos aparecen vinculados a una ontología. Así, la estructura conceptual subyacente sirve de nexo de unión entre los términos en las distintas lenguas. Los datos quedan representados y almacenados en un gestor de conocimiento terminológico llamado *Ontoterm*<sup>41</sup>, aplicación informática desarrollada por Moreno (2000). Para su estructuración conceptual se utilizó la ontología *Mikrokosmos* (Mahesh y Nirenburg, 1995; Viegas y otros, 1999) como recurso de representación del conocimiento (Moreno y Pérez, 2000).

La arquitectura de *Ontoterm* tiene dos módulos principales: un editor de ontologías y un editor de la base de datos terminológica. La estructura conceptual se construye en el editor de ontologías, de la que podemos ver un extracto en la figura 2.29.

---

<sup>40</sup> <http://www.ugr.es/~oncoterm/>

<sup>41</sup> <http://tecnolengua.uma.es/ontoterm/index.html>

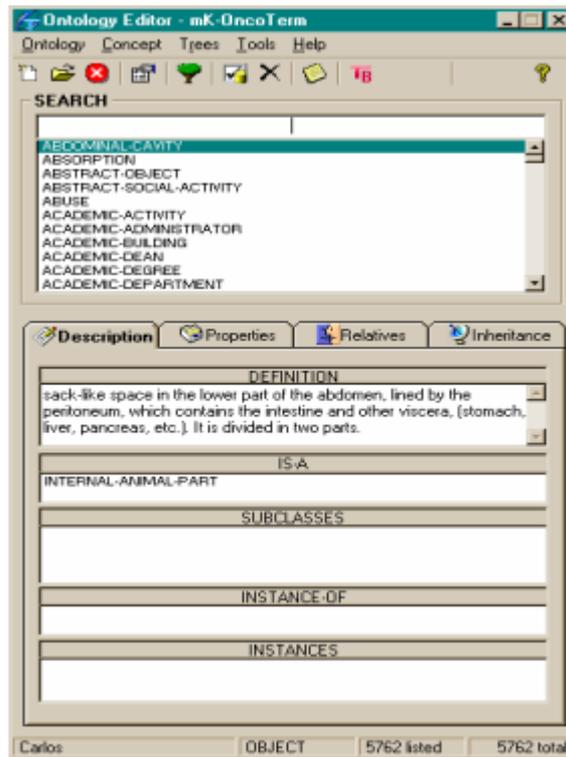


Figura 2.29. El editor de ontologías de *Ontoterm*, en Faber y otros (2001: 174).

En la figura vemos cómo el editor de ontologías muestra la lista de conceptos en orden alfabético. Cada uno aparece con una descripción, propiedades y relaciones con otros conceptos y su herencia. Es posible añadir nuevos conceptos a la ontología a través de un primer submódulo (figura 2.30) en el que se indica el concepto superordinado (*parent concept*) con lo que el concepto queda enmarcado dentro de la ontología.

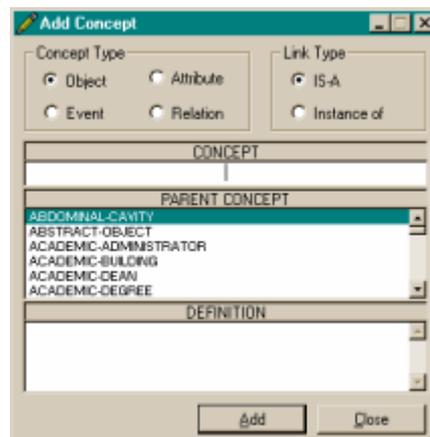


Figura 2.30. Submódulo para la integración de conceptos, en Faber y otros (2001: 174).

A través del segundo submódulo de *Ontoterm* podemos visualizar la jerarquía conceptual que aparece representada por medio de un árbol conceptual dividido en eventos (EVENT), objetos (OBJECT) y propiedades (PROPERTY) (figura 2.31).

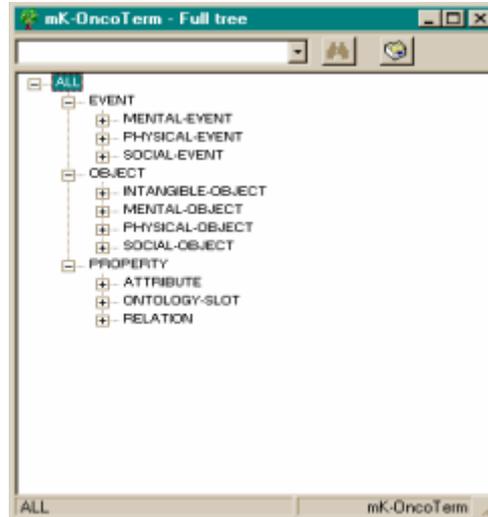


Figura 2.31. Submódulo de visualización de conceptos: conceptos genéricos, en Faber y otros (2001: 175).

El segundo módulo de *Ontoterm* es el editor de la base de datos terminológica, donde se elaboran las entradas terminográficas de los conceptos de la ontología. En este módulo, que complementa al editor de ontologías, se describe el término o términos que representan cada concepto mediante las categorías de datos de la norma ISO 12620 (cf. Faber y otros, 2001).

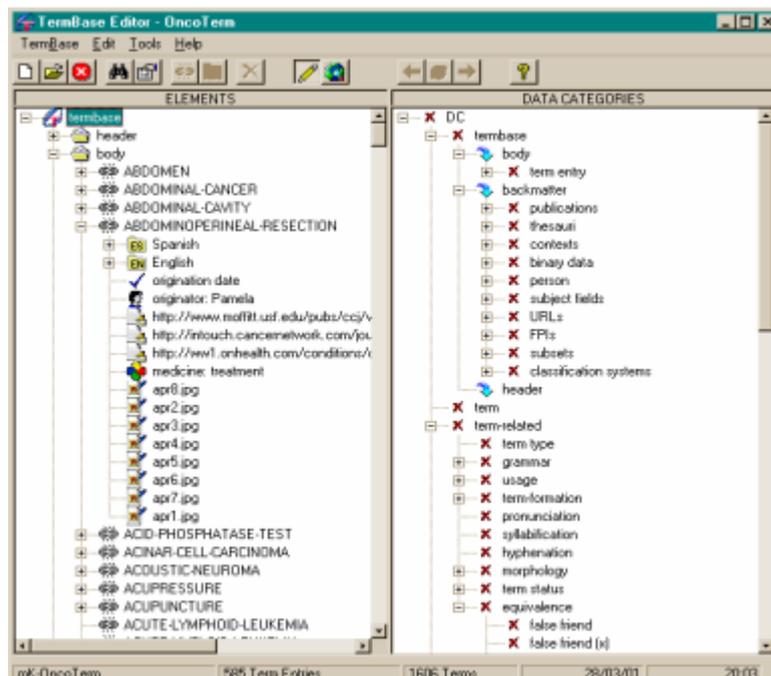


Figura 2.32. Editor de la base de datos terminológica, en Faber y otros (2001: 176).

También es posible representar la información de la base de datos terminológica en formato HTML, donde figuran todos los datos relevantes al concepto en cuestión.

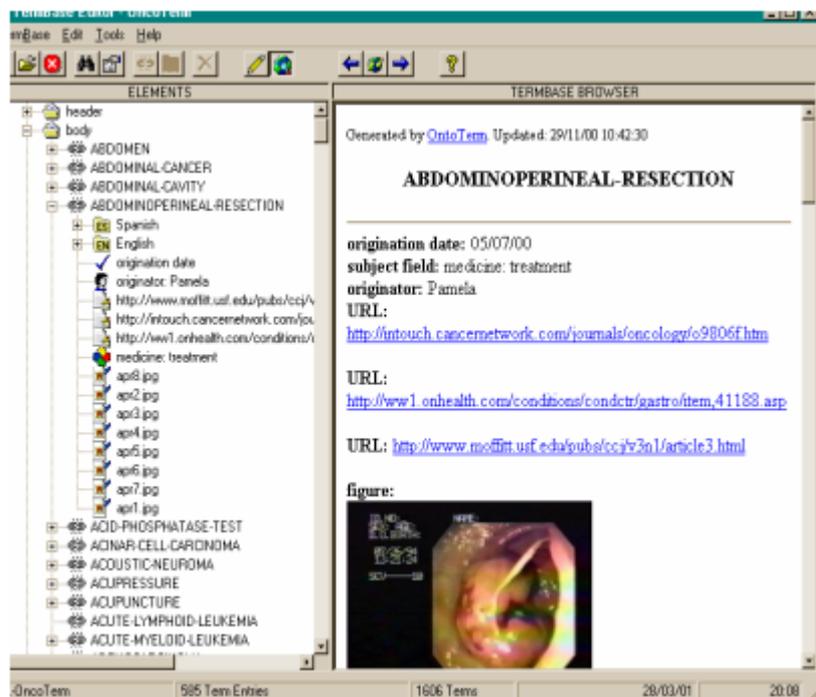


Figura 2.33. Representación de la información conceptual en HTML, en Faber y otros (2001: 177).

### 2.5.4.3. De empresa

El siguiente grupo de ontologías de dominio de nuestra clasificación lo conforman las ontologías de empresa. Normalmente son creadas para definir y organizar conocimiento relevante sobre actividades, procesos, organizaciones, estrategias y marketing, entre otros. El objetivo de este tipo de ontologías es que las empresas utilicen el conocimiento recogido en ellas. A continuación presentamos a modo de ejemplo dos ontologías: la ontología de empresa (*The Enterprise Ontology*) de Uschold y otros (1998) y la ontología *TOVE* (Gómez y otros, 2004: 98).

La *Enterprise Ontology*<sup>42</sup> se desarrolló dentro del marco del proyecto *Enterprise* por el *Artificial Intelligence Applications Institute (AIAI)* en la Universidad de Edimburgo con sus socios *IBM*, *Lloyd's Register*, *Logica UK Limited* y *Unilever*. El proyecto fue financiado por el departamento de comercio e industria británico bajo el programa de

<sup>42</sup> <http://www.aiai.ed.ac.uk/project/enterprise/enterprise/ontology.html>

integración de sistemas inteligentes (proyecto IED4/1/8032). Esta ontología contiene términos y definiciones implementadas en *Ontolingua* relevantes para las empresas. Consta de 92 clases, 68 relaciones, 7 funciones y 10 individuos. La figura 2.34 muestra una vista parcial de la taxonomía de clases.

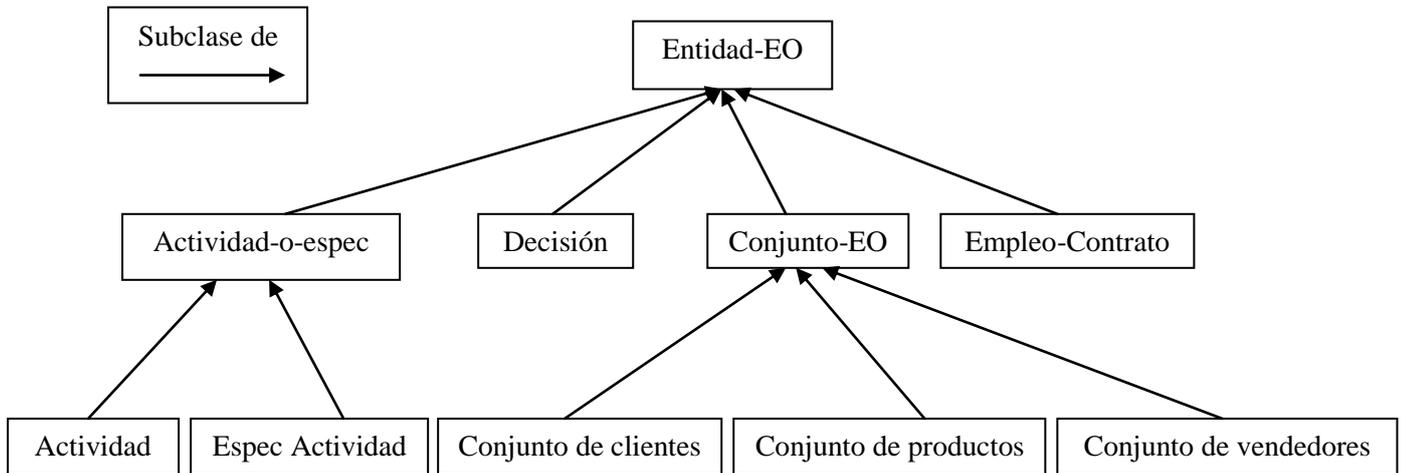


Figura 2.34. Vista parcial de la taxonomía de clases en la *Enterprise Ontology* (adaptado de Gómez y otros, 2004: 99).

Conceptualmente está dividida en cuatro secciones principales que se especifican a continuación:

- Actividades y procesos. “Actividad” es el término principal, cuyo propósito es capturar la noción de todo lo que implique “hacer”, particularmente cuando implica una acción. Está íntimamente relacionada con la idea del “hacedor”, ya sea una persona, unidad organizativa o máquina.
- Organización. Los conceptos centrales de la sección de “organización” son los conceptos de entidad legal y unidad organizativa. La entidad legal incluye a personas y corporaciones. Las entidades legales más grandes pueden poseer la totalidad de entidades legales más pequeñas. Una unidad organizativa puede ser grande y compleja e incluso trascender entidades legales. Las unidades organizativas grandes normalmente se conciben como formadas por unidades más pequeñas. La más pequeña se puede corresponder con una sola persona. Una máquina es una entidad no humana y no legal que puede realizar distintos papeles que de otro modo serían realizados por una persona o unidad organizativa (por ejemplo, desarrollar una actividad).
- Estrategia. El concepto central de esta sección es el “propósito”, que refleja o bien la idea de algo que un plan puede ayudar a conseguir, o bien algo de lo que una unidad organizativa puede ser responsable. De hecho incluye cualquier tipo

de propósito, ya sea en un nivel de organización y escala de tiempo que normalmente será llamada estratégica, o detallado y a corto plazo.

- Marketing. El concepto central de la sección de marketing es “venta”. Una “venta” es un acuerdo entre dos entidades legales para el intercambio de un producto a un precio de venta. Normalmente el producto es una mercancía o un servicio, y el precio de venta es monetario; sin embargo, puede haber otras posibilidades. Las entidades legales tienen los papeles de vendedor y cliente. Una venta puede haber sido concertada en el pasado, y una venta potencial futura se puede prever, aunque el producto real no sea identificado o ni siquiera exista.

Esta ontología ha tenido un papel relevante en el desarrollo de los fundamentos de creación de ontologías según la aproximación de Uschold y King (1995) que desarrollamos en la sección 2.6.

En cuanto a la otra ontología de empresa, el proyecto *TOVE*<sup>43</sup> (*Toronto Virtual Enterprise*) (Fox, 1992) es conducido por el laboratorio de integración de empresas, el *Enterprise Integration Laboratory (EIL)* en la Universidad de Toronto. Tiene como propósito crear un modelo de datos capaz de ofrecer una terminología compartida para la empresa que los agentes pueden entender y utilizar, definir el significado de cada término de forma precisa, implementar la semántica en un conjunto de axiomas que permiten a *TOVE* deducir automáticamente la respuesta de muchas preguntas de sentido común sobre la empresa, y definir una simbiología para describir un término o el concepto construido del mismo en un contexto gráfico.

Las ontologías *TOVE* están implementadas en los lenguajes *C++* en su parte estática y *Prolog* en los axiomas. La figura 2.35 muestra la estructura de las ontologías *TOVE*. Estas ontologías cubren actividades, estados, causalidad, tiempo, recursos, inventario, requisitos de orden y partes. También han axiomatizado las definiciones para porciones del conocimiento de la actividad, estado, tiempo, y recursos. Los axiomas están implementados en *Prolog* y ofrecen respuestas a cuestiones de sentido común a través de un procesamiento de preguntas deductivo. Esta ontología se ha probado como

---

<sup>43</sup> <http://www.eil.utoronto.ca/tove/toveont.html>

fundamento experimental en el método de Grüninger y Fox (1995) de desarrollo de ontologías que se describe en la sección 2.6.

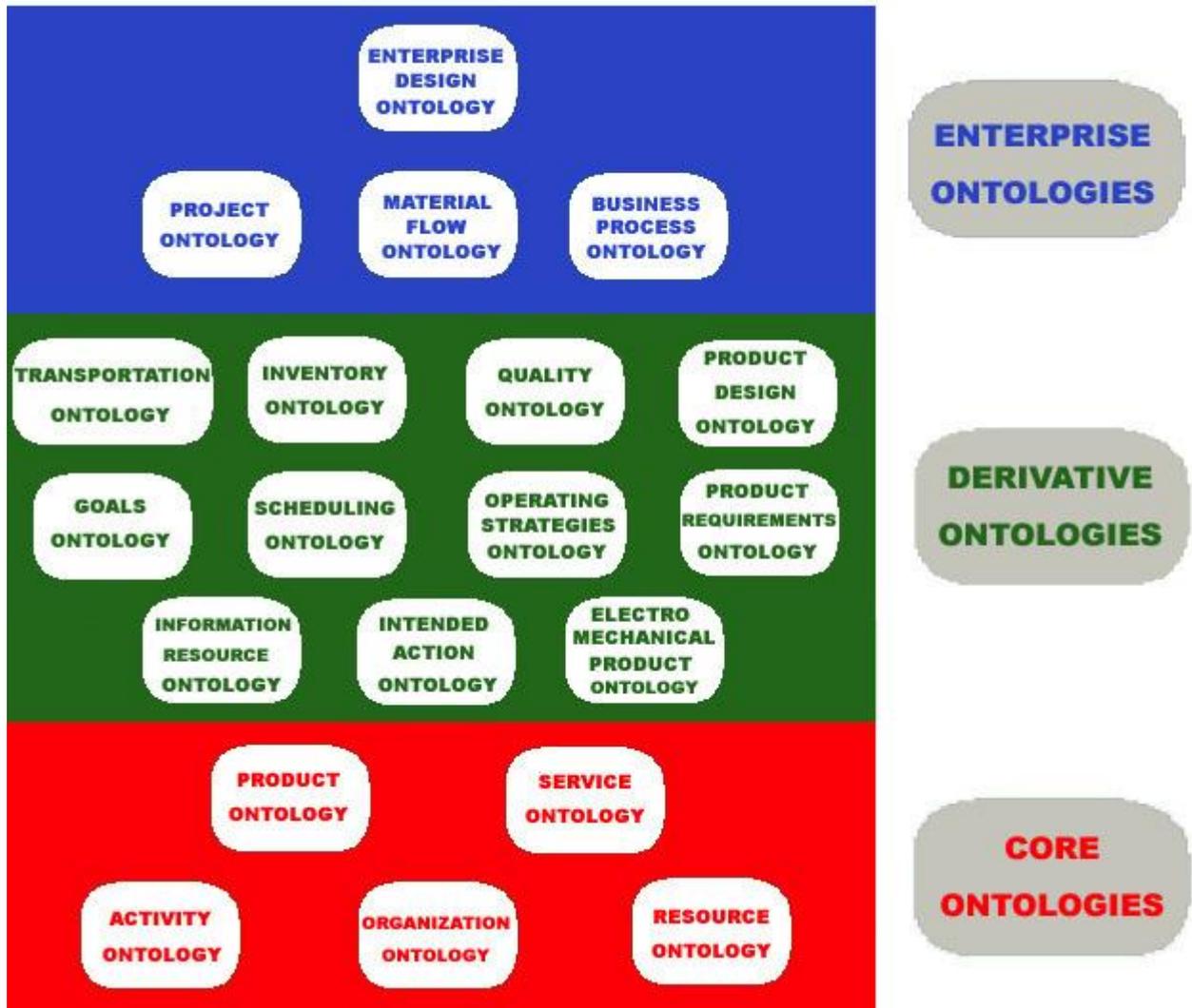


Figura 2.35. Estructura de las ontologías TOVE.

#### 2.5.4.4. De gestión del conocimiento

Especial atención merecen las ontologías de gestión del conocimiento. Los objetivos de la gestión del conocimiento en una organización son promover el crecimiento del conocimiento, la comunicación y la preservación del conocimiento (Steels, 1993). Para conseguir estos objetivos pueden crearse memorias corporativas. Una memoria corporativa es una representación de conocimiento e información en una organización explícita, incorpórea y persistente (van der Spek y otros, 1996). De acuerdo a Dieng-Kuntz y colegas (1998), una memoria corporativa puede construirse de acuerdo a varias técnicas que pueden combinarse: basadas en documentos, basadas en el conocimiento,

basadas en casos, basadas en aplicaciones de trabajo en grupo, basadas en volumen de trabajo y distribuidas. Las ontologías se incluyen entre las técnicas basadas en el conocimiento para construir memorias corporativas.

Básicamente, Abecker y otros (1998: 44) distinguen tres tipos de ontologías de gestión del conocimiento:

- Ontologías de la información. Describen diferentes tipos de fuentes de información, su estructura, permisos de acceso y propiedades de formato.
- Ontologías de dominio. Modelan el contenido de las fuentes de información.
- Ontologías de empresa. Modelan el contexto de una organización, proceso de negocio u organización de la empresa (como se ha mencionado más arriba).

La figura 2.36 muestra un ejemplo de relaciones existentes entre las mencionadas ontologías. Las ontologías presentadas se han desarrollado para las memorias corporativas del laboratorio de inteligencia artificial de la Universidad Politécnica de Madrid (UPM). Hay dos ontologías de dominio: *Hardware & Software* (que modela el equipo de hardware del laboratorio y el software instalado en él) y *Documentation* (que modela todos los documentos generados en el laboratorio: publicaciones, tesis, faxes, etc). Hay cuatro ontologías de empresa: *Organization* y *Groups* (que modelan la estructura del laboratorio y su organización en diferentes grupos de investigación), *Projects* (que modela los proyectos que se desarrollan en el laboratorio), y *Persons* (que modela los miembros del laboratorio: personal de la investigación, personal administrativo, estudiantes, etc). Tres de estas ontologías también son ontologías de la información: *Documentation*, *Projects* y *Persons*, ya que describen las fuentes de información de la memoria corporativa y almacenan información.

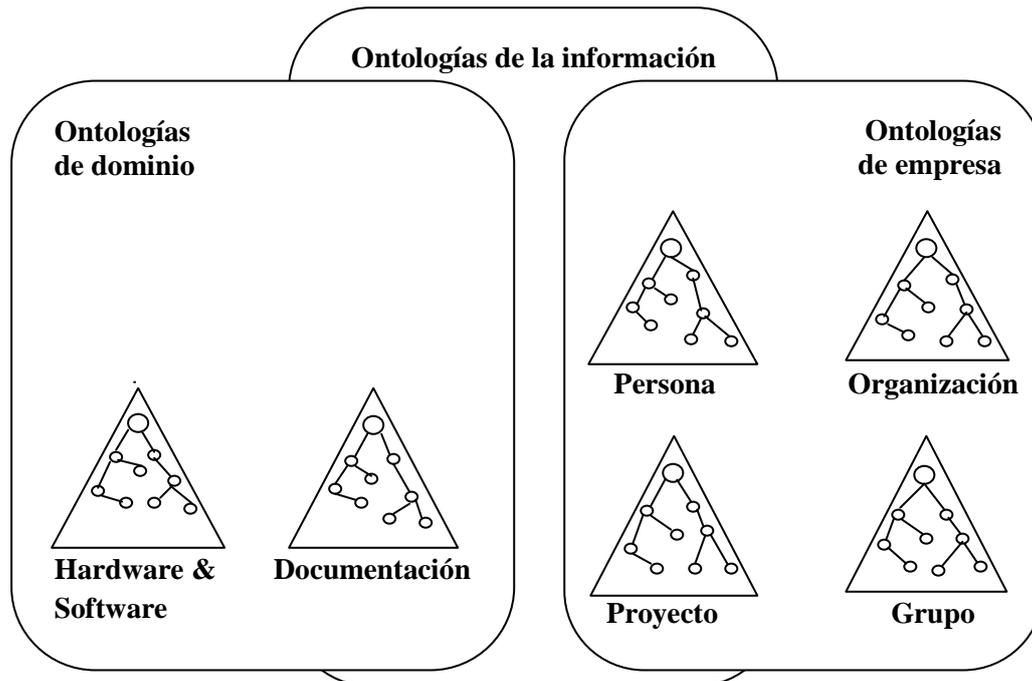


Figura 2.36. Las ontologías de información, dominio y empresa en una memoria corporativa (adaptado de Gómez y otros, 2004: 103).

Otro ejemplo de ontologías de gestión del conocimiento son las ontologías  $(KA)^2$  (Decker y otros, 1999). Se construyeron dentro del marco de la iniciativa de anotación del conocimiento de la comunidad de adquisición del conocimiento (Benjamins y otros, 1999), también conocida como iniciativa  $(KA)^2$ . Surgió con el propósito de modelar la comunidad de adquisición del conocimiento con las ontologías creadas por 15 grupos de personas en diferentes localizaciones. Cada grupo se centró en un tema particular de las ontologías  $(KA)^2$  como métodos de resolución de problemas, ontologías, etc. Como resultado se obtuvieron siete ontologías relacionadas entre sí: una ontología de organización, una ontología de proyecto, una ontología de persona, una ontología de publicación, una ontología de eventos, una ontología de temas de investigación y una ontología de productos de investigación. Estas ontologías conformaron la base para anotar documentos WWW de la comunidad de adquisición del conocimiento, y así permitir el acceso inteligente a estos documentos Web.

La primera versión de  $(KA)^2$  se construyó en el lenguaje *FLogic* (Kifer y otros, 1995), que se describe en la sección 2.6. Se tradujeron estas ontologías en *FLogic* a *Ontolingua* con traductores *ODE* (Blázquez y otros, 1998) para hacerlas accesibles a toda la comunidad a través del servidor espejo europeo de *Ontolingua* en la Universidad

Politécnica de Madrid<sup>44</sup>. Más tarde se construyó una versión actualizada en *DAML+OIL* mantenida por el *AIFB (Institute for Information and Formal Description Methods)* de la Universidad de Karlsruhe<sup>45</sup>.

En el contexto del proyecto europeo IST *Esperanto*, se desarrollaron cinco ontologías en *WebODE*<sup>46</sup> para describir los proyectos de investigación y desarrollo: “documentación”, “persona”, “organización”, y “reunión”. Estas ontologías describen los proyectos de investigación y desarrollo y su estructura, los documentos que se generan en un proyecto, las personas y organizaciones que participan en él, las reuniones, y las reuniones (administrativas, técnicas, etc.) que tienen lugar durante la duración de un proyecto.

En la figura 2.37 mostramos parte de la taxonomía de conceptos de la ontología “documentación”. Como muestra la figura, se distinguen cuatro tipos de documentos en un proyecto: gestión de documentos, documentos técnicos, publicaciones y documentos adicionales. Algunos de los documentos técnicos que se generan en un proyecto son: entregas, manuales y presentaciones. Las publicaciones pueden ser en libros o artículos, y hay distintos tipos de artículos, dependiendo de donde sean publicados (en un seminario, como parte de las actas de una conferencia, en libros o en revistas o publicaciones de investigación). Se puede acceder a estas ontologías a través de la Web de *ODESeW*<sup>47</sup>.

---

<sup>44</sup> <http://www-ksl-svc-lia.dia.fi.upm.es:5915/> Acceso con el usuario “ontologías-ka2” y con contraseña “adieu007”, aunque actualmente no está disponible.

<sup>45</sup> <http://ka2portal.aifb.uni-karlsruhe.de/> actualmente no disponible.

<sup>46</sup> <http://mayor2.dia.fi.upm.es/oeg/index.php/es/downloads/60-webode>

<sup>47</sup> <http://mayor2.dia.fi.upm.es/oeg/index.php/es/downloads/74-odesew>

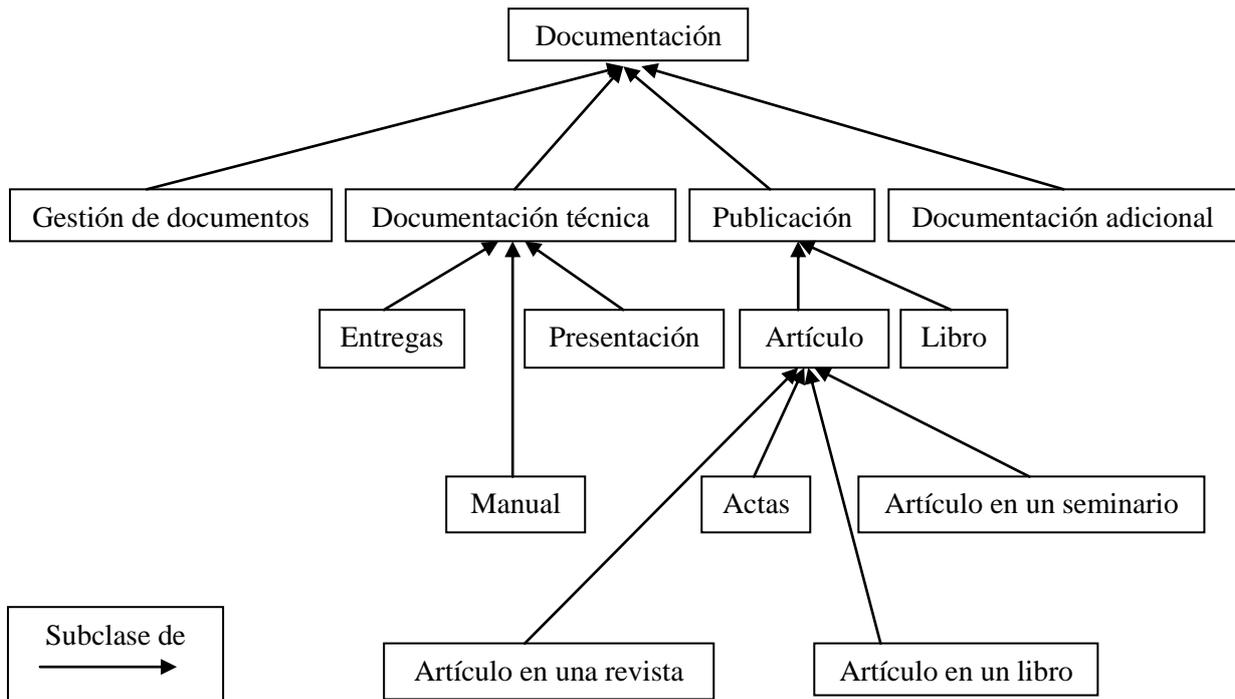


Figura 2.37. Fragmento de la ontología de documentación de los proyectos de investigación y desarrollo (adaptado de Gómez y otros 2004: 105).

#### 2.5.4.5. Legales

La última categoría de las ontologías de dominio corresponde a las ontologías legales, pero ofrecemos aquí una breve descripción ya que en la sección 2.7 las exponemos en más detalle.

La problemática de clasificación de las ontologías también se presenta al abordar la tarea de clasificación de las ontologías legales. Así, encontramos ontologías legales que representan el conocimiento legal (*FBO*, Van Kralingen, 1995; Visser, 1995), que analizan el discurso legal (*LLD*, McCarty, 1989), que reflejan la normativa legal (*NORMA*, Stamper, 1991; 1996), y de resolución de problemas legales (*FOLaw*, Valente, 1995). A éstas hay que añadir ontologías más generales con base léxica en *Wordnet* (*Jur-WordNet*, Gangemi, Sagri y Tiscornia, 2003; Gangemi, Prisco, Sagri, Steve y Tiscornia, 2003) y la ontología *LKIF* (Hoekstra y otros, 2008; 2007), basada en la ontología *LRI-Core* (Breuker y Hoekstra, 2004) que supone una fusión de conceptos legales fundamentales y categorías del sentido común.

Aunque en un primer momento las ontologías legales estaban dirigidas a compartir la información, otras se basaban en la verificación de una base de conocimiento o en la ingeniería del conocimiento, adquisición del conocimiento, reutilización del

conocimiento y desarrollo de una teoría del dominio específico. En los últimos años ha habido una gran proliferación de ontologías legales en el ámbito de la *Web 2.0*, centradas en el usuario en un entorno web, con una mayor especificidad en el ámbito legal: se ha añadido conocimiento legal profesional (*OPJK* de Casanovas y otros, 2004), contextos multimedia y globales (ontología sobre los derechos de autor de Delgado y otros, 2003; 2002), y conceptualización legal para no expertos (Klein y otros, 2006).

A continuación ofrecemos un resumen de la clasificación presentada adaptada de Gómez y otros (2004) y de lo expuesto anteriormente en esta sección:

1. Ontologías de representación del conocimiento
  - 1.1. *Frame Ontology (FO)* y *OKBC*
  - 1.2. *RDF* y *RDF Schema*
  - 1.3. *OIL*
  - 1.4. *DAML + OIL*
  - 1.5. *OWL*
2. Ontologías de alto nivel
  - 2.1. De universales y particulares
  - 2.2. *Sowa*
  - 2.3. *Cyc*
  - 2.4. *The Standard Upper Ontology (SUO)*
3. Ontologías lingüísticas
  - 3.1. *WordNet*
  - 3.2. *EuroWordNet*
  - 3.3. *The Generalized Upper Model (GUM)*
  - 3.4. *Mikrokosmos*
  - 3.5. *SENSUS*
  - 3.6. *DOLCE*
4. Ontologías de dominio
  - 4.1. De comercio electrónico
  - 4.2. Médicas
  - 4.3. De empresa
  - 4.4. De gestión del conocimiento
  - 4.5. Legales

## 2.6. Metodologías, lenguajes y herramientas de diseño de ontologías

### 2.6.1. Metodologías

En la década de los noventa y en los primeros años del siglo XXI, en el marco de la inteligencia artificial y el procesamiento del lenguaje natural ha surgido un gran interés por el desarrollo de metodologías para la creación de ontologías y la reutilización de ontologías ya existentes. No ha sido hasta comienzos del presente siglo cuando la construcción de ontologías se ha convertido en una actividad de ingeniería. Hasta entonces el proceso de creación de ontologías era un arte en el que cada grupo de investigación seguía sus propios principios, criterios de diseño y fases.

Desde las primeras conferencias sobre Inteligencia Artificial en 1996 y 1997, en conferencias internacionales posteriores del campo de la ingeniería ontológica (*EKAW*, *FOIS*, *C-CAP*, *KAW*, etc.) y en seminarios realizados conjuntamente con *AAAI*, *ECAI* e *IJCAI* se ha debatido ampliamente sobre metodologías de diseño y de evaluación de ontologías. En esta sección presentamos las metodologías, lenguajes y herramientas más comunes en el diseño de ontologías.

Como no hay consenso en el uso de los términos “metodología”, “método”, “técnica”, “proceso” y “actividad” adoptamos aquí las definiciones establecidas por el *IEEE* (1990) para la ingeniería de software. Para el *IEEE* una metodología es “un conjunto global de técnicas o métodos integrados que crea una teoría de sistemas generales de cómo ha de ser desarrollado lo que se considera un trabajo intenso”. Las metodologías se utilizan ampliamente en la ingeniería del conocimiento (Watermann, 1986; Wielinga y otros, 1992; Gómez y otros, 1997). De acuerdo a esta definición de “metodología”, los métodos y las técnicas son parte de la metodología. Un “método” es “un conjunto de procesos ordenados o procedimientos utilizados en la ingeniería de un producto o desarrollo de un servicio”. Una técnica es “un procedimiento técnico y directivo utilizado para conseguir un objetivo concreto”.

Según la definición de metodología, los “métodos” y “técnicas” están relacionados. De hecho, los “métodos” y las “técnicas” se utilizan para realizar tareas de los distintos procesos en los que consiste la metodología. Para el *IEEE* (1996) un “proceso” es “una función que ha de realizarse en la duración del software. Está compuesto de actividades.

Una “actividad” es “una tarea constituyente de un proceso”. Una “tarea” es la “unidad más pequeña de trabajo sujeta a la gestión de contabilidad”. Así una “tarea” es “un trabajo bien definido de uno o varios miembros del proyecto”. “Las tareas relacionadas se agrupan para conformar actividades”. Las relaciones entre estas definiciones aparecen en la figura 38, donde se observa que la metodología está compuesta de métodos y técnicas. Los métodos están compuestos de procesos y se detallan con técnicas. Los procesos están compuestos de actividades. Finalmente, las actividades forman grupos de tareas (cf. Gómez y otros, 2004: 107-109).

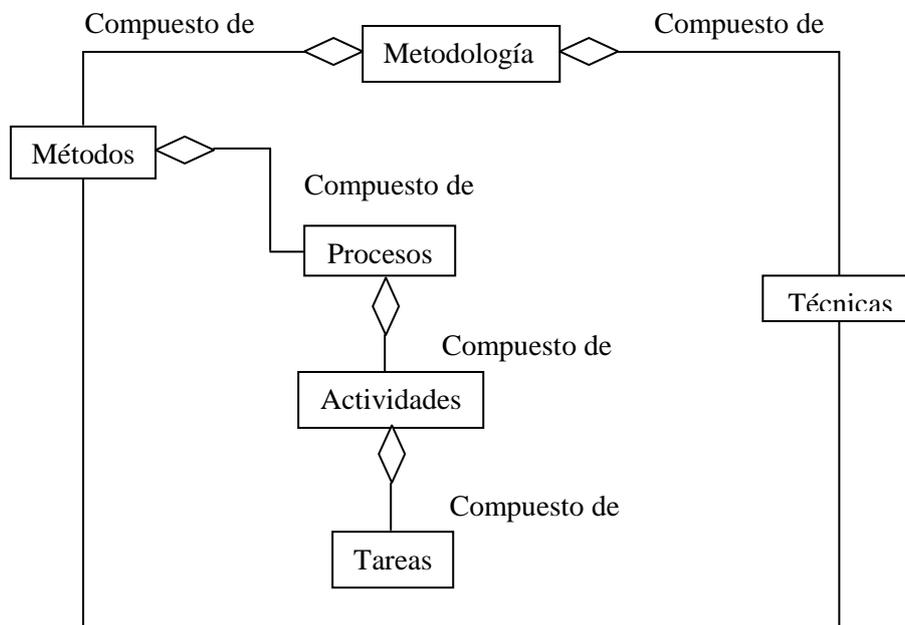


Figura 2.38. La representación gráfica de las relaciones terminológicas en las metodologías (adaptado de Gómez y otros, 2004: 109).

El proceso de desarrollo de ontologías se refiere a las actividades que se desarrollan en el su proceso de creación. Es crucial identificar estas actividades, sobre todo si es necesario un acuerdo entre grupos de trabajo que están distantes geográficamente. Es recomendable seguir tres tipos de actividades: actividades de gestión (planificación, control y garantía de calidad), actividades de desarrollo (pre-desarrollo, desarrollo y post-desarrollo) y actividades de apoyo (conjunto de actividades desarrolladas al mismo tiempo que las actividades de desarrollo sin las cuales la ontología no podría construirse).

### 2.6.1.1. El método Cyc

Cyc fue desarrollada por la *Microelectronics and Computer Technology Corporation (MCC)* como una gran base de conocimiento que implementaba conocimiento del sentido común. Implementa el lenguaje *CycL*, lenguaje híbrido que combina marcos y cálculos de predicados. El motor de inferencias de *CycL* permite la herencia múltiple, la clasificación automática, mantenimiento de enlaces inversos, comprobación de limitaciones, etc.; también ofrece un sistema de mantenimiento de certeza, la detección de contradicciones y un módulo de resolución. Cyc puede considerarse una ontología porque puede ser utilizada como base para la creación de diferentes sistemas de inteligencia que pueden comunicarse e interactuar.

Durante el desarrollo de CyC (Lenat y Guha, 1990) se siguieron los siguientes procesos:

1. Codificación manual de los artículos y piezas de conocimiento. Este primer proceso se realizó de forma manual porque los sistemas de lenguaje natural y máquina de aprendizaje existentes no podían manejar suficiente conocimiento del sentido común para realizar búsquedas de este tipo de nuevo conocimiento. El conocimiento se adquirió de tres formas diferentes:
  - Codificación del conocimiento requerido para entender libros y periódicos en búsqueda del conocimiento del sentido común.
  - Examen de los artículos que resultaban increíbles, para estudiar la decisión que hace que algunos artículos sean increíbles.
  - Identificación de preguntas que alguien tiene que ser capaz de responder tan sólo leyendo el texto.
2. Codificación del conocimiento con la ayuda de herramientas que utilizan el conocimiento almacenado en la base de conocimiento Cyc. Este segundo proceso es posible cuando hay almacenado suficiente conocimiento del sentido común como para buscar más.
3. Codificación del conocimiento principalmente desarrollado por herramientas que utilizan el conocimiento almacenado en la base de conocimiento Cyc. Las herramientas Cyc sirven al usuario para recomendar al sistema las fuentes de conocimiento a leer y a explicar las partes más difíciles del texto.

Hay varios módulos integrados en la base de conocimiento de *Cyc* y en el motor de inferencias *CycL*. Uno de ellos es el sistema de integración de bases de datos heterogéneas que vincula el vocabulario de *Cyc* con esquemas de bases de datos. Otro módulo es el buscador de conocimiento mejorado de información subtitulada, que permite realizar preguntas sobre imágenes que utilizan leyendas en lenguaje natural. Otro módulo es la integración guiada de terminología estructurada, *Guides Integration of Structured Terminology (GIST)*, que permite a los usuarios importar y manejar e integrar simultáneamente tesauros múltiples. Además, el módulo de recuperación de información WWW utiliza el lenguaje natural para acceder a la base de conocimiento *Cyc* y permite ampliar la base de conocimiento con conocimiento disponible en la Web (ver Gómez y otros, 2004: 113-115).

### **2.6.1.2. El método de Uschold y King**

Uschold y King (1995) propusieron el primer método de construcción de ontologías, ampliado en Uschold y Grüninger (1996). Plantearon unas directrices basadas en su experiencia para desarrollar la ontología de empresa. Como mencionamos en la sección 2.5.4, esta ontología se desarrolló como parte del proyecto de empresa del *Artificial Intelligence Applications Institute* de la Universidad de Edimburgo con sus socios *IBM*, *Lloyd's Register*, *Logica UK Limited* y *Unilever*.

De acuerdo a la aproximación de Uschold y King, hay que seguir los siguientes procesos para construir una ontología: (1) identificar el propósito de la ontología, (2) construirla, (3) evaluarla y (4) documentarla. Los autores proponen capturar el conocimiento, codificarlo e integrar otras ontologías dentro de la ontología actual durante el proceso de creación de la ontología (figura 2.39).

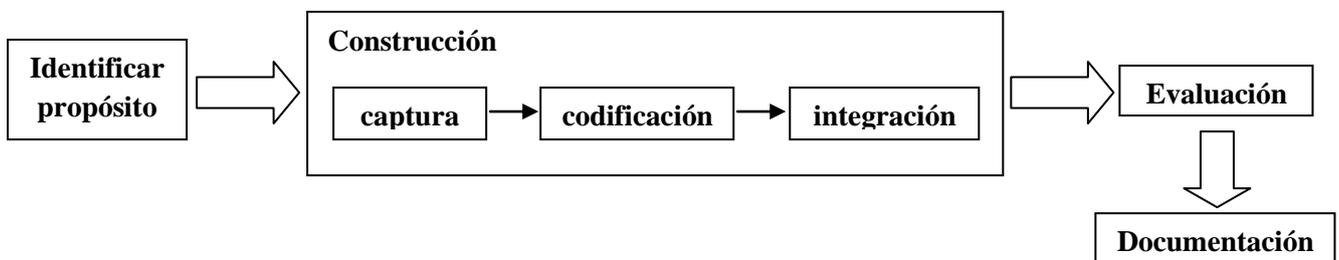


Figura 2.39. Los procesos de la metodología de Uschold y King (adaptado de Gómez y otros, 2004: 115)

Para identificar los conceptos de la ontología, Uschold y Grüninger (1996) propusieron tres estrategias: de abajo-arriba, de arriba-abajo y central. La estrategia de abajo-arriba propone comenzar por la identificación de los conceptos más específicos y después generalizar en conceptos más abstractos. Afirman que esta aproximación ofrece un gran nivel de detalle. En la estrategia de arriba-abajo los conceptos más abstractos se identifican primero y después los más específicos. El principal resultado de esta aproximación es un mayor control del nivel de detalle; sin embargo, comenzar por arriba puede resultar en elegir e imponer categorías de alto nivel posiblemente no necesarias. Finalmente, la estrategia central recomienda identificar primero el núcleo de los conceptos básicos para después especificarlos y generalizarlos como sea necesario. De este modo, si comenzamos por identificar los conceptos más importantes y definimos los conceptos superiores en base a éstos, las categorías de alto nivel surgirán de manera natural y serán así más susceptibles de ser estables.

Según Uschold y Grüninger (1996), los cuatro procesos mencionados (propósito, construcción, evaluación y documentación de la ontología) no son suficientes para obtener una metodología. Toda metodología debe también incluir un conjunto de técnicas, métodos y principios para cada uno de los cuatro procesos e indicar qué relaciones existen entre las etapas (por ejemplo, el orden recomendado, intercalado, input/output). La principal desventaja de este método es la falta de un proceso de conceptualización antes de implementar la ontología, lo que provoca (según Fernández y otros, 1999) que: (1) los expertos del dominio, los usuarios y los ingenieros del conocimiento tengan muchas dificultades para comprender las ontologías implementadas en lenguajes ontológicos; y (2) los expertos del dominio no podrán construir ontologías en su dominio. Para solucionar el problema (Gómez y otros, 2004: 119) proponen el proceso de re-ingeniería de ontologías.

### **2.6.1.3. El método de Grüninger y Fox (*TOVE*)**

Basado en la experiencia del proyecto *TOVE* en el dominio de la empresa, desarrollado por la Universidad de Toronto. Grüninger y Fox (1995) publicaron una aproximación formal para construir y evaluar ontologías. Esta metodología se ha seguido en la creación de ontologías *TOVE*, que son el pilar del banco de trabajo del diseño de empresas, un entorno de diseño que permite al usuario explorar una variedad de diseños de empresa.

Esta metodología está basada en el desarrollo de sistemas de conocimiento que utilizan la lógica de primer orden. Proponen identificar los principales escenarios de manera intuitiva, esto es, las posibles aplicaciones en las que se utilizará la ontología. Se utiliza un conjunto de preguntas en lenguaje natural, llamadas preguntas de competencia, para determinar el alcance de la ontología. Tales preguntas y sus respuestas se utilizan para extraer los conceptos principales y sus propiedades, relaciones y axiomas formales de la ontología. Por otro lado, el conocimiento se expresa formalmente en lógica de primer orden. Esta metodología es muy formal ya que se sirve de la robustez de la lógica clásica y puede ser utilizada como guía para transformar escenarios informales en modelos computacionales.

A continuación mostramos los procesos identificados en esta metodología en la figura 2.40.

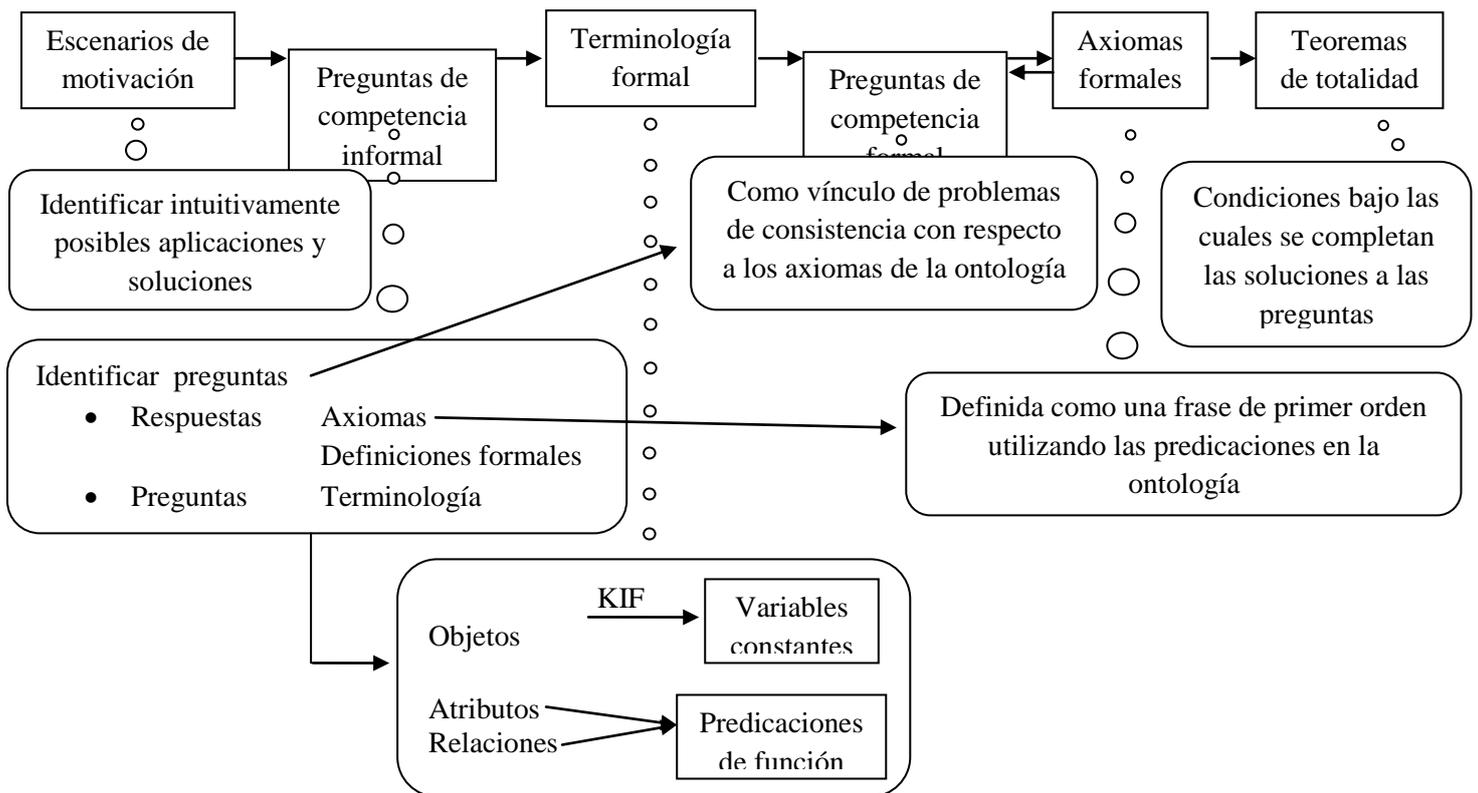


Figura 2.40. Los procesos de la metodología de Grüniger y Fox (adaptado de Gómez y otros, 2004: 120)

Esta ontología está bien fundamentada para la construcción y evaluación de ontologías, aunque carece de algunas actividades de gestión y apoyo.

#### 2.6.1.4. KACTUS

La metodología *KACTUS* fue propuesta por Bernaras y otros (1996) dentro del marco del proyecto *Esprit KACTUS* (*KACTUS*, 1996). Un objetivo de este proyecto era investigar la posibilidad de la reutilización del conocimiento en sistemas técnicos complejos y el papel de las ontologías para apoyarlos (Schreiber y otros, 1995).

Esta metodología está condicionada por el desarrollo de aplicaciones. Por tanto, cada vez que se construye una aplicación, se refina la ontología que representa el conocimiento requerido para la aplicación. La ontología se puede desarrollar por la reutilización de otras y puede integrarse en ontologías de aplicaciones posteriores. Por consiguiente, cada vez que se desarrolla una aplicación tienen lugar los siguientes procesos (Bernaras y otros, 1996):

Proceso 1. La especificación de la aplicación ofrece el contexto de una aplicación y una perspectiva de los componentes que la ontología pretende modelar. En este proceso hay que ofrecer una lista de términos y tareas.

Proceso 2. Diseño preliminar basado en categorías ontológicas relevantes de alto nivel. En este proceso se utiliza la lista de términos y tareas obtenida en el proceso anterior como *input* para obtener varias perspectivas del modelo global, de acuerdo a las categorías ontológicas de alto nivel, por ejemplo, concepto, relación, atributo, etc. Este proceso de diseño implica la búsqueda de ontologías ya desarrolladas (quizás para otras aplicaciones). Estas ontologías son refinadas y ampliadas posteriormente, como se muestra en la figura 2.41.

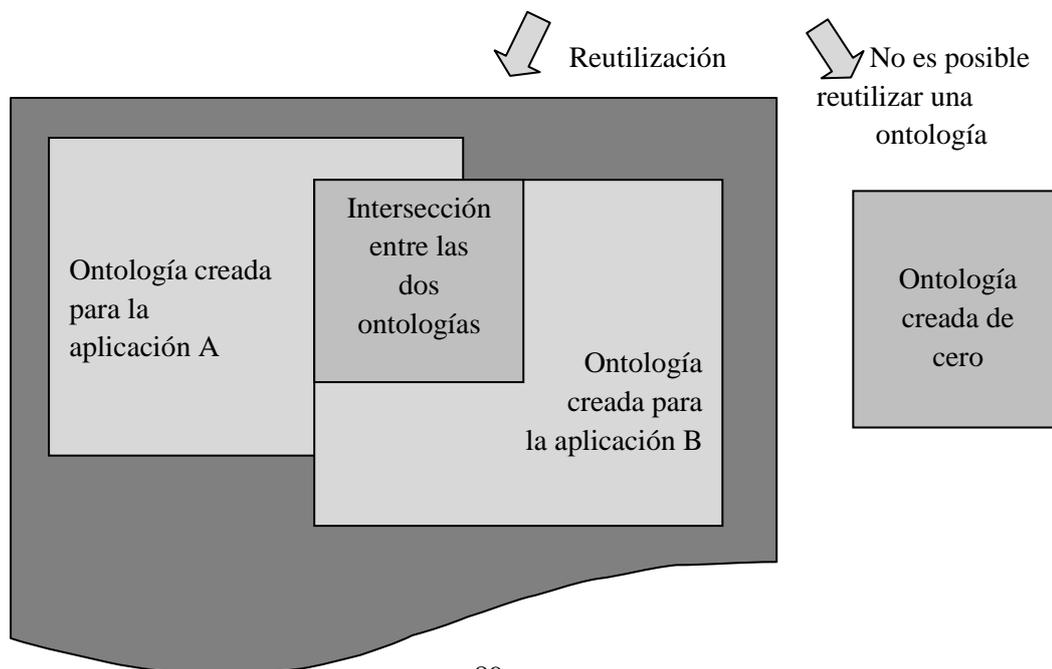


Figura 2.41. La metodología *KACTUS* (adaptado de Gómez y otros 2004: 124).

Proceso 3. Refinamiento y estructuración de la ontología para conseguir un diseño definitivo de acuerdo a la modularización y los principios de organización jerárquica. Para su aplicación los diseñadores de la ontología han de seguir las recomendaciones y metodologías tradicionales de la ingeniería del conocimiento.

Los autores de esta metodología presentan el desarrollo de tres ontologías como resultado del desarrollo de tres aplicaciones. El propósito de la primera aplicación es diagnosticar fallos en la red eléctrica. La segunda tiene la finalidad de planificar la reanudación del servicio tras un fallo en la red. Para cuando empezó el desarrollo de la ontología de esta aplicación, la ontología de fallos ya se había construido. Entonces los desarrolladores refinaron y aumentaron los conceptos del dominio y modificaron los conceptos relevantes del dominio ya identificados en la ontología de diagnóstico. Más tarde refinaron la estructura de la ontología para obtener su diseño definitivo. La tercera aplicación controla la red eléctrica. Se basa en las otras dos. A la ontología de esta aplicación se la puede considerar como la unión entre la ontología de dominio para el diagnóstico y la planificación del servicio de recuperación. La unión de las ontologías reside en un conjunto de subontologías que pertenecen a la intersección y otros conjuntos utilizados para una aplicación u otra, pero no para ambas a la vez. Las subontologías de la intersección tienen más posibilidades de ser reutilizadas ya que las adaptaciones relevantes en estas ontologías deben llevarse a cabo en, al menos, dos aplicaciones diferentes (cf. Gómez y otros, 2004: 124-125).

#### **2.6.1.5. METHONTOLOGY**

La metodología *METHONTOLOGY* se desarrolló dentro del Grupo de Ingeniería Ontológica de la UPM (Universidad Politécnica de Madrid). *METHONTOLOGY* (Fernández y otros, 1997; Gómez, 1999a; Fernández y otros, 1999) permite la construcción de ontologías en el nivel de conocimiento. Esta metodología se elaboró partiendo de las actividades principales identificadas en el proceso de desarrollo de software (IEEE, 1997) y de otras metodologías de ingeniería del conocimiento (Gómez y otros, 1997). Incluye la identificación del proceso de desarrollo de ontologías (mencionado al comienzo de esta sección), el ciclo de vida de la ontología, basada en

prototipos de evolución, y las técnicas para realizar cada actividad en la gestión, orientación al desarrollo y actividades de apoyo.

Las herramientas *ODE* (Blázquez y otros, 1998) y *WebODE* (Arpírez y otros, 2003) se crearon para dotar a *METHONTOLOGY* de apoyo tecnológico. Sin embargo, también es posible utilizar esta metodología con otras herramientas de creación de ontologías y suites, como por ejemplo *Protégé-2000* (Noy y otros, 2000), *OntoEdit* (Sure y otros, 2002) y *KAON* (Maedche y otros, 2003). *METHONTOLOGY* propone un proceso de desarrollo, un ciclo de vida de la ontología y una especificación para cada una de las actividades que se realizan durante el proceso de desarrollo (ver Gómez y otros 2008: 174).

A continuación describimos qué entienden Gómez y otros (2008: 174-178) por proceso de desarrollo de ontologías y por ciclo de vida.

Según estos autores, el proceso de desarrollo de ontologías consiste en la realización de las actividades necesarias para construir ontologías, que pueden clasificarse en actividades de gestión, desarrollo y apoyo (figura 2.42).

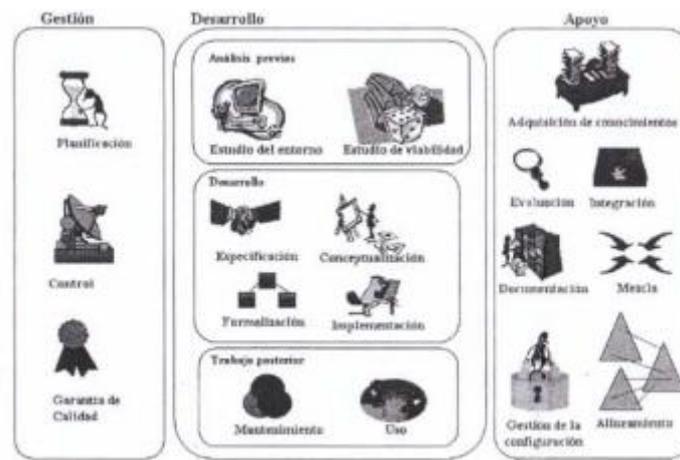


Figura 2.42. Proceso de desarrollo de ontologías en *METHONTOLOGY* (Gómez y otros, 2008: 175).

Las **actividades de gestión** incluyen “planificación”, “control” y “garantía de calidad”. La actividad de “planificación” incluye las tareas a realizar y su orden, además del tiempo y los recursos para completarlas. La planificación es esencial para las ontologías que en su creación utilizan ontologías ya existentes almacenadas en librerías de

ontologías, o para las ontologías que requieren un alto nivel de abstracción y generalidad. La actividad de “control” garantiza que las tareas programadas se realicen en tiempo y forma, y la actividad de “garantía de calidad” certifica la calidad satisfactoria de la ontología, programas y documentación.

Las **actividades orientadas al desarrollo** se clasifican en tres grupos: actividades previas al desarrollo, desarrollo propiamente dicho y actividades posteriores. Durante las actividades previas al desarrollo, un “estudio del entorno” identifica, entre otros, el problema a resolver por la ontología y las aplicaciones en las que se integrará; además, dentro del marco de las actividades previas también tiene lugar un “estudio de viabilidad” para responder a preguntas como: ¿es posible construir una ontología?, ¿es conveniente construirla?, etc. Durante las “actividades de desarrollo propiamente dichas” la tarea de “especificación” establece el propósito de la ontología, sus usos y los usuarios finales; mientras que la tarea de “conceptualización” estructura el conocimiento del dominio o bien desde cero o bien con modelos ya existentes. Por otro lado, la tarea de “formalización” consiste en transformar el modelo conceptual a un modelo formal; y, por último, la “actividad de implementación” crea modelos computables en un lenguaje de ontologías. Tras el desarrollo, en la actividad de “mantenimiento” se actualiza y corrige la ontología en caso necesario. Es en esta fase cuando la ontología puede reutilizarse con otras ontologías o aplicaciones.

Para finalizar, las **actividades de apoyo** incluyen una serie de tareas que se pueden realizar durante las actividades de desarrollo, sin las cuales no se podría construir la ontología. Entre estas tareas están: la “adquisición de conocimientos”, “evaluación”, “integración”, “mezcla”, “alineamiento”, “documentación” y “gestión de la configuración”. La tarea de “adquisición de conocimientos” tiene como objetivo adquirir o bien los conocimientos de distintos expertos sobre la materia, o bien a través de un tipo de proceso semiautomático que Gómez y otros (2008) y otros autores han llamado “aprendizaje de ontologías” (Kietz y otros, 2000). La actividad de “evaluación” consiste en emitir un juicio técnico sobre las ontologías, el software asociado a ellas y su documentación. Por otro lado, se realiza la actividad de “integración” cuando se construye una ontología a partir de otras disponibles. La actividad de “mezcla” es especialmente significativa ya que para obtener una ontología nueva se combinan varias

ontologías del mismo dominio. Cabe destacar que esta actividad se puede llevar a cabo también durante el período de ejecución o de diseño. Durante la actividad de “alineamiento” se establecen correspondencias entre las ontologías implicadas. La actividad de “documentación” sirve para detallar los estadios ya completados y los productos creados. En cuanto a la “gestión de configuración”, se utiliza para registrar todas las versiones de la documentación y del código de la ontología. Por último, Gómez y otros (2008: 176) señalan que también es posible una “actividad multilingüe” para establecer correspondencias entre las ontologías y las descripciones formales de conocimientos lingüísticos, herramienta especialmente útil en las ontologías en red disponibles a través de la Web Semántica.

Con respecto al ciclo de vida de la ontología, Gómez y otros (2008: 177) lo definen como el proceso que “se encarga de expresar cuándo se tienen que hacer, es decir, establece el conjunto de etapas o fases que atraviesa la ontología durante su vida útil, y describe la actividad que hay que realizar en cada fase y la relación entre las fases”. La primera versión del ciclo de vida de la ontología (figura 2.43) propone comenzar por una “planificación” de las actividades a realizar, para luego seguir con la “especificación” del porqué de la construcción de la ontología, posibles usos, destinatarios,...etc. Una vez terminada la “especificación”, en la “conceptualización” se organizan y estructuran los conocimientos adquiridos. Una vez que se ha obtenido el modelo conceptual se formaliza e implementa (cf. Gómez, 2004).

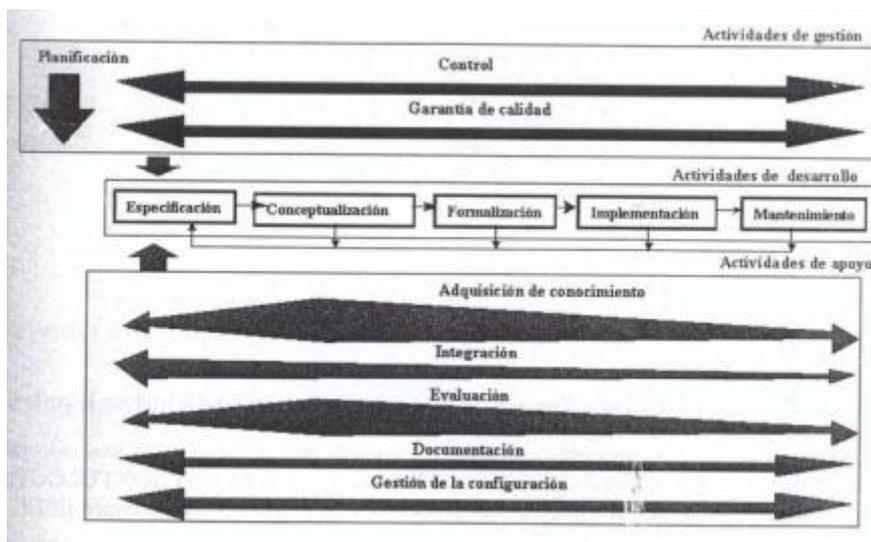


Figura 2.43. Ciclo de vida de la ontología en *METHONTOLOGY* (Gómez y otros, 2008: 177).

En la fase de conceptualización, incluye el conjunto de tareas de conocimiento estructurado que aparecen en la figura 2.44. Se destacan los distintos componentes de la ontología (conceptos, atributos, relaciones, constantes, axiomas formales, reglas e instancias) construidos dentro de cada tarea y el orden a seguir.

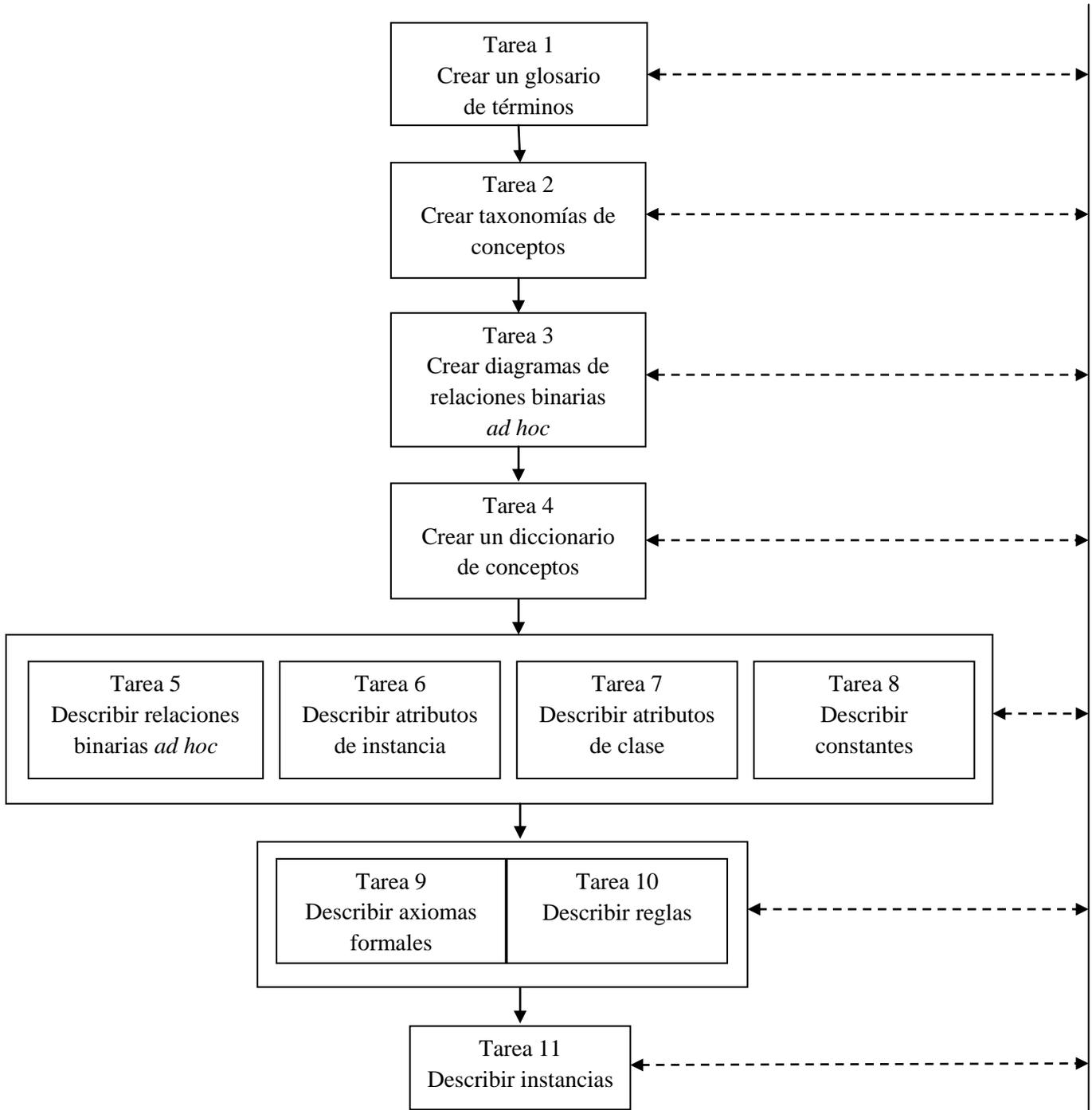


Figura 2.44. Tareas de conceptualización en *METHONTOLOGY* (Gómez y otros, 2008: 179).

A continuación describimos sucintamente cada una de las tareas reflejadas en la figura y expuestas por Gómez y otros (2008).

**Tarea 1: Construir un glosario de términos.** Este glosario ha de incluir todos los términos relevantes del dominio (conceptos, instancias, atributos, relaciones entre conceptos, etc.), su descripción en lenguaje natural y sus sinónimos y/o antónimos.

**Tarea 2: Construir taxonomías de conceptos.** Para elaborar la jerarquía se seleccionan los conceptos del glosario de términos en cuestión. Para ello *METHONTOLOGY* propone utilizar las relaciones taxonómicas definidas en la *Frame Ontology* (Farquhar y otros, 1997) y en la Ontología *OKBC*, que son “subclase de”, “descomposición disjunta” y “descomposición exhaustiva”. Una vez estructurada la taxonomía de conceptos se comprobará que las taxonomías no contienen errores (Gómez, 2006).

**Tarea 3: Construir diagramas de relaciones binarias *ad hoc*.** El objetivo de este diagrama es establecer relaciones *ad hoc* entre los conceptos de la misma o diferente taxonomía. Antes de continuar con la siguiente tarea hay que comprobar que no existen errores.

**Tarea 4: Construir el diccionario de conceptos.** En el diccionario de conceptos se especifican las relaciones y propiedades de cada concepto de la taxonomía. El diccionario de conceptos contiene todos los conceptos de la taxonomía, sus relaciones, instancias y sus atributos de instancia y de clase.

**Tarea 5: Definir minuciosamente las relaciones binarias *ad hoc*.** En esta tarea se describen todas las relaciones binarias *ad hoc* incluidas en el diccionario de conceptos para así elaborar una tabla. En cada relación binaria *ad hoc* hay que especificar tanto su nombre, como los nombres de los conceptos origen y destino, su cardinalidad y su relación inversa.

**Tarea 6: Definir detalladamente los atributos de instancia.** Aquí se definen todos los atributos de instancia incluidos en el diccionario de conceptos. Los atributos de instancia describen las instancias del concepto. Es necesario especificar, entre otros, los siguientes campos para cada atributo de instancia: nombre, concepto al que pertenece, tipo de valor, intervalo de valores (en el caso de los valores numéricos), cardinalidad mínima y máxima, atributos de instancia, atributos de clase y constantes utilizadas para inferir los valores del atributo, ...etc.

**Tarea 7: Definir minuciosamente los atributos de clase.** Mediante una tabla se describen los atributos de clase ya incluidos en el diccionario de conceptos. Hay que

incluir la siguiente información en cada atributo de clase: nombre, nombre del concepto en el que se define el atributo, tipo de valor, valor(es), cardinalidad,...etc.

**Tarea 8: Definir las constantes minuciosamente.** Aquí se define cada una de las constantes incluidas en el glosario de términos. En la tabla de constantes hay que especificar el nombre, tipo de valor, el valor, la unidad de medida para las constantes numéricas y los atributos que se pueden inferir con la constante.

**Tarea 9: Definir axiomas formales.** Una vez identificados los axiomas formales necesarios en la ontología se describen detalladamente. Para ello se especifica el nombre, la descripción en lenguaje natural, la expresión que describe formalmente el axioma con lógica de primer orden, los conceptos, los atributos y las relaciones *ad hoc* a las que el axioma se refiere y las variables utilizadas.

**Tarea 10: Definir las reglas.** Primero hay que identificar qué reglas se necesitan en la ontología para luego describirlas en la tabla de reglas. Para definir las reglas hay que especificar el nombre, la descripción en lenguaje natural, la expresión que la regla describe formalmente, los conceptos, los atributos y relaciones a los que la regla se refiere y las variables que se utilizan.

**Tarea 11: Definir instancias.** Una vez creado el modelo conceptual de la ontología, se definen las instancias relevantes que aparecerán en el diccionario de conceptos y en la tabla de instancias. Para ello entre la información que se especifica está el nombre, el nombre del concepto al que pertenece y sus valores en los atributos.

Si bien METHONTOLOGY establece las tareas a realizar, no propone el orden en el que deben llevarse a cabo.

#### **2.6.1.6. SENSUS**

El método *SENSUS* propone vincular términos específicos de un dominio a la gran ontología *SENSUS*, y allí eliminar los términos irrelevantes para la nueva ontología que se pretende construir. Como resultado se obtiene la estructura de la nueva ontología, que se genera de forma automática utilizando el proceso que se describe más abajo y la herramienta *OntoSaurus*, que se describe en la sección 2.6.3.

De acuerdo a este método, para construir una ontología de un dominio específico hay que realizar los siguientes procesos:

**Proceso 1: Identificar los términos clave.** Por ejemplo, si queremos construir una ontología de vuelos podemos tomar los términos presentados en la figura 2.5 como términos clave: vuelo Europa-África, vuelo Europa-América, vuelo Londres-Liverpool, vuelo Madrid-Barcelona.

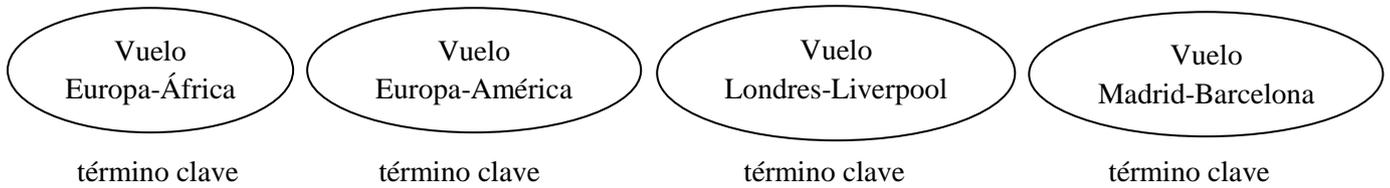


Figura 2.45. Proceso 1. Identificar términos clave (adaptado de Gómez y otros, 2004: 143).

**Proceso 2: Vincular de forma manual los términos clave a SENSUS.** Como se ve en la figura 2.46, los dos primeros tipos de enfermedades del ejemplo anterior irían vinculados a SENSUS como hipónimos (subclases).

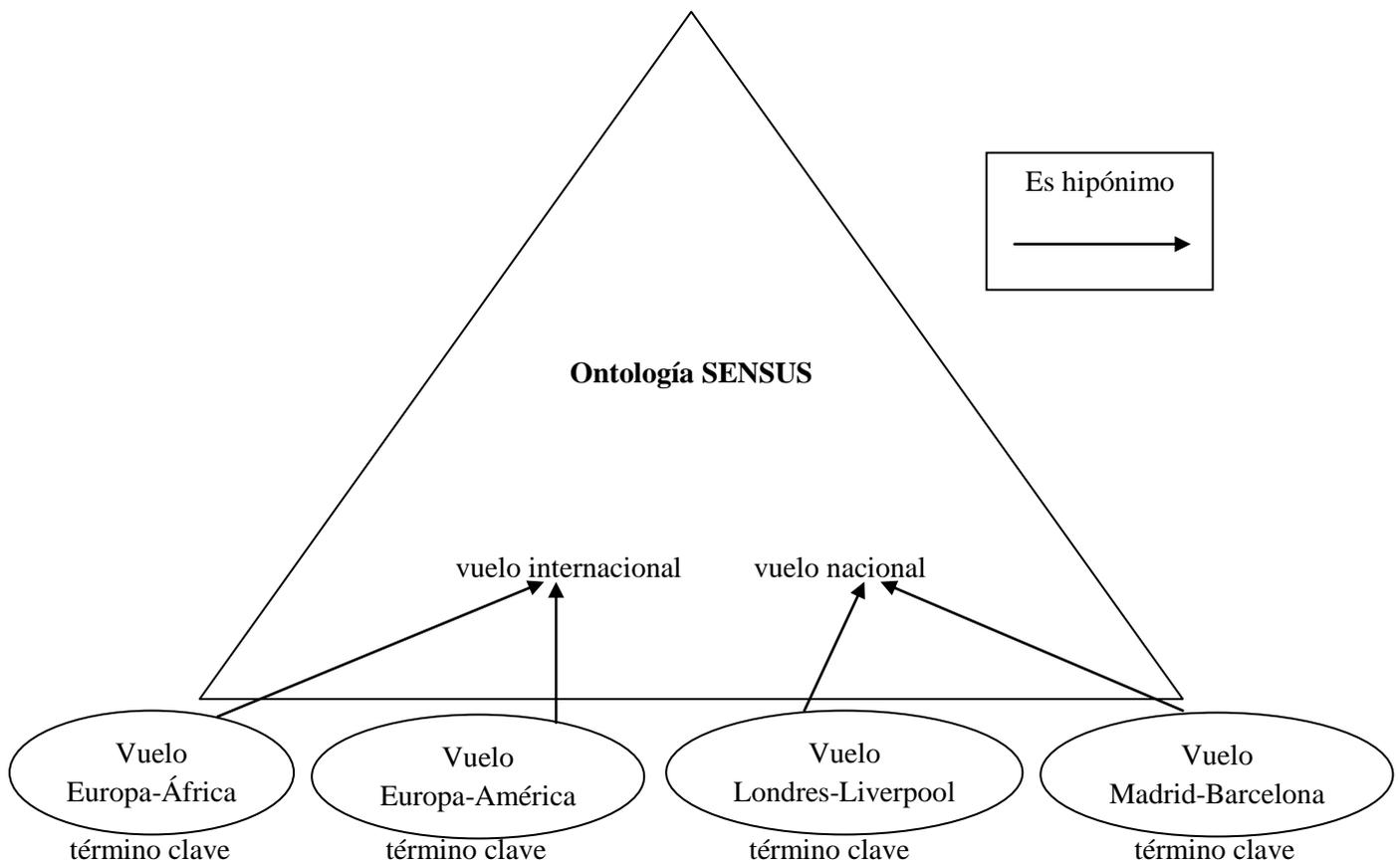


Figura 2.46. Proceso 2. Vincular los términos clave a SENSUS (adaptado de Gómez y otros, 2004: 144).

**Proceso 3: Añadir caminos al núcleo.** Se incluyen todos los conceptos del camino, desde los términos clave al núcleo de *SENSUS*. *OntoSaurus* ofreció el conjunto de conceptos presentados en la figura 2.47.

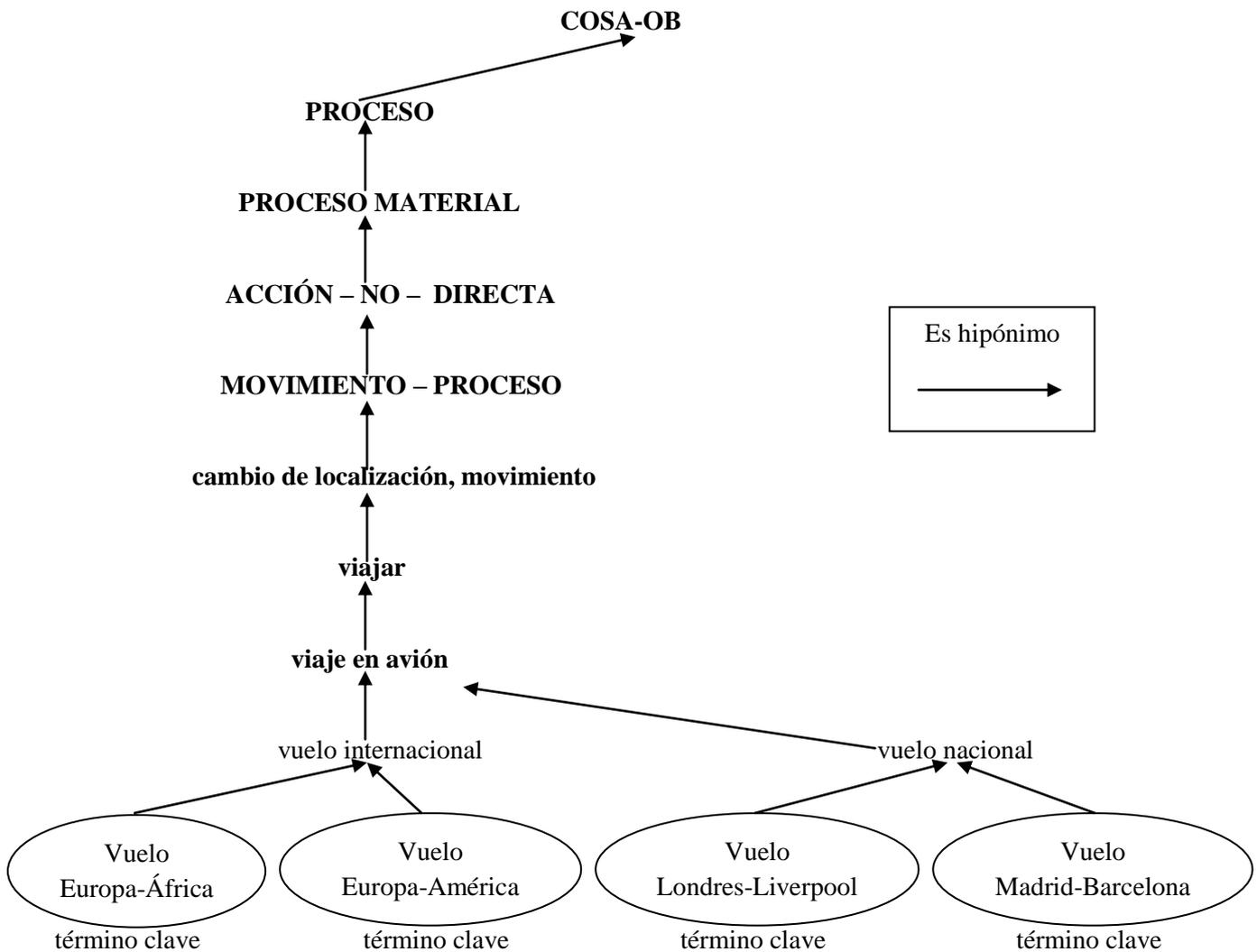


Figura 2.47. Proceso 3. Añadir caminos al núcleo (adaptado de Gómez y otros, 2004: 144).

**Proceso 4: Añadir nuevos términos del dominio.** En este proceso se añaden de forma manual los términos relevantes para el dominio. Por tanto se repiten los procesos 2 y 3 para incluir conceptos en el camino que va desde los nuevos conceptos al núcleo de *SENSUS*. En nuestro ejemplo podemos añadir los términos “origen” y “destino” como se muestra en la figura 2.48. También es aconsejable añadir los términos ascendentes de los nuevos términos.

**Proceso 5:** Añadir subárboles completos. En este proceso los ingenieros del conocimiento han de prestar especial atención a los nodos con gran número de caminos en el árbol resultante. Para el subárbol bajo cada uno de los nodos, el ingeniero de conocimiento ha de decidir si añade un nodo o no. El criterio a seguir es, si se ha encontrado que muchos de los nodos de un subárbol son relevantes, entonces los otros nodos del subárbol tienen altas probabilidades de ser relevantes también. Como se aprecia en la figura 2.49, podemos añadir un subárbol entero de viaje en vuelo y los conceptos vuelo nocturno y vuelo directo. Este proceso se realiza de forma manual, dado que el ingeniero del conocimiento ha de contar con un conocimiento del dominio suficiente como para tomar la decisión de añadir nodos nuevos. Obviamente, los nodos de muy alto nivel en la ontología siempre tendrán muchos caminos hasta ellos, pero no es apropiado incluir subárboles enteros bajo estos nodos.

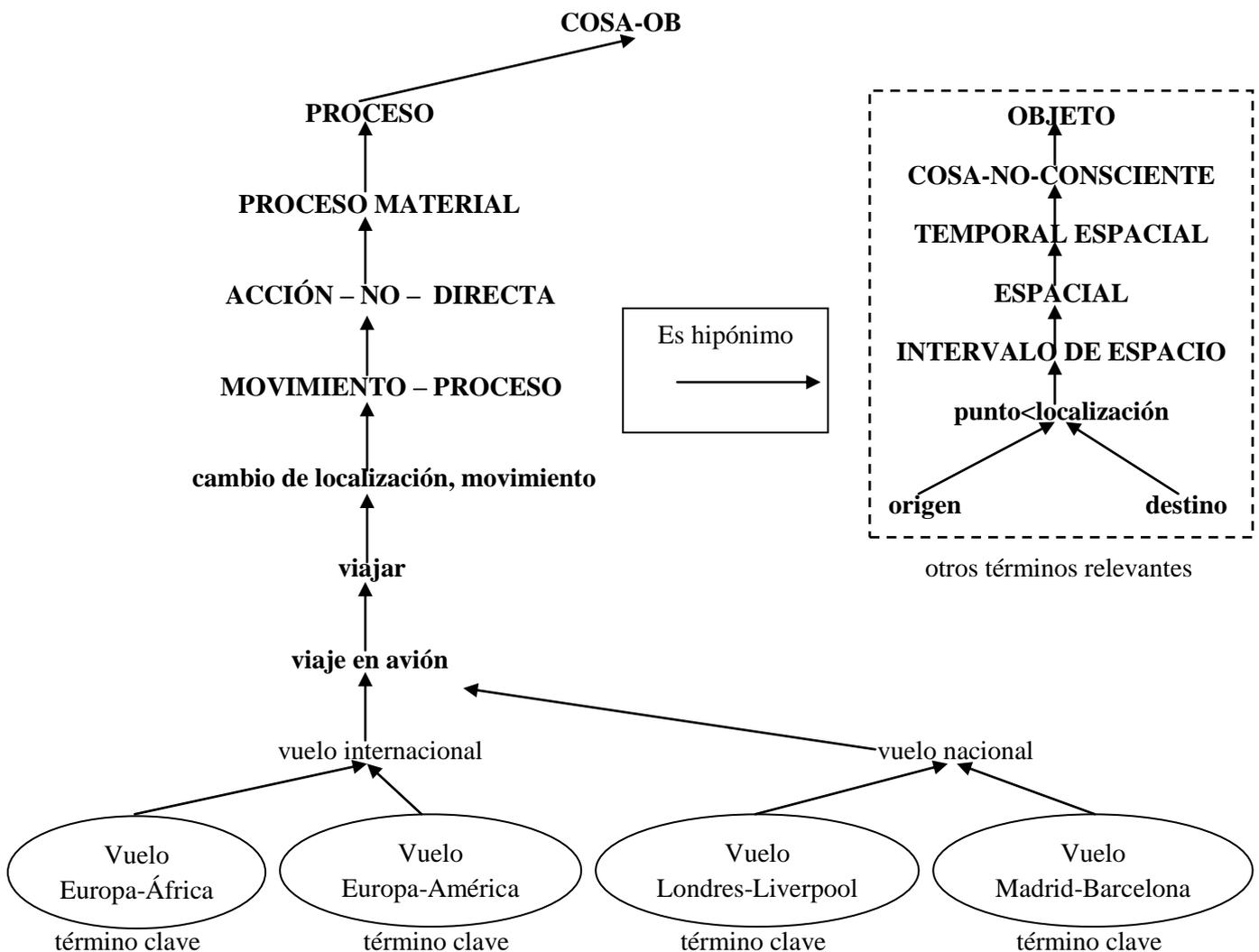


Figura 2.48. Proceso 4. Añadir nuevos términos del dominio (adaptado de Gómez y otros, 2004: 145).

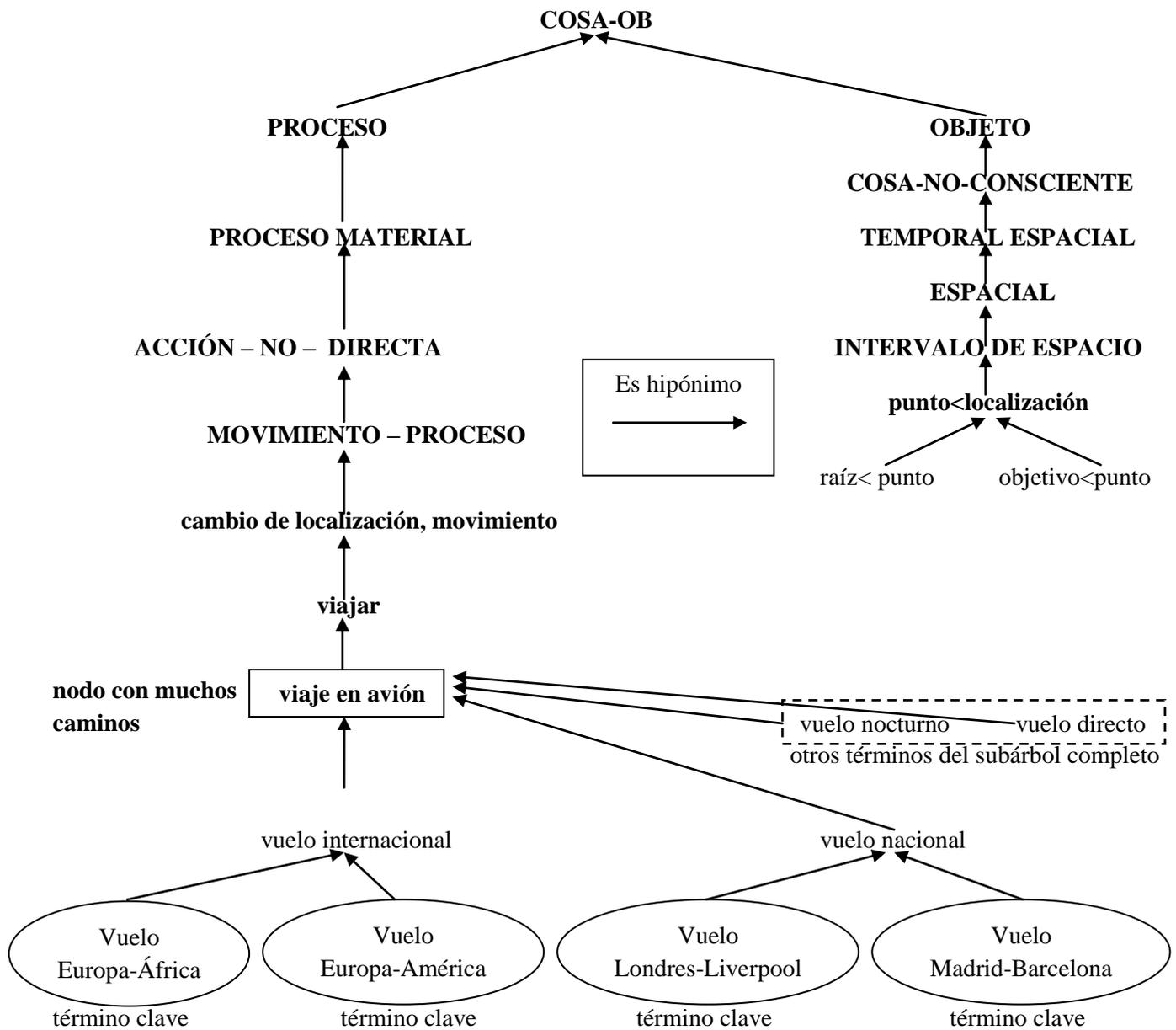


Figura 2.49. Proceso 5. Añadir subárboles completos (adaptado de Gómez y otros, 2004: 145).

La metodología de *SENSUS* promueve compartir el conocimiento ya que la misma base ontológica se utiliza para desarrollar ontologías de dominio. Una gran ventaja de esta metodología es que si dos ontologías se desarrollan de forma independiente la amplia cobertura de *SENSUS* hace de bisagra que une la terminología y la organización de una ontología con la otra (Swartout y otros, 1997).

### 2.6.1.7. *On-To-Knowledge*

El objetivo principal del proyecto *On-To-Knowledge* (Staab y otros, 2001) es aplicar ontologías a información disponible electrónicamente para mejorar la calidad de la gestión de conocimiento en grandes organizaciones. Algunos socios de este proyecto son el Instituto *AIFB* de la Universidad de Karlsruhe, la *Vrije Universiteit* de Amsterdam y *British Telecom*. En este proyecto se desarrolló una metodología y herramientas de acceso inteligente a grandes volúmenes de fuentes de información semi-estructurada y textual en entornos intra-, extra- y basados en internet. *On-To-Knowledge* incluye una metodología para construir ontologías susceptibles de ser utilizadas en la aplicación de gestión del conocimiento. Por tanto, esta metodología propone construir la ontología teniendo en cuenta cómo se va a utilizar la ontología en aplicaciones futuras. Por consiguiente, las ontologías desarrolladas con esta metodología son altamente dependientes de la aplicación. Otra característica principal es que esta metodología propone el aprendizaje de ontologías para reducir esfuerzos a la hora de desarrollar la ontología.

*On-To-Knowledge* ofrece un conjunto de técnicas, métodos y principios para cada uno de sus procesos e indica las relaciones entre ellos. Los procesos propuestos por esta metodología aparecen en la figura 2.50.

**Proceso 1: Estudio de viabilidad.** El tipo de estudio de viabilidad adoptado por *On-To-Knowledge* es el descrito en la metodología *CommonKADS* (Schreiber y otros, 1999). Según *On-To-Knowledge*, el estudio de viabilidad es aplicable a toda la aplicación y, por tanto, ha de ser llevado a cabo antes del desarrollo de las ontologías. De hecho, el estudio de viabilidad sirve de base para el inicio de la ontología.

**Proceso 2: Inicio de la ontología.** El resultado de este proceso es el documento de especificación de requisitos ontológicos que describe el dominio y propósito de la ontología, las directrices de diseño (por ejemplo las convenciones sobre nomenclatura), las fuentes de conocimiento disponibles (libros, revistas, etc.), usuarios potenciales y casos de uso, al igual que las aplicaciones compatibles con la ontología. En este proceso los ingenieros ontológicos buscarán ontologías ya desarrolladas potencialmente reutilizables.

**Proceso 3: Refinamiento.** El propósito de este proceso es producir una ontología de aplicación meta de acuerdo a las especificaciones del proceso anterior. El proceso de refinamiento se divide a su vez en dos actividades:

- Actividad 1. Proceso de obtención de conceptos del dominio en cuestión con la ayuda de expertos. En el proceso de refinamiento el primer borrador de la ontología obtenida en el proceso 2 se refina con la ayuda de la interacción de los expertos del dominio. Durante el proceso de obtención de conceptos, se recogen los conceptos en un lado y los términos para etiquetar a los conceptos en otro. Después se unen. La metodología de *On-To-Knowledge* propone el uso de representaciones intermedias para modelar el conocimiento.
- Actividad 2. Formalización. Se implementa la ontología utilizando un lenguaje de representación. El lenguaje se elige de acuerdo a los requisitos específicos de la aplicación prevista. Para realizar la formalización, *On-To-Knowledge* recomienda el editor de ontologías *OntoEdit*, que genera automáticamente el código ontológico en varias lenguas.

**Proceso 4: Evaluación.** El proceso de evaluación sirve para comprobar la utilidad de las ontologías desarrolladas y su entorno de software asociado. Al producto resultante se le llama aplicación basada en ontologías. Durante este proceso se realizan dos tareas:

- Actividad 1. Comprobación de los requisitos y cuestiones de competencia. Los ingenieros ontológicos comprueban si la ontología cumple los requisitos y si puede “responder” a preguntas de competencia.
- Actividad 2. Evaluación de la ontología en el entorno de la aplicación meta. En esta actividad puede surgir la necesidad de un mayor refinamiento.

**Proceso 5: Mantenimiento.** Según *On-To-Knowledge* es necesario el mantenimiento de la ontología como parte del software del sistema.

Se ha utilizado *On-To-Knowledge* para crear empresas virtuales, para organizar memorias corporativas y como funcionalidad de soporte técnico en centros de atención al cliente. Esta última aplicación ofrece a los clientes la información apropiada de los productos y servicios, y concluye sobre la capacidad de la compañía de establecer relaciones estables y de éxito con sus clientes.

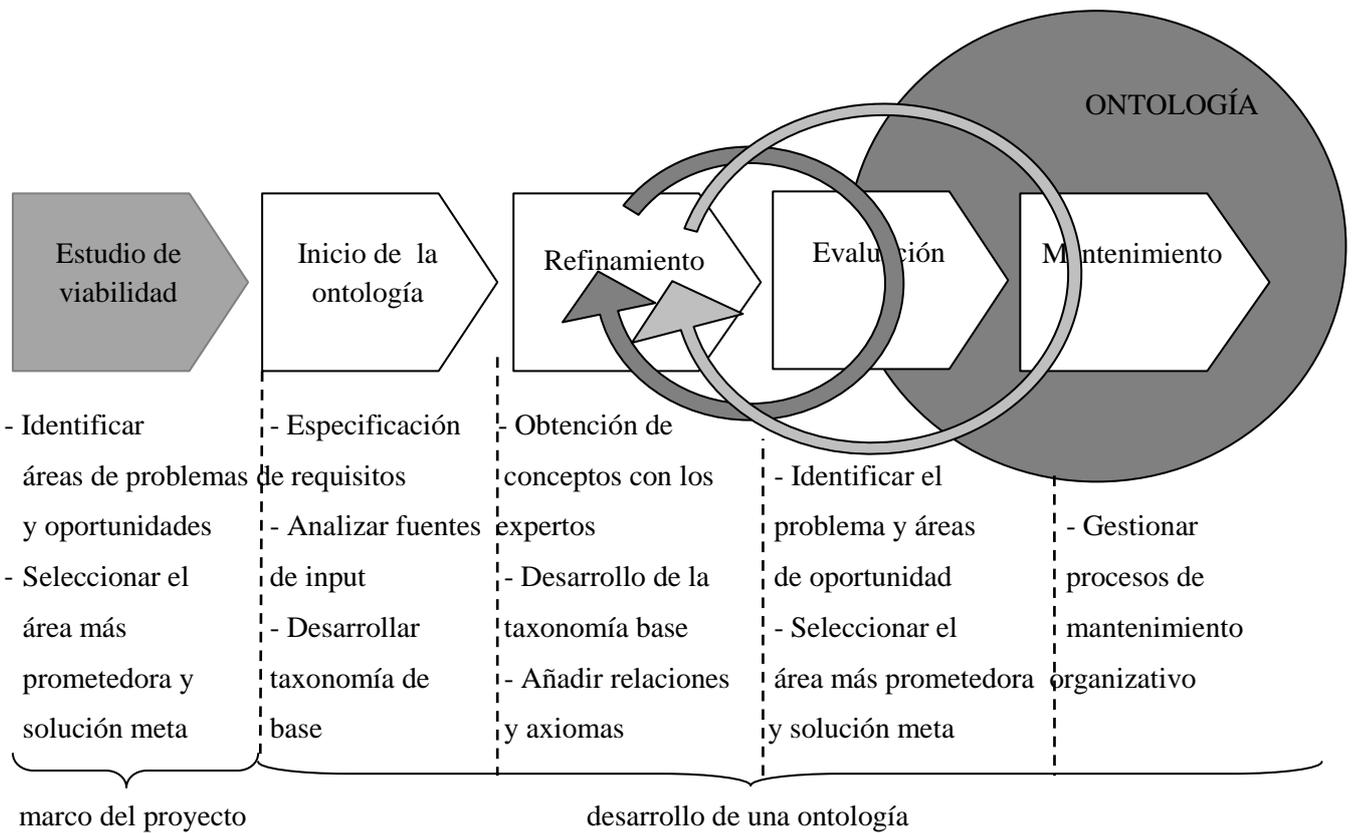


Figura 2.50. Los procesos en *On-To-Knowledge* (adaptado de Staab y otros, 2001).

### 2.6.1.8. COHERENT

Por último, pero no por ello menos importante, presentamos la metodología *COHERENT* (*C*onceptualization + *H*iErarchization + *R*Emodelling + *r*efinemeNT) de Periñán y Mairal (2011), utilizada para la construcción del nivel conceptual básico de la ontología nuclear de la base de conocimiento *FunGramKB* y base para el desarrollo de ontologías satélite vinculadas a ella.

En un primer paso se identificaron los conceptos básicos del *Longman Defining Vocabulary (LDV)* del *Longman Dictionary of Contemporary English* (Procter, 1978), el cual se ha probado como una fuente útil para el vocabulario básico de un lenguaje artificial. Sin embargo, fue necesaria una revisión profunda para poder desarrollar un mapa conceptual. Particularmente, tanto la población como estructuración del nivel conceptual básico en la ontología nuclear se desarrollaron de forma manual siguiendo la metodología *COHERENT* en cuatro fases que se muestran en la figura 2.51.

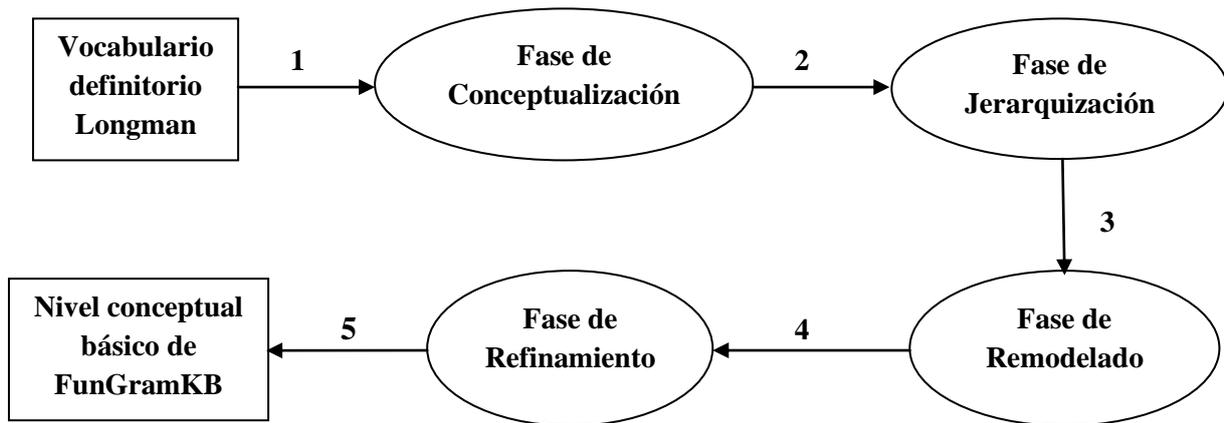


Figura 2.51. La metodología *COHERENT* [adaptado de Perrián y Mairal (2011: 20)].

- (1) Lista de las unidades léxicas en inglés
- (2) Inventario de unidades conceptuales en varias lenguas
- (3) Taxonomía jerárquica de conceptos básicos incluyendo sus Marcos Temáticos (MT) y Postulados de Significado (PS)
- (4) Taxonomía conceptual que incluye a los subconceptos
- (5) Nivel básico refinado de la ontología nuclear

**1. La fase de conceptualización.** Se tomó como punto de partida el *LDV*, esto es, un inventario de 2.197 unidades léxicas inglesas que facilitan la descripción semántica de cualquier tipo de palabra. La motivación fue realizar un mapa conceptual del *LDV*, es decir, había que convertir la lista de palabras inglesas a un inventario de unidades conceptuales interlingüísticas. Desde el principio era evidente que no se trataba de una asignación unívoca, por lo que se realizaron varias tareas para: (1) rechazar unidades léxicas (esto es, algunas palabras del *LDV* no se asignaron a conceptos básicos, sino a conceptos terminales) y (2) realizar grupos cognitivos (esto es, se agruparon algunas palabras del *LDV* en el mismo concepto básico). En lo que respecta al rechazo de unidades léxicas, se realizaron las siguientes tareas:

Tarea 1.1. No sólo se rechazaron palabras funcionales (conjunciones, preposiciones, determinantes y pronombres), sino también los partitivos (pieza, cantidad, par, conjunto, etc.), verbos modales y numerales. También se descartaron las instancias léxicas de cuantificación, aspectualidad, temporalidad y modalidad en el *LDV*, ya que este tipo de significado se expresa a través de los operadores de *COREL* (cf. Perrián y Mairal, 2010b).

Tarea 1.2. También se descartaron las palabras con contenido pleno pertenecientes al metalenguaje lexicográfico: “adjetivo”, “artículo”, “gramática”, “sustantivo”, “verbo”, etc. A diferencia de las definiciones de los diccionarios, donde se incluyen referencias al uso y la gramática, los *PS* de *FunGramKB* sólo ofrecen conocimiento semántico.

Tarea 1.3. Cuando dos o más unidades léxicas del *LDV* están relacionadas morfológicamente por el fenómeno de la derivación, se desechan todas excepto una de acuerdo al siguiente criterio de prioridad: Verbo > nombre > adjetivo. Esto es, si tenemos que elegir entre un nombre y un verbo, elegiremos este último (por ejemplo, *advice-advise*, *agreement-agree*, *sale-sell*, etc.). Por el contrario, si la relación tiene lugar entre un adjetivo y un verbo/nombre, rechazamos el adjetivo (por ejemplo, *asleep-sleep*, *successful-success*, etc). Finalmente, cuando nos encontramos con los tres tipos de palabras, entonces escogemos el verbo (por ejemplo, *obedience-obedient-obey*, etc.). De este modo se reduce drásticamente la redundancia, ya que no tiene sentido tener dos conceptos básicos que representan el mismo estado de las cosas, como sucedería en las frases *arsenic is a poison* y *arsenic is poisonous*. Este criterio de prioridad se sustenta en el poder descriptivo de los conceptos en las predicaciones en *COREL*, donde los eventos introducen esquemas cognitivos completos en forma de MT, las entidades son representaciones típicas de los participantes y las cualidades quedan prácticamente restringidas al argumento que tiene función de atributo.

Tarea 1.4. *FunGramKB* describe las cualidades con significados opuestos por su localización en espacios cognitivos, donde se determinan los conceptos positivos y negativos (cf. Periñán y Arcas, 2008). Aquí “positivo” y “negativo” no se refieren a una connotación de significado, sino a la presencia o ausencia del operador de negación en la representación del significado. En otras palabras, el concepto negativo se define por la negación del positivo, por ejemplo *lie* significa *not say the truth*. Si hubiera gradación en la dimensión semántica<sup>48</sup>, todos los conceptos involucrados se definirían en base a dos conceptos centrales, que se determinan a su vez al comparar los índices de frecuencia (se seleccionan los conceptos positivos y negativos de mayor índice) de las unidades léxicas vinculadas a todos los conceptos pertenecientes a la dimensión cognitiva en cuestión. Para el resto de los conceptos de la dimensión semántica se

<sup>48</sup> Una cualidad es graduable (por ejemplo, *cheap*) cuando, para la misma instancia de la entidad, la cualidad puede tener varios grados de intensidad en el tiempo. De otro modo la cualidad no es graduable (por ejemplo, *broken*).

utilizan los operadores de cuantificación *m* (*many/much*) y *p* (*few/little*) para describir los distintos grados de intensidad. Cuando dos o más adjetivos del *LDV* pertenecen a la misma dimensión semántica, el adjetivo considerado el concepto foco positivo se almacena como concepto básico. Encontramos casos similares en los opuestos semánticos *alive-dead*, *male-female*, *true-false*, etc.

De acuerdo a los grupos cognitivos, se desarrollan las siguientes tareas:

Tarea 2.1. Los sinónimos y los casi-sinónimos se aglutinan bajo el mismo concepto (por ejemplo, *answer-reply*, *country-nation*, *do-make*, *say-tell*, etc.).

Tarea 2.2. En el caso de los verbos también se aglutinan los que describen el mismo escenario cognitivo. Este es el caso por ejemplo de *buy* y *sell* que no presentan diferencias en el MT ni en el PS representados en (1) y (2) respectivamente, ya que ambos están vinculados al concepto +SELL\_00. La diferencia semántica radica en el perfil de alguno de sus participantes en el esquema cognitivo: en el caso de *buy*, el *Goal* es el único argumento en la primera actividad, pero el *Agent* define la primera actividad en la *CLS* (*Conceptual Logical Structure*) de *sell* (3). Otro caso similar sería el de *remember-remind*.

- (1) (x1: +HUMAN\_00)Agent (x2)Theme (x3)Origin (x4: +HUMAN\_00) Goal
- (2) +(e1: +GIVE\_00 (x1)Agent (x2)Theme (x3)Origin (x4)Goal  
(f1: (e2: +PAY\_00 (x4)Agent (x5: +MONEY\_00)Theme (x4)Origin  
(x1)Goal))Condition)
- (3) buy: do (x-Goal, [+sell\_00 (x-Goal)]) & inGR +sell\_00 (x-Goal, y-Theme)  
sell: do (x-Agent, [+sell\_00 (x-Agent)]) & inGR +sell\_00 (x-Agent, y-  
Goal, z-Theme)

**2. La fase de jerarquización.** La ontología de *FunGramKB* contiene a su vez tres subontologías cuyos metaconceptos son *#ENTITY*, *#EVENT* y *#QUALITY*, que permiten la organización interna de nombres, verbos y adjetivos respectivamente. Por tanto, los conceptos de la fase anterior se distribuyeron en estas tres subontologías, y después se organizaron jerárquicamente de acuerdo a su relación *IS-A*. En este proceso fue necesario introducir algunos “conceptos paraguas”, principalmente entidades, para explotar de manera más eficiente el mecanismo de herencia en el nivel medio de la ontología. Encontramos un ejemplo de “conceptos paraguas” en la taxonomía de animales donde se introdujeron los conceptos básicos

+CRUSTACEAN\_00, +MAMMAL\_00, +MOLLUSK\_00, +INVERTEBRATE\_00, +REPTILE\_00 y +VERTEBRATE\_00. Además, como todos los conceptos subordinados han de compartir el mismo PS de sus conceptos superordinados (principio de similaridad), se dotó a los conceptos de MT y PS para comprobar este compromiso ontológico (cf. Perinián y Arcas 2010b). En el caso de las entidades, que tienen una taxonomía más profunda, se aplicó el método de evaluación de ontologías *OntoClean* (Guarino y Welty, 2000a, 2000b, 2002; Welty y Guarino, 2001) por el que meta-propiedades formales como “rigidez”, “identidad”, “unidad” y “dependencia” pueden ayudar al ingeniero ontológico a utilizar una relación de subsunción más estricta. A estas meta-propiedades se las puede definir de la siguiente manera:

- (i) Una propiedad es “rígida” si es esencial para la entidad, esto es, no puede cambiar en cualquier instancia de la entidad.
- (ii) Una propiedad es de “identificación” si es única para la instancia al completo, distinguiendo a una instancia específica de cierta clase de otras instancias de esa clase.
- (iii) Una propiedad conlleva “unidad” si hay una relación unificadora común tal como que todas las instancias de una propiedad son completas bajo esa relación.
- (iv) Una propiedad es “dependiente” si cada instancia de la propiedad implica la existencia de otra entidad.

En la figura 2.52 se muestran las propiedades “rigidez” (R), “identidad” (I, O), “unidad” (U) y “dependencia” (D) que se combinan para formar los siguientes tipos semánticos (Guarino y Welty 2000a; Welty y Guarino, 2001: 63).

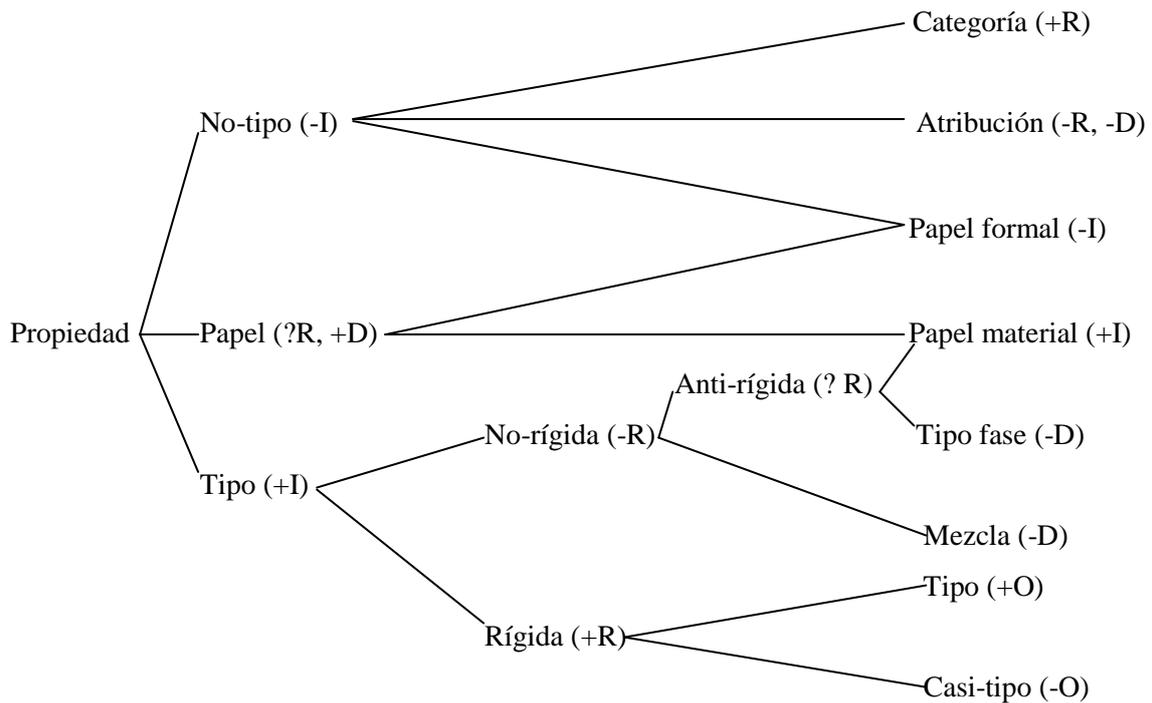


Figura 2.52. Tipología de propiedades basada en las meta-propiedades de *OntoClean* (adaptado de Perrián y Mairal, 2011: 26).

Dado que las meta-propiedades imponen limitaciones en la relación de subsunción, estos tipos formales también imponen restricciones tales como “los tipos sólo se pueden subsumir por categorías o casi-categorías”, y “los tipos fase se subsumen por los tipos”. Como en *OntologyWorks* (Guarino y Welty, 2002: 65), *FunGramKB* contiene un verificador que comprueba la consistencia de la ontología automáticamente una vez que se ha dotado a los conceptos de propiedades formales.

**3. La fase de remodelado.** En esta fase algunos conceptos básicos se degradaron a “subconceptos”, lo cual afectó a la jerarquía ontológica. Tal fue el caso de *-BURY* (6), *-DRESS* (7), *-FLOOD* (8) y *-PAINT* (9) que se establecieron como subconceptos de *+COVER\_00*, cuyo MT y PS podemos ver en (4) y (5) respectivamente, porque, aunque comparten el mismo PS (5), presentan diferencias en las preferencias de selección del argumento *Theme* en sus MTs:

(4) (x1)Agent (x2)Theme (x3)origin (x4)Goal

(5) +(e1: +PUT\_00 (x1)Agent (x2)Theme (x3)origin (x4)Goal

(f2: +ON\_00)Position(f1)instrument (f3: (e2: +HiDE\_00 (x1)Theme

(x4)Referent))Purpose ^ (f4: (e3: +PROTECT\_00 (x1)Theme

(x4)Referent))Purpose)

- (6) –BURY\_00: (x1: +HUMAN\_00 ^ +ANIMAL\_00)Agent (x2: +GROUND\_00 | +LEAF\_00 | +STONE\_00) Theme (x3)origin (x4)Goal
- (7) –DRESS\_00: (x1: +HUMAN\_00)Agent (x2: +CLOTHING\_00) Theme (x3)origin (x4: +HUMAN\_00 ^ +ANIMAL\_00)Goal
- (8) –FLOOD\_00: (x1)Agent (x2: m +WATER\_00)Theme (x3)origin (x4)Goal
- (9) –PAINT\_00: (x1)Agent (x2: +PAINT\_00)Theme (x3)origin (x4)Goal

Así, +COVER\_00 funciona de superconcepto de los subconceptos –BURY, –DRESS, –FLOOD y –PAINT. Los subconceptos no son unidades conceptuales plenas sino refinamientos léxicos de las preferencias de selección en el MT de conceptos básicos o terminales ya existentes. Los subconceptos comparten todas las propiedades conceptuales de sus superconceptos, excepto algunas de las preferencias de selección de sus MTs (incluso debe ser idéntico el número y el papel de los participantes en esos MTs).

**4. La fase de refinamiento.** Esta fase ampliará el ciclo de vida de la Ontología y consistirá principalmente en quitar los conceptos básicos que han resultado muy poco productivos. Es decir, si un concepto básico determinado no forma parte de los PS de un gran número de conceptos, entonces ese concepto se fusiona en su superordinado o se le degrada a concepto terminal, dependiendo de la presencia o ausencia de los conceptos subordinados respectivamente. Cabe señalar que esta fase se aplicó cuando el nivel terminal de *FunGramKB* ya estaba considerablemente poblado.

Como resultado de estas cuatro fases se obtuvo un inventario de aproximadamente 1.300 conceptos básicos que ha servido de base para poblar la ontología de conceptos terminales, proceso que sigue aún en desarrollo. En el momento de realización de esta tesis, la ontología nuclear de *FunGramKB* constaba ya de 402 eventos, 932 entidades y 320 cualidades.

De las distintas metodologías existentes en la creación de ontologías presentadas aquí podemos concluir que es difícil encontrar una metodología que combine todos los procesos implicados en el proceso de desarrollo y de evaluación de ontologías. De todas ellas *METHONTOLOGY* y *COHERENT* son las más completas ya que incluyen tanto una fase de refinamiento como de evaluación de la ontología.

## **2.6.2. Métodos de evaluación de ontologías**

El contenido de las ontologías (esto es, definiciones de los conceptos, taxonomía y axiomas) ha de ser evaluado antes de (re)utilizarlo en otras ontologías o aplicaciones. Los primeros trabajos sobre evaluación de ontologías datan de mediados de los años noventa (Gómez, 1994a; 1994b). En los últimos años ha crecido el interés por este tema. Los principales esfuerzos en la evaluación de ontología han sido de Gómez (1996; 2004) en la metodología *METHONTOLOGY* y los de Welty y Guarino (2001) en el método *OntoClean*.

### **2.6.2.1. Terminología de evaluación de ontologías**

En su trabajo Gómez (1994b) estableció las líneas principales en la evaluación de ontologías, que incluían: (1) la división de la evaluación en dos tipos: técnica (llevada a cabo por los desarrolladores) y la evaluación del usuario; (2) la provisión de un conjunto de términos para el proceso de evaluación de la ontología y la definición estándar de esos términos; (3) la definición de un conjunto de criterios para llevar a cabo los procesos de evaluación tanto técnicos como del usuario; (4) la inclusión de las actividades de evaluación en metodologías de construcción de ontologías; (5) la construcción de herramientas de evaluación de ontologías ya existentes; y (6) la inclusión de módulos de evaluación en herramientas utilizadas para construir ontologías.

El propósito principal del proceso de evaluación de ontologías es determinar qué define la ontología correctamente, qué no y qué hace de manera incorrecta. Para evaluar el contenido de una ontología determinada Gómez (1996) identificó los siguientes criterios: consistencia, integridad y concisión.

**Consistencia** se refiere a si es posible obtener conclusiones contradictorias de las definiciones válidas de la ontología. Una definición determinada es consistente si y sólo si la definición individual es consistente y no se puede inferir conocimiento contradictorio de otras definiciones y axiomas.

**Integridad.** La falta de integridad es un problema fundamental en las ontologías, incluso más cuando están disponibles en un entorno tan abierto como la Web Semántica. De hecho, no podemos probar la integridad total de una ontología o de sus

definiciones, pero sí la falta de integridad de definiciones individuales, y, por tanto, podemos deducir así la falta de integridad de la ontología si al menos falta una definición en el marco de referencia establecido. Así, una ontología es íntegra si y solo si:

- Todo el conocimiento que se supone debe estar incluido en la ontología está incluido de forma explícita o se puede inferir.
- Cada definición es completa. Esto se establece de dos maneras: (a) qué conocimiento especifica la definición y si se hace una definición explícita del mundo; y (b) si el conocimiento requerido pero no explícito se puede inferir de otras definiciones o axiomas. Si es así, entonces la definición es completa, de lo contrario no lo será.

**Concisión.** Una ontología es concisa si: (a) no almacena definiciones innecesarias o inútiles, (b) no existe redundancia en las definiciones, y (c) las redundancias no se pueden inferir de otras definiciones o axiomas.

#### **2.6.2.2. Evaluación de la taxonomía**

Gómez (2004: 261) también identificó y clasificó diferentes tipos de errores en las taxonomías. Elaboró una lista de errores posibles en el modelado de conocimiento de taxonomías en una ontología basada en marcos. Estos errores se clasifican en: inconsistencia, falta de integridad y errores de redundancia, como se muestra en la figura 2.53. Recomienda que los mecanismos de control se realicen durante la construcción de la taxonomía.

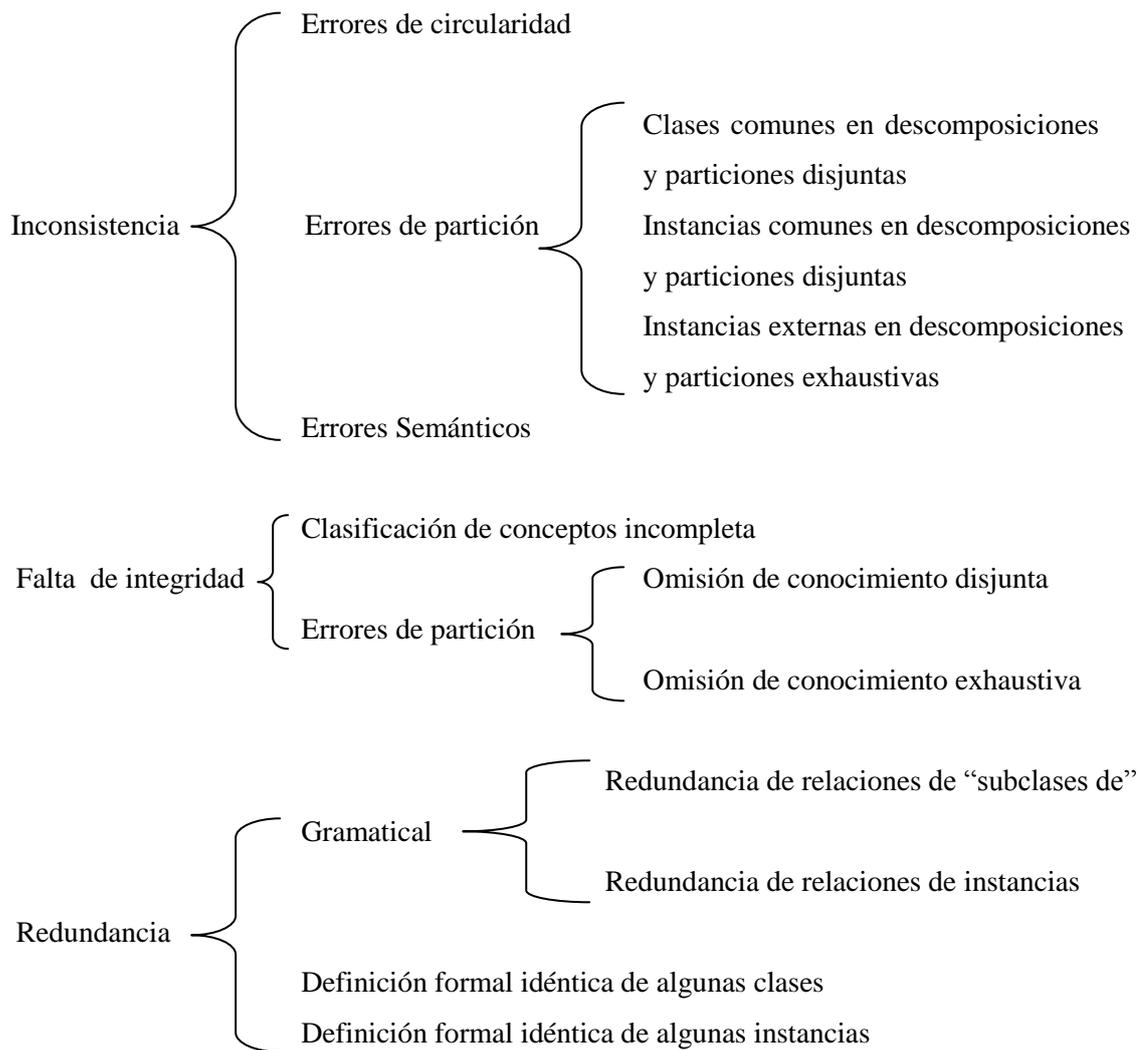


Figura 2.53. Tipos de errores susceptibles de realizar cuando se desarrollan taxonomías con marcos (adaptado de Gómez, 2004: 180).

### 2.6.2.3. *OntoClean*

*OntoClean*<sup>49</sup> es un método de evaluación de ontologías desarrollado por el grupo de investigación ontológica del CNR en Padua (Italia). Tiene como propósito eliminar de las taxonomías las relaciones “subclase de” erróneas de acuerdo a algunas nociones filosóficas como “rigidez”, “identidad” y “unidad”. Estas nociones se pueden aplicar a las propiedades, pero también se pueden extender a los conceptos. Anteriormente explicamos las nociones de “rigidez” e “identidad” en 2.5.2. A continuación

---

<sup>49</sup> <http://www.ontoclean.org/>

comentamos las cinco actividades que propone *OntoClean* para limpiar las taxonomías (Welty y Guarino, 2001).

**Actividad 1. Asignación de meta-propiedades a cada propiedad.** En esta actividad se propone asignar valores a las meta-propiedades de cada concepto de la taxonomía. Los símbolos a utilizar por el ingeniero ontológico son: “rígido” (+*R*), “anti-rígido” (~*R*), “no rígido” (-*R*), “implica el criterio de identidad” (+*I*), “ofrece el criterio de identidad” (+*O*), “implica unidad” (+*U*), “anti-unidad” (~*U*). Aunque (-*U*) puede utilizarse para cualquier concepto que no implica unidad, se reserva para los conceptos que ni implican unidad ni anti-unidad. Las *ontologías de alto nivel de universales y particulares* presentadas en la sección 2.5.2. se pueden utilizar como ejemplo de clasificación de estas meta-propiedades.

**Actividad 2. Centrarse solamente en las propiedades rígidas.** Esta actividad tiene como propósito identificar las partes esenciales de la taxonomía. A la taxonomía que sólo tiene propiedades rígidas se le llama “taxonomía vertebral” porque es la base del resto de la taxonomía.

**Actividad 3. Evaluación de la taxonomía de acuerdo a principios basados en meta-propiedades.** Para desarrollar esta actividad Welty y Guarino (2001) hablan de tres propiedades relacionadas con la identificación de conflictos entre criterios de identidad y entre anti-unidad y unidad.

*Regla 1: Comprobar la identidad*

De acuerdo a esta regla, dos propiedades con criterios de identidad incompatibles tienen individuos comunes. Los individuos heredarán dos criterios de identidad incompatibles. Así, los conceptos candidatos que podrían violar la regla son:

- Conceptos que no son disjuntos.
- Conceptos que implican identidad (+*I*).

Además, si hay incompatibilidad entre los criterios de identidad implicados en ambos conceptos, podemos afirmar que los dos conceptos no cumplen la regla.

*Regla 2: Comprobar la unidad*

Las propiedades que llevan criterios de unidad incompatibles son necesariamente disjuntas. De acuerdo a esta regla, dos propiedades que llevan criterios de unidades incompatibles no pueden tener individuos comunes. Si dos conceptos con criterios de unidad incompatibles tienen individuos comunes, los

individuos heredarán dos criterios de unidad que son incompatibles. Por tanto los conceptos candidatos que podrían violar la regla son:

- Conceptos que no son disjuntos.
- Conceptos que implican unidad (+U).

Si hay incompatibilidad entre los criterios de unidad de los dos conceptos, entonces podemos afirmar que los dos conceptos no cumplen la regla.

*Regla 3: Comprobar la unidad y la anti-unidad*

Una propiedad que implica anti-unidad ( $\sim U$ ) tiene que ser disjunta con una propiedad que implica unidad (+U). Si sabemos que  $P_1$  implica anti-unidad ( $\sim U$ ), ningún individuo común será completo bajo una relación unificadora común. Si sabemos que  $P_2$  implica unidad (+U), todos los individuos comunes serán completos bajo una relación unificadora común (criterio de unidad). En otras palabras, cualquier individuo común será completo o no, lo cual es una contradicción. Los conceptos que podrían violar la regla son:

- Conceptos que no son disjuntos, y
- Uno de los conceptos que implica anti-unidad ( $\sim U$ ) y otro que implica unidad (+U).

**Actividad 4. Considerar las propiedades no-rígidas.** Cuando se ha examinado la taxonomía vertebral, el ingeniero ontológico ha de evaluar las propiedades no-rígidas (-R) y rígidas (+R). Una regla útil para ello es la siguiente:

*Regla 4: Comprobar la rigidez (+R) y antirigidez ( $\sim R$ ).*

La propiedad de anti-rigidez no puede subsumir a una propiedad rígida, lo que significa que un concepto rígido no puede ser subclase de uno anti-rígido. Por tanto, los conceptos que podrían violar la regla son:

- El concepto padre tiene la etiqueta  $\sim R$ .
- El concepto hijo tiene la etiqueta +R.

**Actividad 5. Completar la taxonomía con otros conceptos y relaciones.** Podemos introducir nuevos conceptos por diversas razones. Una es la transformación de los conceptos en relaciones. Al introducir conceptos nuevos se hace necesario repetir todo el proceso desde la actividad 1.

*OntoClean* no es sólo útil para limpiar la taxonomía de conceptos, sino también para hacer explícitos los compromisos ontológicos subyacentes en las definiciones de los términos de la ontología.

Aunque la evaluación de ontologías es un área en pleno desarrollo actualmente, desde una perspectiva metodológica se hace necesaria la implantación de actividades de evaluación en el desarrollo de la ontología para acreditar la consistencia de la misma.

### 2.6.3. Lenguajes de implementación de ontologías

Los lenguajes de implementación de ontologías comenzaron a surgir a comienzos de los años noventa dentro del campo de la inteligencia artificial como evolución de los lenguajes de representación del conocimiento. Los paradigmas de representación del conocimiento que albergaban tales lenguajes ontológicos se basan en lógica de primer orden como *KIF*, en marcos combinados con lógica de primer orden, como *CycL*, *Ontolingua*, *OCML* y *FLogic*, y en lógica descriptiva (*LOOM*). También se creó *OKBC* como protocolo de acceso a ontologías implementadas en otros lenguajes de marcos. La figura 2.54 presenta un esquema general de estos lenguajes (ver Gómez y otros 2004: 200).

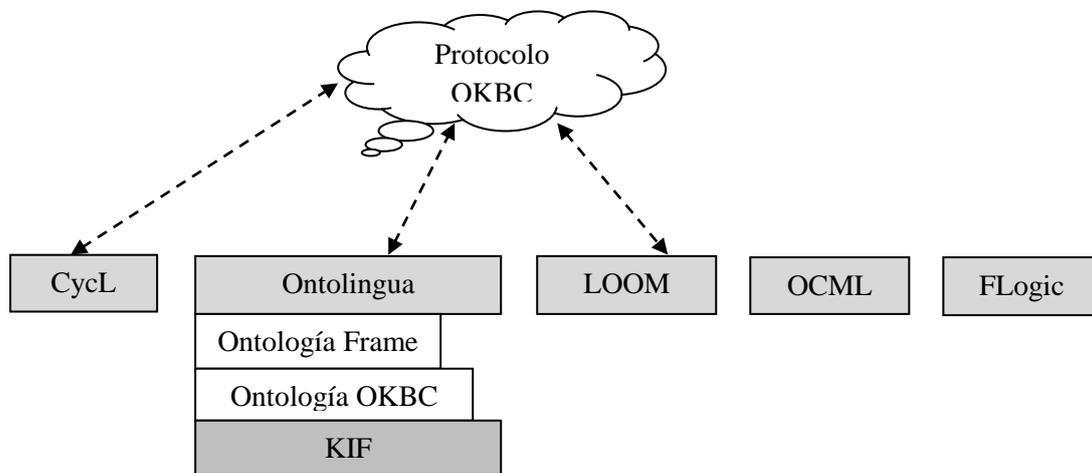


Figura 2.54. Lenguajes tradicionales (adaptación de Gómez y otros, 2004: 200).

De todos estos lenguajes, el primero que se creó fue *CycL* (Lenat y Guha, 1990) que utiliza marcos y lógica de primer orden y se utilizó en la creación de la ontología *Cyc*. *KIF* (*Knowledge Interchange Format*) (Genesereth y Fikes, 1992) se creó después, en 1992, como formato de intercambio de conocimiento. Como era difícil crear ontologías directamente en *KIF*, se creó *Ontolingua* (Farquhar y otros, 1997) por encima de ésta. Así, *Ontolingua* construye sobre *KIF*, que es el lenguaje compatible con el servidor de

*Ontolingua*. Lo que es más, en los años noventa se consideró a *Ontolingua* como el estándar *de facto* en la comunidad ontológica.

*LOOM* (MacGregor, 1991) se construía a la vez pero no para implementar ontologías, sino para bases de conocimiento generales. Se basa en la lógica descriptiva y en reglas de producción, a la vez que ofrece la clasificación automática de conceptos. *OCML* (Motta, 1999) se desarrolló después, en 1993, como un tipo de “*Ontolingua* operacional”. De hecho, un gran número de definiciones en *OCML* son muy similares a sus correspondientes en *Ontolingua*. *OCML* se construyó para desarrollar ontologías y modelos de métodos de resolución de problemas. Finalmente, en 1995 se desarrolló *FLogic* (Kifer y otros, 1995) como lenguaje que combina marcos y lógica de primer orden. Por otro lado, en 1997 se creó el protocolo *OKBC* (*Open Knowledge Base Connectivity*) (Chaudhri y otros, 1998) que permite el acceso a las bases de conocimiento almacenadas en sistemas de representación del conocimiento distintos, que pueden a su vez estar basados en paradigmas diferentes de representación del conocimiento. De las lenguas mencionadas anteriormente, *Ontolingua*, *LOOM* y *CycL* son compatibles con *OKBC*.

Por otro lado, con el auge de Internet se crearon también lenguajes de implementación de ontologías que explotan las características de la Web. A estos lenguajes se les conoce como lenguajes de ontologías basados en la Web o lenguajes de marcado de ontologías. Su sintaxis se basa en lenguajes de marcado existentes como *HTML* (Raggett y otros, 1999) y *XML* (Bray y otros, 2000), cuyo propósito es no sólo el desarrollo ontológico, sino también la presentación de datos y el intercambio de datos respectivamente (Gómez y otros, 2008). Las relaciones entre estos lenguajes aparecen en la figura 2.55.

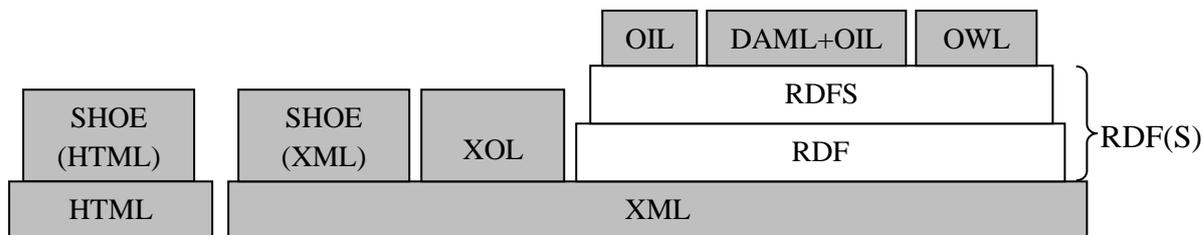


Figura 2.55. Lenguajes de marcado ontológico (en Gómez y otros, 2004: 201).

Los ejemplos más destacados de estos lenguajes son: *SHOE* (Luke y Heflin, 2000), que combina marcos y reglas. Se construyó en 1996 como una extensión de *HTML*. Utilizaba un etiquetado diferente al de *HTML*, con lo cual permitía la inserción de ontologías en documentos *HTML*. Más adelante se adaptó esta sintaxis a *XML*.

Los otros lenguajes se basan en *XOL* (Karp y otros, 1999). Este lenguaje se desarrolló como *XMLization* de un pequeño subconjunto de primitivos del protocolo *OKBC*, llamado *OKBC-Lite*. Por su parte *RDF* (Lassila y Swick, 1999) fue desarrollado por el *W3C (the World Wide Web Consortium)* como lenguaje basado en la red semántica para describir recursos Web. Comenzó su desarrollo en 1997 y *RDF* se propuso como recomendación del *W3C* en 1999. El lenguaje *RDF Schema* (Brickley y Guha, 2000) también fue construido por el *W3C* como extensión del *RDF* con primitivos basados en marcos. A la combinación del *RDF* y el *RDF Schema* se la conoce como *RDF(S)*. Este lenguaje sólo sirve para representar conceptos, taxonomías de conceptos y relaciones binarias. Recientemente se está desarrollando la ontología *WSML (Web Service Modelling Language)*<sup>50</sup> en el marco de *WSMO (Web Service Modelling Ontology)*<sup>51</sup> (Bruijn, 2006). Tiene como propósito servir de representación de los servicios Web semánticos además de representar ontologías. Está dividido en capas, cada una de ellas se basa en un formalismo de representación del conocimiento distinto, ya sea lógica descriptiva, programación lógica o lógica de primer orden.

Estos lenguajes han marcado los fundamentos de la Web Semántica (Berners-Lee, 1999). En este contexto se han desarrollado otros tres lenguajes como extensiones de *RDF(S)*: *OIL*, *DAML+OIL* y *OWL*. *OIL (Ontology Interchange Language)* (Fensel y otros, 2000) se desarrolló a comienzos del 2000 en el marco del proyecto *On-To-Knowledge*<sup>52</sup>. Añade primitivos basados en marcos y su semántica formal se basa en lógica descriptiva. Se basa en las propuestas de *OKBC*, *XOL* y *RDF(S)*. *DAML+OIL* (van Harmelen y otros, 2001) se creó después en el contexto del proyecto *DAML*<sup>53</sup>. En 2001, el *W3C* formó un grupo de investigación llamado *Web-Ontology (WebOnt)*. Este

<sup>50</sup> <http://www.wsmo.org/wsml/>

<sup>51</sup> <http://www.w3.org/Submission/WSMO/>

<sup>52</sup> <http://www.ontotext.com/research/otk>

<sup>53</sup> <http://www.daml.org/>

grupo tenía como objetivo la creación de un nuevo lenguaje de marcado ontológico para la Web Semántica. Como resultado crearon el lenguaje *OWL (Ontology Web Language)* (Dean y Schreiber, 2004). *OWL* cubre la mayoría de los rasgos de *DAML+OIL* y amplía la expresividad de los primitivos que ya aparecían en ese lenguaje, lo que permite representar expresiones complejas y describir conceptos y relaciones. *OWL* está dividida en tres capas (*Lite*, *DL* y *Full*) que sirven para representar distintos niveles de expresividad.

De todos los lenguajes mencionados, sólo algunos como *Ontolingua*, *OCLM*, *FLogic*, *RDF(S)* y *OWL (OIL y DAML+OIL* también respaldan esta noción pero ya no están activos) están lo suficientemente equipados con primitivos semánticos como para explotar las ontologías en red.

Por último destacar que la suite *FunGramKB*<sup>54</sup> en la que se enmarca el proyecto al que se adscribe la presente tesis, implementa el lenguaje de representación *COREL (Conceptual Representation Language)* común a los tres niveles de conocimiento representados en la base de conocimiento (descritos en el capítulo 3). Como lenguaje de representación conceptual, *COREL* permite definir todas las unidades conceptuales de la ontología nuclear a través de los mecanismos de herencia e inferencia sobre los Postulados de Significado (PS) (cf. Perrián y Arcas, 2005). Como es común a los lenguajes de representación, Perrián y Mairal (2010b: 16) distinguen también dos componentes:

- 1) Una colección de unidades conceptuales que se organizan jerárquicamente para permitir el mecanismo de herencia y así facilitar la brevedad definitoria a la vez que se aumenta el poder expresivo.
- 2) Un sistema de notación, es decir, la sintaxis a la que se tienen que ajustar las representaciones interlingüísticas para poder considerarlas estructuras bien formadas.

---

<sup>54</sup> <http://www.fungramkb.com>

El poder expresivo de la ontología nuclear de *FunGramKB* viene marcado por el hecho de que con un catálogo limitado de conceptos básicos se pueden definir todos los conceptos terminales de la ontología.

Recientemente se ha comprobado que tales componentes son también aplicables de forma satisfactoria tanto a sistemas de traducción automática (Periñán y Mairal, 2010a) como a ontologías satélite vinculadas a la ontología nuclear de *FunGramKB* (Faber y otros, 2011).

#### **2.6.4. Herramientas de desarrollo de ontologías**

Las herramientas de desarrollo de ontologías han sufrido una importante evolución desde su aparición a mediados de los noventa, según la cual podemos distinguir dos grupos diferentes (Gómez y otros, 2004; 2008):

- Herramientas cuyo modelo de conocimiento se corresponde completamente con un lenguaje de representación de ontologías. Así, estas herramientas se desarrollan para un lenguaje específico. En este grupo se incluyen: *Ontolingua Server* (Farquhar y otros, 1997), que sirve de base para la construcción de ontologías con *Ontolingua* y *KIF*; *Ontosaurus Server* (Benton y Kambhampati, 1997; Swartout y otros, 1997), que utiliza el lenguaje *Loom*; *WebOnto* (Domingue, 1998), que utiliza *OCML*; y *OilEd* (Bechhofer y otros, 2001), primero con *OIL*, luego con *DAML+OIL* y ahora con *OWL*.
- Suites de herramientas integradas cuya característica principal es que tienen una arquitectura extensible y cuyo modelo de conocimiento es normalmente independiente del lenguaje de representación de ontologías. Además, estas herramientas ofrecen un conjunto básico de servicios relacionados con las ontologías y pueden ampliarse fácilmente con otros módulos para ofrecer más funciones. A este grupo pertenecen *Protégé-2000* (Noy y otros, 2000), *WebODE* (Arpírez y otros, 2003), *OntoEdit* (Benton y Kambhampati, 2002; Sure y otros, 2000), y *KAON* (Maedche y otros, 2003).

Como hemos señalado anteriormente, las herramientas del primer grupo están íntimamente relacionadas con el lenguaje de representación de ontologías para el que se desarrollan. Así, estas herramientas se desarrollaron para permitir la edición y exploración de ontologías en sus correspondientes lenguajes y la importación y exportación de ontologías de/hacia otros lenguajes de ontologías, pero requieren que el usuario tenga conocimientos suficientes de los lenguajes subyacentes.

El *Ontolingua Server* (Farquhar y otros, 1997) fue la primera herramienta de desarrollo de ontologías. Apareció a mediados de los noventa y se construyó para facilitar el desarrollo de las ontologías *Ontolingua* con una interfaz en forma Web. Inicialmente la primera aplicación en el servidor de *Ontolingua* fue el editor de ontologías. Después se incluyeron otros sistemas, como *Webster*, un solucionador de ecuaciones, un servidor *OKBC*, etc.

Paralelo al desarrollo del servidor de *Ontolingua*, se implementó *OntoSaurus* (Swartout y otros, 1997) como editor Web y explorador para las ontologías *LOOM*. *OntoSaurus* consta de dos módulos principales: un servidor de ontologías, que utiliza el sistema de representación del conocimiento<sup>55</sup> incluido en el lenguaje *LOOM* y el navegador Web, que permite editar y explorar las ontologías *LOOM* con forma *HTML*.

En 1997 apareció *WebOnto* (Domingue, 1998). Se trata de un editor de ontologías para las ontologías *OCML*. La principal diferencia con las otras ontologías reside en el hecho de que el editor de ontologías no está basado en *HTML*, sino en *Java applets*. Sin embargo, su gran ventaja con respecto a las otras herramientas de desarrollo de ontologías era su gran apoyo en la edición colaborativa de ontologías, que permitió discusiones sincrónicas y asíncronas sobre las ontologías construidas por grupos de usuarios.

---

<sup>55</sup> Levesque (1984) define un sistema de representación del conocimiento como un servidor compatible con dos tipos de operaciones: dotar al sistema información adicional y realizar preguntas. También señala que, a diferencia de otros sistemas, los sistemas de representación del conocimiento utilizan no sólo la información ofrecida explícitamente, sino también la información inferida a través de las respuestas a las preguntas.

*OilEd* (Bechhofer y otros, 2001) se desarrolló en 2001 como editor de las ontologías *OIL*. Con la creación de *DAML+OIL*, *OilEd* se adaptó primero para gestionar las ontologías *DAML+OIL* y luego para *OWL*. *OilEd* ofrece funciones de comprobación de consistencia y clasificaciones de conceptos automáticas por medio del motor de inferencias *FaCT*, aunque también acepta otros motores de inferencias de lógica descriptiva como *RACER*.

Recientemente se ha desarrollado una nueva generación de entornos de ingeniería ontológica. Se construyen como suites de herramientas integradas que ofrecen apoyo tecnológico para una gran variedad de actividades del proceso de desarrollo de ontologías. Para tal fin, constan de una arquitectura extensible basada en componentes, donde se pueden añadir fácilmente nuevos módulos que ofrecen más funciones a la suite. Además, los modelos de conocimiento que subyacen a estos entornos son normalmente independientes de la lengua y ofrecen traducciones de y hacia varias lenguas y formatos.

*Protégé-2000* (Noy y otros, 2000) es una aplicación autónoma de código abierto con arquitectura extensible. El núcleo de *Protégé-2000* es un editor de ontologías, que es extensible a través de *plug-ins* que dotan al entorno de más funciones, tales como importación y/o exportación de lenguajes de ontologías (*FLogic*, *Jess*, *OIL*, *XML*, *Prolog*), diseño de lenguajes (Knublauch y otros, 2004), acceso a *OKBC*, creación y ejecución de restricciones (*PAL*), fusión de ontologías (Noy y otros, 2000), etc.

*WebODE* (Arpírez y otros, 2003) es también una suite de arquitectura extensible basada en un servidor de aplicaciones, cuyo desarrollo comenzó en 1999. El núcleo de *WebODE* es su servicio de acceso ontológico que es utilizado por todos los servicios y aplicaciones conectados al servidor. El editor de ontologías de *WebODE*, que permite editar y explorar las ontologías *WebODE*, está basado en *HTML* y *Java applets*. El entorno de trabajo integra servicios que permiten la importación y/o exportación de lenguajes de ontologías (*XML*, *RDF(S)*, *OIL*, *DAML+OIL*, *OWL*, *CARIN*, *FLogic*, *Jess*, *Prolog*), la edición de axiomas con *WAB* (*WebODE Axiom Builder*), para documentación, evaluación, evolución, aprendizaje, fusión y motor de inferencias. *WebODE* puede interoperar con *Protégé-2000*.

*OntoEdit* (Benton y Kambhampati, 2002; Sure y otros, 2002) es un entorno extensible y flexible basado en una arquitectura de extensiones. Su editor de ontologías es una aplicación autónoma que permite editar y explorar ontologías e incluye funciones de construcción de ontologías colaborativas, inferencias, manejo de lexicones de dominio, etc. Este editor exporta e importa ontologías en diferentes formatos (*XLM*, *FLogic*, *RDF(S)* y *DAML+OIL*). Hay dos versiones: la gratuita, *OntoEdit Free* y la profesional, *OntoEdit Professional*, cada una con un conjunto de funciones diferente.

La suite de herramientas *KAON* (Maedche y otros, 2003) es un entorno de ingeniería ontológica de código abierto y extensible. El núcleo de esta herramienta es la ontología *API*, que define el modelo de conocimiento subyacente basado en extensiones de *RDF(S)*. *OI modeler* es el editor de ontologías encargado de la evolución de las ontologías, de encontrar las correspondencias entre ellas y de generar ontologías a partir de bases de datos, etc.

A modo de conclusión señalar que sólo *OntoEdit* y *WebODE* ofrecen metodologías para construir ontologías (*On-To-Knowledge* y *METHONTOLOGY* respectivamente), aunque esto no impide que puedan utilizar otras metodologías.

Desde el paradigma de la representación del conocimiento, las herramientas *KAON* se basan en redes semánticas y marcos. Por su parte, *WebODE*, *Protégé*, *OntoEdit* y *KAON* pueden representar conocimiento de acuerdo a un sistema combinado de marcos y lógica de primer orden. En cuanto a la expresividad del modelo de conocimiento subyacente en la herramienta, todas las herramientas pueden representar clases, relaciones, atributos e instancias, pero sólo *KAON* y *Protégé* tienen componentes flexibles (metaclases). A la hora de decidir sobre qué herramienta de desarrollo de ontologías elegir, es importante conocer, entre otras, si consta de mecanismos para examinar restricciones y la consistencia del tipo de herencia (sencilla, múltiple, monotónica o no-monotónica). De todas estas herramientas la única que no tiene motor de inferencias es *KAON*. En cuanto a la evaluación de ontologías, *Protégé* y *WebODE* tienen un módulo de evaluación de *OntoClean* (Guarino y Welty, 2002; Welty y Guarino, 2001). Finalmente, *Protégé* (con las extensiones de *OWL*) se conecta a un razonador de lógicas descriptivas para realizar clasificaciones automáticas.

En cuanto a la cooperación en la construcción de ontologías, *Protégé* incorpora funcionalidades de sincronización. A este respecto, *XML.com*<sup>56</sup> ofrece un inventario de herramientas de desarrollo de ontologías.

La ventaja que presenta la base de conocimiento *FunGramKB* con respecto a otras es que consta de su propia metodología y herramientas de desarrollo de ontologías, además de cumplir unos compromisos ontológicos mínimos: universalidad y motivación lingüística, modelo ontológico en 3 niveles, unidades conceptuales no-atómicas, los PS como organizadores de la ontología, relación taxonómica IS-A, herencia múltiple y herencia no monotónica (Periñán y Arcas, 2010b). Así, la metodología *COHERENT* y las herramientas de desarrollo de ontologías (*FunGramKB Term Extractor*, lenguaje de representación *COREL*, y el editor de ontologías) utilizadas para el desarrollo de la ontología nuclear son reutilizables en el desarrollo de ontologías satélite, como mostraremos en el capítulo 4.

## 2.7. Estado de la cuestión de las ontologías legales

Como hemos señalado anteriormente en este capítulo, podemos clasificar las ontologías de acuerdo a diferentes criterios: propósito, tema, nivel de generalidad, nivel de formalidad o riqueza en su estructura interna, además de por la metodología, las herramientas y el lenguaje utilizados en su desarrollo.

Las ontologías aquí presentadas son especialmente relevantes en la historia del desarrollo de las ontologías legales como así atestigua la literatura al respecto. En base a los criterios, metodología, lenguajes y herramientas anteriormente presentados en este capítulo, analizaremos las ontologías legales existentes hasta el momento de acuerdo a los siguientes criterios:

1. Uso y función de la ontología.
2. Tema de la ontología.
3. Nivel de generalidad.

---

<sup>56</sup> <http://www.xml.com/pub/a/2004/07/14/onto.html>

4. Nivel de formalidad en su estructura interna.
5. Metodología, herramientas y lenguaje de representación utilizados en el desarrollo de la ontología.
6. Si es monolingüe, bilingüe o multilingüe.
7. Posibles aplicaciones.
8. Otras características: disponibilidad, coherencia, reutilización, etc.

A continuación presentamos en detalle el estado de la cuestión de las ontologías legales más significativas hasta el momento.

Según Breuker y otros (2009b: 11), las ontologías legales representan conocimiento legal. El problema es definir qué entendemos por conocimiento legal. A mediados de los noventa, los trabajos de McCarty, Stamper, van Kralingen, Visser, Breuker, Winkels y Valente pretendían unir conceptos computacionales y legales. Entonces las ontologías legales se entendían como el punto de unión entre la teoría legal y la Inteligencia Artificial y el derecho (Valente y Breuker, 1994b). Esos esfuerzos iniciales han dado lugar a ontologías generales más desarrolladas como *LKIF*.

Como señalan Visser y Bench-Capon (1998), las primeras ontologías legales eran de cuatro tipos: (i) de discurso legal (*LLD* de McCarty); (ii) de normativa legal (*NOR(Norma)* de Stamper); (iii) funcional (*FOLaw* de Valente y Breuker) y (iv) de marcos (*FBO* de van Kranlingen y Visser). A estas se les añadieron más tarde bases de conocimiento léxicas más generales de alto nivel basadas en *WordNet* (*JurWordNet* de Gangemi y otros). Finalmente los conceptos legales fundamentales y categorías del sentido común se han fusionado recientemente en la ontología general *LKIF* basada en la ontología *LRI-Core* desarrollada en los noventa.

## 2.7.1. Primeras conceptualizaciones legales

### 2.7.1.1. NORMA

Stamper (1991) señalaba que el uso de la lógica tradicional presentaba importantes problemas semánticos para la representación del conocimiento (legal). Según él, las representaciones simbólicas “confían” de forma inválida en nociones como certeza, individualidad e identidad. Por consiguiente, expresar conocimiento legal en forma de reglas es una simplificación excesiva. Por ello propuso el formalismo *NORMA*, una lógica de normas y permisos que se basaba en dos suposiciones: (1) que no hay conocimiento sin un “conocedor” (Stamper, 1991), y que (2) el conocimiento del “conocedor” depende de su comportamiento (Stamper, 1996). Al utilizar *NORMA* (también conocida como *NOR*) las entidades del mundo se describen por su comportamiento más que por asignarles valores de certeza e individualidad (Visser y Bench-Capon 1998: 35).

Los principales conceptos ontológicos son: (a) “agentes” (organismos que se encuentran en el centro de la realidad), (b) “invariables de comportamiento” (descripción de una situación cuyas características permanecen invariables) y (c) “realizaciones” (una realización de una situación, un agente que desarrolla acciones o la combinación de un agente y una invariable de comportamiento). A continuación los describimos con más detalle.

- (a) Agentes. Un agente es un organismo que se encuentra en el centro de la realidad. Alcanza conocimiento, regula y modifica el mundo a través de acciones. Asume responsabilidades por sus acciones. El concepto agente puede ampliarse hasta incluir grupos, equipos, compañías, agentes sociales o incluso naciones.
- (b) Invariables de comportamiento. Las entidades del mundo se describen como características que permanecen invariables a lo largo del tiempo. También se considera que estas características de comportamiento de las entidades. Una invariable de comportamiento es una descripción (utilizando verbos, sustantivos y adjetivos) de una situación cuyas características permanecen invariables a lo largo del tiempo. Así, una situación denota ampliamente un cierto conocimiento del mundo, por ejemplo un objeto (una copa, una guitarra) o un estado de las cosas (dormir, trabajar).

- (c) Realizaciones. Los agentes desarrollan situaciones al realizar acciones. La realización de una situación se especifica como la combinación de (1) un agente y (2) una invariable de comportamiento (de forma abreviada  $Ax$ ). Un ejemplo de  $Ax$  sería “María duerme”.  $Ax$  significaría que  $A$  (María) tiene la habilidad de realizar  $x$  (dormir). Las invariables de comportamiento se pueden combinar en realizaciones compuestas, como por ejemplo,  $Axy$  ( $A$  no puede realizar  $y$  sin primero realizar  $x$ ),  $A.x.y$ . ( $x$  es parte de  $A$  e  $y$  es parte de  $x$ ).

Para concluir con la descripción de *NORMA* señalar que Visser y Bench-Capon (1998: 32) consideran que no es una ontología propiamente dicha, sino una discusión de suposiciones ontológicas especificadas en un lenguaje semi-formal.

#### **2.7.1.2. LLD**

McCarty (1989) propuso un lenguaje para el discurso legal (*LLD*). Consideró este lenguaje como un primer paso hacia un lenguaje de representación general aplicable al ámbito legal. Para ello planteó el desarrollo de modelos conceptuales profundos del dominio legal y de un lenguaje de representación del conocimiento que representara un conjunto determinado de categorías del sentido común para una aplicación legal determinada, porque, según afirmaba, “hay muchas categorías del sentido común subyacentes a la representación de problemas del dominio legal: espacio, tiempo, masa, acción, permiso, obligación, causalidad, propósito, intención, conocimiento, creencia, etc”.

*LLD* consiste en tres componentes de conocimiento básicos: “fórmulas atómicas” (relaciones de predicación utilizadas para expresar afirmaciones objetivas), “reglas” (se forman al conectar las fórmulas atómicas con conectores lógicos) y “modalidades” (expresiones de segundo orden). Estas modalidades son: “tiempo” (basada en lógica temporal), “eventos y acciones” (basados en una lógica intuitiva) y “expresiones deónticas” permitido ( $P$ : *permitted*), prohibido ( $P$ : *forbidden*), obligatorio ( $O$ : *obligatory*) y posible ( $E$ : *enabled*). Las afirmaciones deónticas se forman por la combinación de un nombre, un operador modal (posiblemente negado), una condición, y una acción (Visser y Bench-Capon, 1998). Para expresar afirmaciones temporales *LLD* reconoce “estados”. Un “estado” es esencialmente la cosificación (temporal) de

una relación de predicación. Las relaciones de predicación pueden cosificarse o con puntos en el tiempo o con intervalos (dos puntos en el tiempo). Los eventos realizan cambios en los estados. Los eventos son elementales o complejos. Una acción es la realización entre un actor y un evento. Con respecto a las afirmaciones deónticas, *LLD* reconoce cuatro operadores deónticos: permitido (*P: permitted*), prohibido (*P: forbidden*), obligatorio (*O: obligatory*) y posible (*E: enabled*). Las afirmaciones deónticas se forman por la combinación de un nombre, un operador modal (posiblemente negado), una condición, y una acción (Visser y Bench-Capon, 1998: 34).

Para concluir con la descripción de *LLD*, señalar que Valente y Breuker (1992) plantean la cuestión de si a este lenguaje del discurso jurídico se le puede clasificar como ontología típica, ya que la gran mayoría de los términos que incluye no son específicos del dominio legal, a lo que Visser y Bench-Capon (1998: 32) señalan que el trabajo de McCarty ha de ser considerado como una conceptualización temprana del dominio legal a través de un lenguaje (semi-)formal.

## **2.7.2. Ontologías legales existentes**

### **2.7.2.1. *FOLaw***

La *Functional Ontology of Law (FOLaw)* de Valente y Breuker (1994b) se basa en una perspectiva funcional del sistema legal, al que se le considera como un instrumento cuya función principal es reaccionar ante el comportamiento social (Valente y Breuker, 1996). Suponen que el conocimiento que especifica cómo funciona el sistema legal conforma las fuentes legales, y que los diferentes papeles que estas fuentes pueden realizar conforman las diferentes funciones que el conocimiento legal puede realizar. En la figura 2.56 se observa el conjunto de las principales funciones.

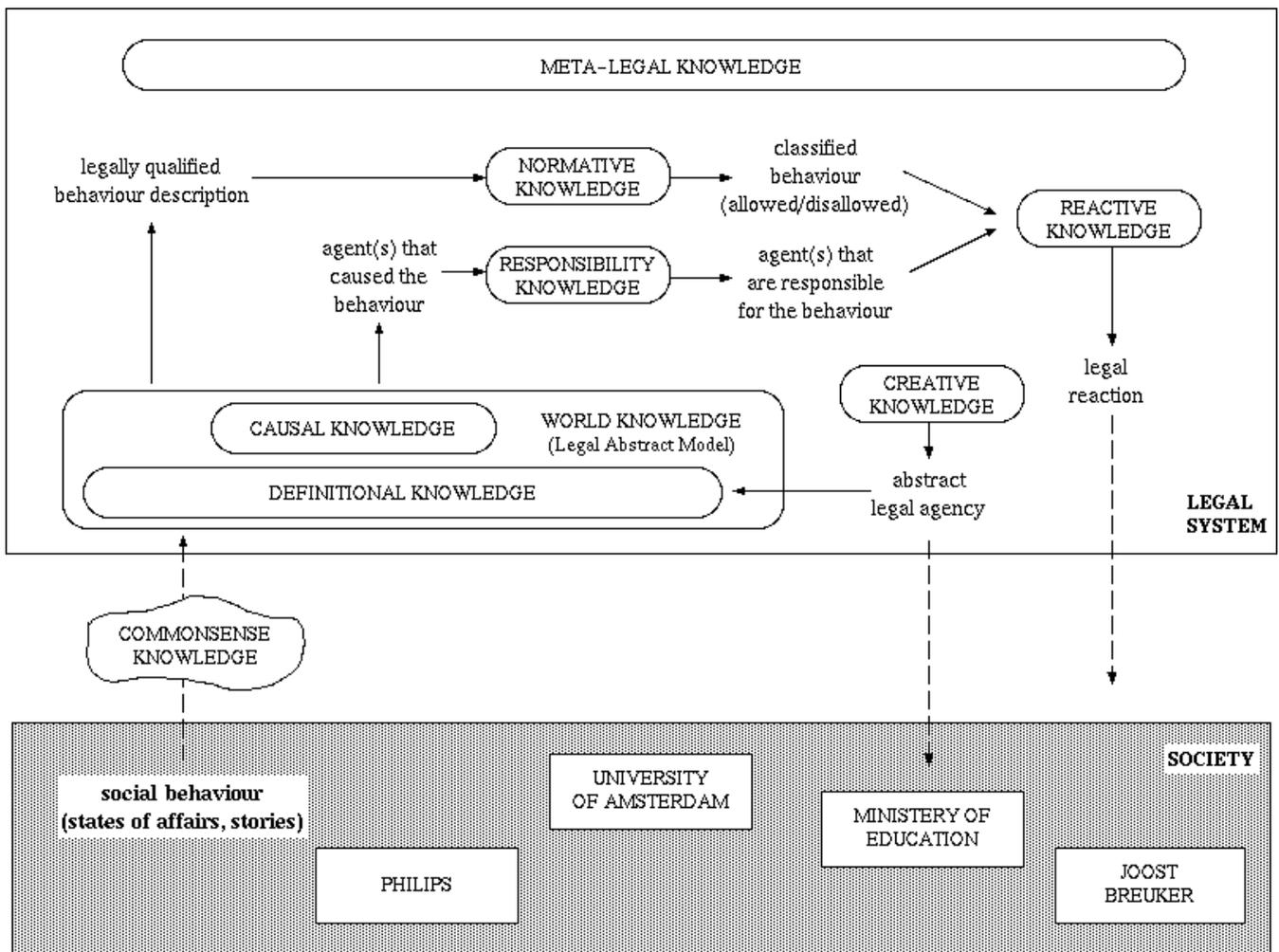


Figura 2.56. Papeles funcionales del conocimiento legal en el funcionamiento del sistema legal. (Valente y Breuker, 1996).

La función principal del sistema legal se puede descomponer en seis funciones primarias, donde cada una se corresponde con una categoría del conocimiento legal en *FOLaw* (figura 2.57): (a) conocimiento normativo, (b) conocimiento del mundo, (c) conocimiento de responsabilidad, (d) conocimiento reactivo, (e) conocimiento meta-legal y (f) conocimiento creativo, que detallamos a continuación.

- (a) Conocimiento normativo. Se caracteriza por ser el conocimiento que define un estándar de comportamiento. Prescribe el comportamiento de las personas en la sociedad. El estándar se define por el dictamen de normas individuales que expresan lo que debe ser. Distinguen tres tipos de normas: de orden (ordenan un comportamiento determinado), de poder (otorgan poder a un agente determinado para poner o aplicar normas) y derogatorias (derogan otra norma). Esta es una versión simplificada de la

tipología de normas de Kelsen, 1991). Tanto la versión de Valente y Breuker (1994a) como la de Kelsen (1991) se basan en las funciones realizadas por las normas en el sistema legal.

- (b) Conocimiento del mundo. El conocimiento del mundo es el conocimiento legal que describe el mundo objeto de ser regulado. Define el comportamiento posible de (personas e instituciones) en la sociedad y, por tanto, ofrece un marco para definir cómo debe o no debe ser tal comportamiento. Se considera como una interfaz entre el conocimiento del sentido común de las personas en la sociedad y el conocimiento normativo. Valente y Breuker (1994) distinguen dos tipos de conocimiento del mundo: (i) el “conocimiento terminológico” y el (ii) “conocimiento causal”. El conocimiento terminológico es la parte estática ya que consiste en definiciones de conceptos legales (agentes, objetos), relaciones legales (cualificaciones legales de las acciones), un caso (el caso bajo investigación), circunstancias (hechos fundados), casos generales (casos legales típicos) y condiciones. Por otra parte, el “conocimiento causal” es la parte dinámica, ya que describe el comportamiento de las personas en la sociedad en base al conocimiento terminológico.
- (c) Conocimiento de responsabilidad. Es el conocimiento legal que o amplía (asigna) o restringe la responsabilidad de un agente por su comportamiento. Tiene como función ofrecer los medios legales para rechazar la idea de que las personas son sólo responsables de sus actos (responsabilidad legal) ya que también se les considera responsables de sus causas (responsabilidad causal). Los actos son un concepto intermedio entre el conocimiento normativo y reactivo. El conocimiento de responsabilidad conecta la “responsabilidad causal” y la “responsabilidad legal” ya que cada agente es legalmente responsable de las causas de sus actos.
- (d) Conocimiento reactivo y conocimiento creativo. A estos dos conocimientos Valente y Breuker (1994a) los consideraban de menor carácter ya que no son parte del razonamiento legal nuclear. El conocimiento reactivo especifica la acción a desarrollar y cómo ha de ser desarrollada si un agente viola una norma primaria. Normalmente esta reacción suele ser un castigo, aunque también podría ser una recompensa.

El conocimiento creativo (f) es el conocimiento legal que permite la creación de entidades legales previamente inexistentes. Normalmente se confirma en términos imperativos, designando una entidad (por ejemplo, un comité gubernamental o un departamento de una compañía) que previamente no existía para que comience a existir a partir de ese momento concreto.

- (e) Conocimiento meta-legal. Se trata del conocimiento legal sobre el conocimiento legal o el conocimiento legal que se refiere a otro conocimiento legal. Es el conocimiento legal necesario para resolver conflictos entre normas. Hay cuatro categorías de conocimiento meta-legal: (e1) datos normativos, (e2) normas de orden, (e3) normativa por defecto y (e4) conocimiento de validez. Los datos normativos (e1) incluyen información sobre normas, tales como su ámbito de aplicación, su tipo, su lugar en la jerarquía de normas, su poder de origen, su promulgación y el objetivo de la norma. Las normas de orden (e2) son normas que determinan cómo resolver conflictos. La normativa por defecto (e3) determina cuál es el estatus legal en caso de que todas las normas primarias no se apliquen. Finalmente, el conocimiento de validez (e4) determina si el conocimiento legal es válido.

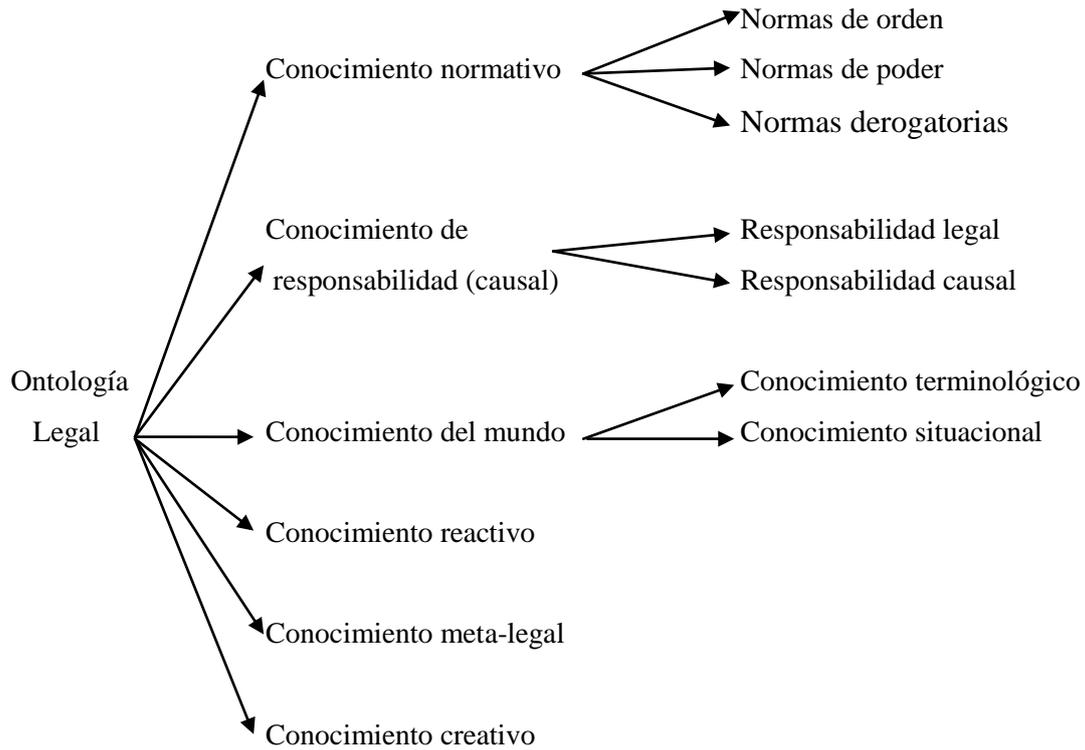


Figura 2.57. Categorías de conocimiento en *FOLaw* (adaptación de Valente y Breuker, 1994a: 5).

Según Valente y Breuker (1994a), *FOLaw* tiene una base teórica fundamentada en la literatura de la teoría legal de Kelsen, Hart y Bentham, por lo que describieron los conceptos primarios que estos autores habían propuesto para representar el conocimiento legal que pensaban podían ser traducidos directamente a una ontología. Formalizaron la ontología en *Ontolingua* con seis categorías primarias.

Para finalizar, señalar que Breuker y Hoekstra (2004) describieron *FOLaw* como un marco epistemológico más que como una ontología general del derecho, puesto que, según ellos, carece de los conceptos abstractos centrales del derecho, aunque reconocieron que permitió un mayor entendimiento del razonamiento legal.

### 2.7.2.2. *FBO*

La *Frame-Based Ontology of Law (FBO)* (van Kralingen, 1995; Visser, 1995) se basa en normas, actos y descripciones de conceptos. La principal distinción ontológica en *FBO* reside en el hecho de que consta de dos ontologías separadas: una ontología legal y una ontología específica de estatutos (Visser y Bench-Capon, 1996).

La ontología legal, en oposición con la ontología de estatutos, es la parte genérica y reutilizable de la ontología, ya que potencialmente se pueden reutilizar todos sus términos en cualquier subdominio legal. Divide el conocimiento legal en tres entidades: “normas”, “actos” y “descripciones de conceptos”. La ontología define una estructura marco para cada una de estas entidades, que incluye todos los atributos relevantes para la entidad. Para van Kralingen (1997), las normas son los elementos más importantes de los sistemas legales. Las “normas” son las reglas generales, estándares y principios de comportamiento que los sujetos están obligados a cumplir. En esta ontología una “norma” incluye los siguientes ocho elementos: (1) un identificador de la norma (utilizado como punto de referencia en la norma), (2) una norma tipo (norma de conducta o norma de competencia), (3) una promulgación (la fuente de la norma), (4) el alcance (el ámbito de aplicación de la norma), (5) las condiciones de aplicación (las circunstancias bajo las cuales es aplicable la norma), (6) el sujeto de la norma (la persona o personas a las que se dirige la norma), (7) la modalidad legal (obligación, no obligación o posibilidad), y (8) el identificador del acto (utilizado como referencia para un acto de descripción separado).

Los “actos” representan los aspectos dinámicos cuyo efecto cambia en el estado del mundo. En esta categoría hay dos distinciones. La primera es entre “eventos” y “procesos”. Los “eventos” representan un cambio instantáneo entre dos estados, mientras que los “procesos” tienen duración. La segunda distinción es entre actos institucionales y actos físicos. Un acto institucional es una cualificación legal de un acto físico. Todos los “actos” comparten los siguientes catorce elementos: (1) el identificador del acto (utilizado como punto de referencia del acto), (2) una promulgación (la fuente de la descripción del acto), (3) el alcance (el ámbito de aplicación de la descripción del acto), (4) el agente (un individuo, un conjunto de individuos, un agregado o un conglomerado), (5) el tipo de acto (tanto actos básicos como actos que se han especificado en otro lugar), (6) la modalidad del medio (los objetos materiales utilizados en el acto o subactos, por ejemplo, un arma), (7) la modalidad del modo (la manera en la que los objetos han sido utilizados o se han desarrollado los subactos, por ejemplo, de forma agresiva), (8) los aspectos temporales (una especificación de tiempo absoluta, por ejemplo, el 1 de mayo, los sábados, por la mañana, etc, pero no durante la película, después de que el rector venga, etc), (9) los aspectos espaciales (una especificación del

lugar donde tiene lugar el acto; por ejemplo, en Grecia, en un avión), (10) los aspectos circunstanciales (una descripción de las circunstancias bajo las cuales tiene lugar el acto, por ejemplo, durante unas elecciones), (11) la causa de la acción (una especificación de la razón o razones para desarrollar la acción, por ejemplo, venganza), (12) el propósito de la acción (el objetivo que visualiza el agente, por ejemplo, apropiarse de forma indebida de un objeto), (13) la intencionalidad de una acción (el estado mental del agente, por ejemplo, voluntario), (14) el estado final (los resultados y consecuencias de una acción, por ejemplo, la muerte de la víctima).

Las “descripciones de conceptos” versan sobre los significados de los conceptos del dominio. Pueden ser definiciones o provisiones a considerar y pueden utilizarse para determinar de forma definitiva el significado de una noción, o, en el primer caso, por ofrecer las condiciones necesarias y suficientes, o, en el segundo caso, por establecer una ficción legal. El factor es otro tipo de concepto, que puede establecer una condición suficiente o indicar algún tipo de contribución a la aplicabilidad del concepto. Finalmente hay metaconceptos, que son provisiones que determinan la aplicación de otras provisiones. Las “descripciones de conceptos” comprenden los siguientes siete elementos: (1) el concepto a ser descrito, (2) el tipo de concepto (definición, provisión a considerar, factor o meta), (3) la prioridad (el peso asignado a un factor), (4) la promulgación (la fuente de la descripción del concepto), (5) el alcance (el ámbito de aplicación de la descripción del concepto), (6) las condiciones bajo las cuales un concepto es aplicable y (7) una enumeración de las instancias del concepto.

Por su parte, la ontología específica de estatutos básicamente consiste en las relaciones de predicación utilizadas para instanciar los marcos para las normas, actos y descripciones de conceptos de modo apropiado en un dominio determinado. Van Kralingen (1995) y Visser (1995) señalan que la ontología legal genérica es reutilizable en cualquier dominio legal, mientras que la ontología específica de estatutos sólo puede reutilizarse para realizar tareas diferentes dentro del mismo dominio. Por tanto, la ontología específica de estatutos no se puede reutilizar en otro dominio, por lo que debe crearse una ontología específica de estatutos nueva para cada subdominio en cuestión. La ontología específica de estatutos establece el vocabulario con el cual se construye la base de conocimiento.

Inicialmente, *FBO* era una ontología “conceptual” informal (van Kralingen y otros, 1993; van Kralingen, 1995; 1997), pero más adelante Visser y Bench-Capon (1996) la formalizaron en *Ontolingua*. Esta ontología se ha utilizado como base del sistema *Framer* que trata dos aplicaciones de la ley de subsidio del desempleo holandés (Visser, 1995), una con respecto a la tarea de evaluación de las personas aptas para recibir el subsidio de desempleo y la otra con respecto a la planificación, para determinar si hay una serie de acciones a realizar con cierta consecuencia legal.

### **2.7.2.3. Knowledge Based Model of Law**

Mommer investigó los puntos de vista filosóficos y de representación del conocimiento de los conceptos legales y los representó con una semántica de situación. Como señaló Valente (1995) el propósito de su ontología era ofrecer una investigación teórica del dominio para facilitar su comprensión.

La ontología consiste en seis tipos básicos (tabla 2.1): entidades, capas de estatus ontológico, papeles epistémicos, relaciones, actos y hechos, y cada uno de ellos tiene subtipos legalmente relevantes y representantes legales. Los tipos básicos y los subtipos representan categorías de cosas y fenómenos típicos del dominio legal. Ofrecen un marco de representación de diferentes afirmaciones ontológicas y epistemológicas en la representación del conocimiento en el dominio legal. Al introducir capas de estatus ontológico diferentes y papeles epistémicos, el modelo puede acomodar diferentes visiones del derecho. Así, al especificar las capas de estatus ontológico y los papeles epistémicos aplicables a un cierto sistema legal, se puede construir un modelo legal que acomode las características específicas de ese sistema legal (Mommers 2001: 1).

En la figura 2.58 se muestra cómo se combinan las diferentes capas de estatus ontológico y los papeles epistémicos en el modelo del dominio legal basado en el conocimiento.

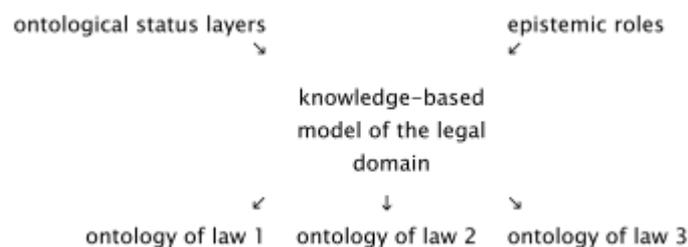


Figura 2.58. Distribución de las capas de estatus ontológico y papeles epistémicos en las ontologías legales (Mommers, 2001: 2).

Basic type	Legally relevant subtypes	Legal subtypes
Entities	sentences statements propositions beliefs artefacts rules concepts questions	legal principles legal norms legal decisions legal systematisations judicial interpretations judicial classifications legal rules legal concepts legal questions
Ontological status layers	existence constitution recognition	legal efficacy legal validity legal constitution legal recognition
Epistemic roles	reasons defeaters factual knowledge practical knowledge	legal reasons legal defeaters factual legal knowledge practical legal knowledge
Relations	causation counting as recognition	legal causation legal counting as legal recognition
Acts	applying rules making decisions making systematisations making interpretations making classifications	applying legal rules making legal decisions making legal systematisations making judicial interpretations making judicial classifications
Facts	brute facts recognised facts conventional facts institutional facts	recognised legal facts conventional legal facts institutional legal facts

Tabla 2.1. Perspectiva general de la ontología del dominio legal basada en el conocimiento (Mommers, 2001: 4-5).

Según Mommers (2001: 5) la ontología del dominio legal basada en el conocimiento se puede utilizar de varias maneras. Puede ser instrumento de investigación teórica, guiando la actividad de hacer distinciones relevantes en el dominio legal, y especificando las afirmaciones ontológicas y epistémicas en la filosofía legal. También se puede utilizar como base de marcado de textos legales con información relevante para desarrollar acciones de búsqueda. Esta ontología se ha aplicado al derecho penal holandés (Mommers, 2002).

#### **2.7.2.4. LRI-Core**

La ontología *LRI-Core* se desarrolla como una ontología legal general sobre nociones del sentido común (Breuker y Hoekstra, 2004a). El propósito de esta ontología no es proponer otra ontología de nivel superior, sino ofrecer una estructura conceptual amplia más profunda para las nociones legales “superiores” típicas o legalmente relevantes (Breuker y otros, 2002). Para estos autores muchos conceptos legales no eran típicos del derecho, sino especializaciones de conceptos del sentido común. El propósito de *LRI-Core* es sustentar la adquisición del lenguaje para las ontologías del dominio legal (Breuker, 2004).

Los autores consideraron reutilizar ontologías superiores existentes, sin embargo desecharon esta idea porque consideraron que estaban centradas en describir el mundo físico matemáticamente formal, en lugar del mundo social/comunicativo más típico del derecho (Breuker y otros, 2003).

Aunque no se construyó tomando como base otra ontología legal existente, incluye *FOLaw* y papeles legales, procedimientos, documentación, comunicación y fuentes legales. De hecho, para establecer una distinción clara entre personas y papeles, entre eventos físicos y mentales, *LRI-Core* incluye muchos términos de ontologías de nivel superior. En este nivel, *LRI-Core* consta de cinco tipos porciones o mundos: conceptos físicos (objetos y procesos), conceptos mentales, conceptos abstractos, papeles y acontecimientos.

*LRI-Core* se utilizó para ayudar al desarrollo de ontologías en el proyecto *e-Court*, que tenía como propósito la recuperación de la información y la generación y especificación de metadatos, específicamente utilizado por la ontología de derecho penal holandés del proyecto *e-Court* para apoyar la adquisición del conocimiento (Breuker y otros, 2003). Además, la idea era “anclar” los conceptos de esta ontología del derecho penal holandés a *LRI-Core* (figura 2.59) para facilitar el proceso de construcción de otras ontologías del derecho penal (derecho polaco e italiano), como solución a las posibles divergencias entre conceptos del derecho penal en distintos sistemas legales (Breuker y otros, 2002). La ontología y las subsiguientes ontologías específicas presentadas en el marco del proyecto *e-COURT* se desarrollaron en *Protégé-2000 (RDF(S))* (Breuker y otros, 2003).

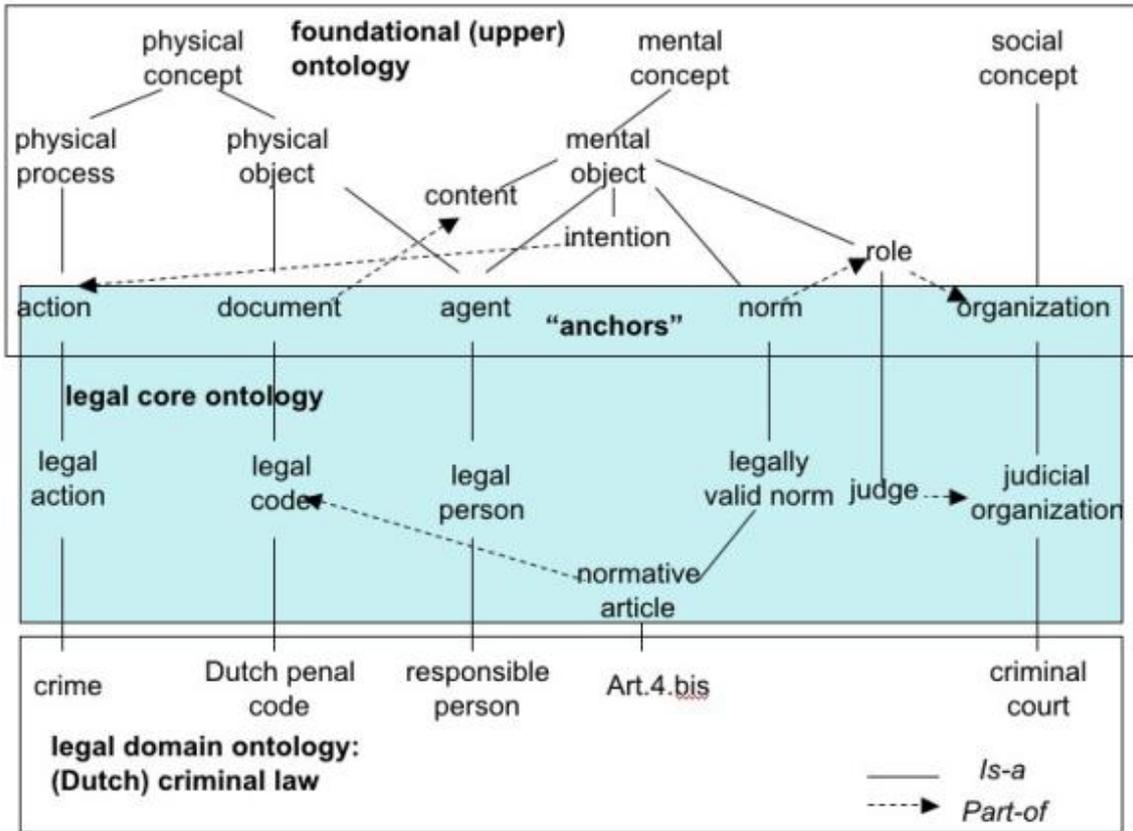


Figura 2.59. Las capas en *LRI-Core* en Breuker y Hoekstra (2004b).

La ontología estuvo disponible hasta 2007, cuando comenzó el desarrollo de la ontología *LKIF-Core*, que se basa en su predecesora. Aunque inicialmente se predijo que *LRI-Core* tendría alrededor de 200 conceptos (Breuker y otros, 2002), finalmente Breuker (2004) habla de 100 en la segunda versión de la ontología. Así, Breuker y otros (2007) concluyeron que el número de conceptos en *LRI-Core* era bastante pequeño y que se trataba más bien de una ontología superior de conceptos abstractos del sentido común que de una ontología general del dominio del derecho.

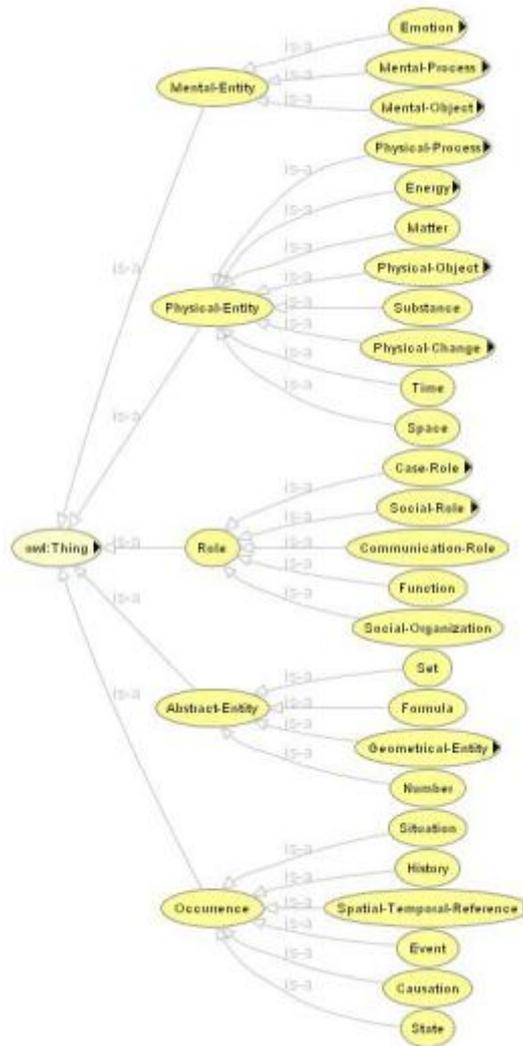


Figura 2.60. Las primeras dos capas de *LRI-Core* (Breuker y Hoekstra, 2004a).

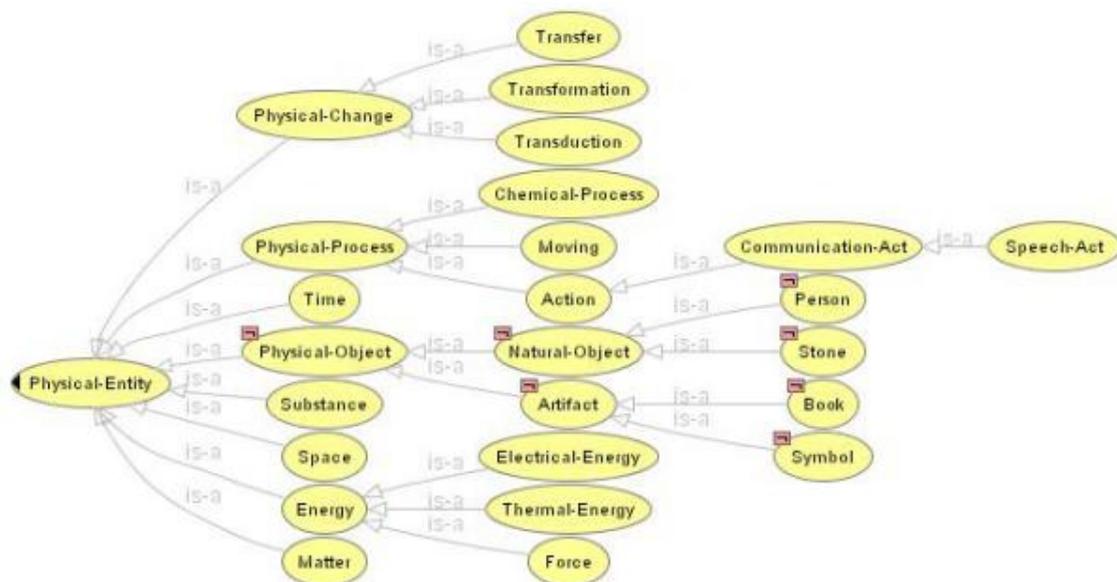


Figura 2.61. Ejemplo de clases físicas en *LRI-Core* (Breuker y Hoekstra, 2004a).

### 2.7.2.5. *LKIF-Core*

*LKIF* (*Legal Knowledge Interchange Format*) se desarrolló dentro del proyecto Estrella para permitir la traducción entre bases de conocimiento legal escritas en lenguajes de representación y formatos diferentes. Funciona como un formalismo de representación del conocimiento que es parte de una arquitectura mayor para desarrollar sistemas de conocimiento legal en la Web Semántica (Hoekstra y otros, 2007).

En el desarrollo de *LKIF*, se construyó *LKIF-Core*<sup>57</sup>, una ontología de conceptos legales básicos desarrollada y formalizada en *OWL-DL*. Está inspirada en la orientación al sentido común de la ontología *LRI-Core* que está suspendida actualmente. Consiste en 14 módulos que describen conceptos que van desde conceptos generales como tiempo, lugar, cambio y proceso, a los conceptos angulares del dominio legal tales como acciones, transacciones, creencias, intenciones, expresiones y normas.

Según Hoekstra y otros (2007), *LKIF-Core* pretende jugar un importante papel en la traducción de bases de conocimiento existentes a otros formatos de representación compatibles en el proceso de adquisición y modelado del conocimiento en el dominio legal. Pero principalmente contiene conceptos básicos del derecho y es parte de una arquitectura genérica que permite el intercambio de conocimiento. Así, *LKIF-Core* está destinada a ser compatible con mecanismos de inferencia legal, adquisición del conocimiento e intercambio de información.

La ontología consta de tres capas: un nivel superior (que toma prestadas la mayoría de clases de *LRI-Core*), un nivel intencional (que incluye conceptos y relaciones que describen el comportamiento de agentes racionales fácilmente influenciados por la ley) (Hoekstra y otros, 2007) y un nivel legal (agentes legales, derechos y poderes de las acciones – modificado de Rubino y otros (2007) – y papeles legales y definiciones de conceptos). Hoekstra y otros (2007) señalan que se acogen a una visión bastante restrictiva de lo que ha de contener una ontología: conocimiento terminológico, esto es,

<sup>57</sup> <http://www.leibnizcenter.org/general/lkif-core-ontology>

definiciones intensionales de conceptos, representados como clases con las que interpretamos el mundo.

Hoekstra y otros (2007) señalan que la construcción de *LKIF-Core* se basó en la combinación de varias metodologías de ingeniería ontológica, una adaptación de Uschold y Grüninger (1996) entre otras, en los siguientes pasos: (1) identificación del propósito y dimensión de la ontología, (2) captura de la ontología, (3) codificación de la ontología, (4) integración de la ontología en otras ontologías existentes y (5) evaluación. Hoekstra y otros (2007: 48-49; 2009: 33) explican el proceso de captura de la ontología de la siguiente manera:

Dentro del proyecto Estrella participaron tres tipos de expertos (ciudadanos, profesionales del sector legal y expertos en la materia). A cada uno se le pidió que elaborara una lista de los 20 conceptos legales más importantes. Éstos, combinados con los términos encontrados en la literatura (jurisprudencia y libros de texto legales) obtuvieron una lista de unos 250 términos. Como esa cantidad no es manejable como conjunto básico para modelar el resto de los términos, pidieron a los participantes en el proyecto que evaluaran cada término de la lista en cinco escalas: nivel de abstracción, relevancia para el dominio legal, el grado en el que un término es legal con respecto al sentido común, el grado en el que un término es general del dominio legal (en oposición a un término que es específico de algún subdominio del derecho), y el grado en el que el experto pensaba que ese término debía ser incluido en la ontología. La posición relativa de un término con respecto a estas escalas indica si es apropiada su inclusión en los grupos de términos básicos. Por ejemplo, cuanto más alta sea su puntuación para la importancia, relevancia legal y término legal común, más apropiada será su inclusión en la ontología. Por otro lado, una puntuación alta o baja con respecto al nivel de abstracción o sentido común frente a legal indica que el término no es básico. Para esas escalas buscaban términos más que una puntuación de media.

Los resultados se utilizaron para seleccionar un conjunto inicial de 50 términos más aquellos reutilizados de otras ontologías (*LRI-Core* y *LLD*)<sup>58</sup> y formaron la base para la identificación de los grupos de términos y el desarrollo de la ontología *LKIF-Core*. A modo de ilustración, se incluye la tabla 2.2 con los diez términos de mayor puntuación para las tres escalas.

Importance	Abstractness	Legal Relevance
Law	Deontic operator	Civil law
Right	Law	Law
Jurisdiction	Norm	Legal consequence
Permission	Obligative Right	Legislation
Prohibition	Permissive Right	Obligation
Rule	Power	Right
Sanction	Right	Authority
Violation	Rule	Deontic operator
Power	Time	Duty
Duty	Anancastic Rule	Jurisdiction
Legal Position	Existential Initiation	Legal Fact
Norm	Existential Termination	Legal Person
Obligation	Potestative Right	Legal Position
Permissive Right	Productive Char.	Legal Procedure
Argument	Absolute Obl. Right	Liability

Tabla 2.2. Los diez primeros términos de acuerdo a su importancia, nivel de abstracción y relevancia legal (Hoekstra y otros 2009: 33).

La versión 1.0.2. de *LKIF-Core* tenía 206 clases y 113 propiedades de objeto. La versión 1.0.3 (a fecha de 10 de noviembre de 2008) tenía 155 clases y 97 propiedades de objeto Casellas (2008: 153). Durante la realización de esta tesis no hemos podido comprobar en qué estado de desarrollo se encuentra la ontología ya que no estaba disponible.

### 2.7.2.6. *JurWordnet*

*ItalWordnet (IWN)* es la sección italiana de *EuroWordnet*, desarrollada en el Instituto de Lingüística Computacional del CNR de Pisa. *Jur-(Ital)Wordnet (JurIWN)* es una ontología extensión del dominio legal de la versión italiana de *EuroWordnet*, vinculada a los registros del índice interlingual (*ILI*) (Gangemi y otros, 2003a).

<sup>58</sup> Los desarrolladores de la ontología esperaban poder reutilizar conceptos de otras ontologías nucleares y de alto nivel aunque resultó que la cantidad de términos susceptibles de ser reutilizados fue bastante limitada. Entre las ontologías que se consideraron estaban *SUMO*, *DOLCE*, *CyC*, *LRI-Core*, *LLD* y *CLO*.

*Jur-IWN* es un lexicón semántico que se puede considerar un tipo de ontología superficial u ontología léxica. Como en la *WordNet* original, en *JurWordNet* los *synsets*<sup>59</sup> se relacionan unos con otros en varias relaciones semánticas tales como hiperonimia, hiponimia, meronimia, papel temático o instancia-de. En oposición a las ontologías formales, a los léxicos semánticos, también llamados ontologías superficiales, son genéricos y se basan en un modelo débil de abstracción, dado que los elementos (clases, propiedades e individuos) de la ontología dependen primariamente de la aceptación de entradas léxicas existentes (cf. Tiscornia, 2007).

La ontología *Jur-IWN* puede ofrecer una base léxica en la construcción de ontologías específicas de dominio (Gangemi y otros, 2003a), compatibilidad con la gestión del contenido y ofrecer una fuente de metadatos para descripción del contenido utilizable para el etiquetado semántico de textos legislativos. Además, la base de datos puede ser una herramienta de apoyo para los sistemas de recuperación de la información para facilitar el acceso a datos heterogéneos y multilingües, y una fuente conceptual de extracción de la información, de etiquetado automático, de compartir el conocimiento, de comparación de normas, etc (Gangemi y otros, 2003b). Específicamente, el objetivo de *Jur-IWN* es ofrecer al sistema *NormeInRete*<sup>60</sup> una base de conocimiento para el etiquetado semántico en el nivel de artículos o incluso disposiciones; esto es, mediante el reconocimiento de entidades normativas (las disposiciones) del texto, que no es necesariamente lo mismo que las entidades estructurales, y asumiendo un doble punto de vista en el texto, como documento y como conjunto de disposiciones (Gangemi y otros, 2003b).

Como fuente de metadatos para la descripción de contenido, utilizable para el etiquetado semántico, necesitamos un estándar de metadatos basado, no en las relaciones léxicas, sino en la definición del contenido, en las unidades léxicas. Tal estándar debe estar basado en la naturaleza ontológica de las entidades del mundo legal;

---

<sup>59</sup> Conjunto de uno o más lemas con la misma forma (sustantivo, verbo, adjetivo o adverbio) que se pueden intercambiar en un contexto determinado. Por ejemplo, acción, juicio, actas, juicio forman un conjunto de sustantivos sinónimos porque se pueden utilizar para referirse al mismo concepto (Tiscornia, 2007).

<sup>60</sup> <http://www.ittig.cnr.it/Ricerca/UnitaEng.php?Id=40>

conceptos como “licencia”, “autorización”, “delegación”, que adquieren un significado especial en el dominio legal, y que aproximadamente se unen a las particiones clásicas de la teoría legal, se organizan en una ontología legal nuclear (Gangemi y otros, 2003b).

El desarrollo de la ontología nuclear tiene en cuenta metodologías propias de las ontologías de alto nivel, y propuestas en el dominio de las ontologías legales. La ontología nuclear organiza los conceptos jurídicos en clases en la base de (meta) propiedades formales definidas en las ontologías fundacionales. Además, las entidades básicas que pueblan el dominio legal se pueden considerar universal y claramente identificables y, como tal, se marcan por una serie mínima de propiedades y relaciones generalizadas. El nivel inferior de la ontología nuclear es especializado de acuerdo al dominio de aplicación y se amplían en un nivel léxico a través de las taxonomías de *Jur-IWN*. En la figura 2.62 se muestra la estructura de *Jur-IWN*.

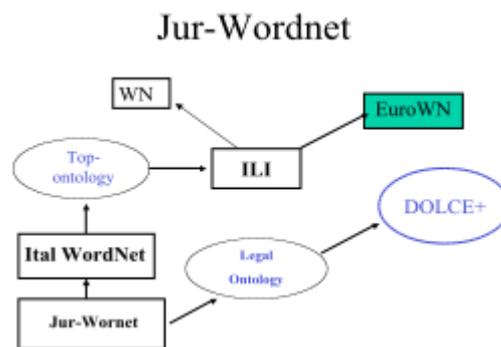


Figura 2.62. La estructura de *JUR-IWN* (Gangemi y otros, 2003b).

### 2.7.2.7. OPLK

La *OPLK* (*Ontology of the Professional Legal Knowledge*) ofrece un nuevo punto de vista en las ontologías legales. Se desarrolló de acuerdo a un cuestionario (tabla 2.3) elaborado a 130 jueces noveles, cuyas respuestas fueron comparadas con las de 141 jueces veteranos. Las preguntas versaron sobre las siguientes áreas:

1. Evaluación de la formación en la universidad, en la academia del poder judicial y en cursos de formación continua.
2. Principales problemas encontrados por los jueces en su actividad profesional desde la inserción laboral a casos judiciales.

3. Los servicios del CGPJ (Consejo General del Poder Judicial) en relación a las nuevas tecnologías y documentación.
4. Las redes profesional, institucional y social a las que pertenece el juez.
5. El significado de justicia para los jueces.

Las preguntas trataban sobre las principales dudas profesionales que los nuevos jueces afrontaban en sus primeros años de actividad profesional. La mayoría de las preguntas eran de respuesta abierta con el objetivo de obtener respuestas espontáneas y literales. La tabla 2.3 resume algunas de esas preguntas.

<p><b>Questions:</b></p> <ul style="list-style-type: none"><li>• <i>What kind of judicial cases (civil or criminal cases) were more frequent in your first destination?</i></li><li>• <i>What were the two most important doubts that you had in your first three months as a judge?</i></li><li>• <i>What were the most complicated civil cases that you had to solve in your first year as a judge? Why?</i></li><li>• <i>What were the most complicated criminal cases that you had to solve in your first year as a judge? Why?</i></li><li>• <i>Could you define the civil case that has given you more work?</i></li><li>• <i>Could you define the criminal case that has given you more work?</i></li><li>• <i>Which are the most frequent problems that you have found in the hearings?</i></li></ul>
---

Tabla 2.3. Extracto del cuestionario de *OPLK* (actividad profesional) en Benjamins y otros (2003).

Se analizaron las preguntas de respuesta abierta de la tabla 2.3 para extraer de ellas los dominios en los que los nuevos jueces tenían dudas. Las preguntas frecuentes tenían que estar orientadas a resolver este tipo de dudas. Para encontrar las áreas problemáticas de los nuevos jueces se utilizaron métodos de estadística textual. Tras identificar la palabra “guardia” como tema importante a considerar en el sistema de preguntas frecuentes, se enumeraron todas las respuestas que mencionaban el sustantivo “guardia” en el subcorpus de sustantivos, que aparecen en la tabla 2.4.

•	1	ON-DUTY PROBLEMS
•	2	ON-DUTY SERVICE
•	3	BEHAVIORS DURING ON-DUTY PERIODS
•	4	TREATMENT OF URGENT FAMILY MATTERS DURING THE DUTY
•	5	ON-DUTY MATTERS, PERSONS INVOLVED IN SPECIFIC ACTIONS
•	6	CRIMINAL: ON-DUTY PERIODS
•	7	ON-DUTY PROBLEMS CONCERNING MINORS PROTECTION
•	8	SOLVING SPECIFIC ON-DUTY MATTERS
•	9	ON-DUTY MATTERS
•	10	ON-DUTY/ CALLS FROM THE POLICE CONCERNING CERTAIN ASPECTS WHICH DO NOT FIGURE IN THE BOOKS/ PRACTICAL ASPECTS DURING ON-DUTY PERIOD)
•	11	DOUBTS ARISING DURING ON-DUTY PERIODS
•	12	WHAT SHOULD BE UNDERSTOOD BY ON-DUTY ACTIONS
•	13	OF THE ON-DUTY COURT
•	14	WHEN THE LEGAL RESPONSIBILITY FILE COULD BE OPENED, ESPECIALLY IF IT COULD OCCUR DURING THE ON-DUTY PERIOD
•	15	AT THE BEGINNING, DURING THE FIRST THREE MONTHS, MY DOUBTS CONCERNED IMPORTANT DECISIONS TO BE TAKEN DURING THE ON-DUTY PERIOD
•	16	CANNOT CONCRETE, THE MOST IMPORTANT DOUBTS ARISEN DURING THE ON-DUTY PERIODS
•	17	IF THE DISTRICT JUDGES HAD TO DO ON DUTY PERIODS

Tabla 2.4. Lista de respuestas con la palabra “guardia” (*on duty*) (Benjamins y otros, 2003).

A partir de estas respuestas, los expertos elaboraron un conjunto de preguntas en relación a los problemas surgidos durante las guardias. Se siguió el mismo proceso para todas las preguntas de este campo para destacar las principales cuestiones de la ley de enjuiciamiento civil, tráfico de drogas, etc.

Según Benjamins y otros (2003), la construcción de una ontología legal tiene más que ver con los modelos legales que con las teorías generales sobre el derecho. Los jueces son expertos, dado que conocen perfectamente las leyes y la jurisprudencia. Así, la función de la *OPLK* era modelar el conocimiento legal profesional, esto es, el tipo de conocimiento compartido por los miembros de la profesión y transmitido en los procesos de formación y medios organizativos. Para estos autores, el conocimiento profesional legal: (i) es conocimiento corporativo (se excluye especialmente a otros profesionales del derecho); (ii) no está distribuido equitativamente entre los miembros del colectivo; (iii) está basado en la experiencia; (iv) es susceptible a la experiencia (dependiente de los lugares, casos e historia personal); (v) es transmitido institucionalmente en facultades de derecho, escuelas de práctica del derecho, escuelas de derecho, escuela de la judicatura, tribunales, bufetes y agencias del estado.

Como explican Benjamins y otros (2003), en *OPLK* la interpretación de los textos legales (estatutos, normas, decretos, etc.) que las ontologías legales intentan reflejar quedan “ancladas” en el conocimiento profesional que ellos destacan, ya que las ontologías legales de dominio se solapan con las ontologías legales nucleares. Así, el

conocimiento legal profesional construye un dominio intermedio en el que los contextos legales y el conocimiento legal compartido se vinculan a los estatutos particulares y normas especiales.

La *OPLK* se basa en la premisa de que cualquier juez novel inexperto consulta sus dudas con los jueces más expertos, quienes comparten sus conocimientos con los anteriores. Así, se infirieron conceptos de la gran cantidad de materiales recogidos (casos difíciles, casos extraños, interpretaciones legales, analogías legales, actitudes profesionales y estándares comunes). La figura 2.63 representa el conocimiento profesional de *OPLK* de forma gráfica.



Figura 2.63. Parte del conocimiento profesional relevante para los jueces noveles (Benjamins y otros, 2003).

En la figura 2.64 aparece la *OPLK* formalizada en *Protégé*, que incluye los siguientes términos: procesos, objeto y actor, con sus correspondientes instancias, que se vinculan a los atributos “generalización”, “equivalencia”, “actor”, “seguido de” y “subproceso de”, que representan relaciones (Benjamins y otros, 2003).

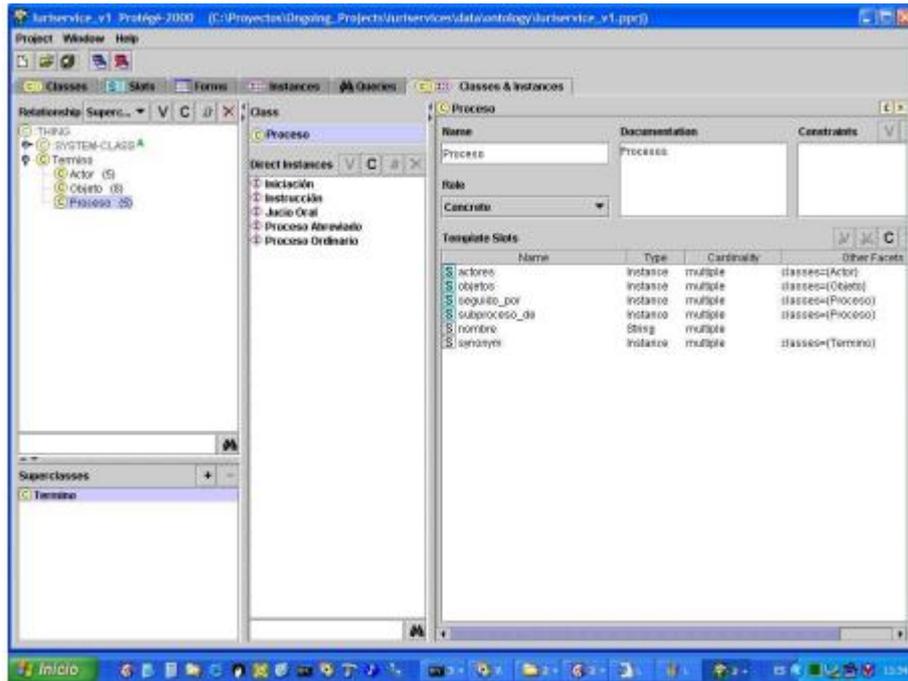


Figura 2.64. La ontología legal en *Protégé* (Benjamins y otros, 2003).

Benjamins y otros (2003) señalan que la ontología *OPLK* se basa en la descripción de la estructura cognitiva que todos los jueces comparten para comprender y gestionar los problemas legales y prácticos a los que normalmente tienen que enfrentarse cuando están de guardia. Destacan que esta primera descripción era susceptible de mejorar para refinar la ontología.

#### 2.7.2.8. *OPJK*

La *OPJK* (*Ontology of Professional Judicial Knowledge*) modela y representa formalmente el conocimiento judicial profesional para hacer accesibles las preguntas del sistema de preguntas frecuentes de la aplicación *Iuriservice* en su segunda versión. El principal objetivo de la ontología era facilitar la recuperación de los pares de preguntas y respuestas relativos a la práctica judicial almacenados de forma semánticamente similar a la pregunta de entrada en lenguaje natural. Como las preguntas versaban sobre varios dominios legales, desde el derecho de la familia hasta el derecho comercial y el derecho penal, la ontología legal tenía que incluir conceptos de varios subdominios legales. Para ello la principal fuente de conocimiento fue el corpus de preguntas y respuestas obtenidas en el cuestionario (el mismo que en la *OPLK*) y almacenadas en la base de datos del sistema. Así, se utilizó el corpus para la extracción automática de términos y la adquisición del conocimiento. Otros criterios para el diseño de la

ontología fueron la estrecha colaboración entre los expertos del dominio y los ingenieros ontológicos en un entorno apropiadamente distribuido para la construcción de la ontología, la naturaleza de la herramienta de recuperación de la información (*Juriservice*), el juez como usuario de la herramienta, y el español como la lengua de uso tanto en el lenguaje natural de entrada como de salida (Casanovas y otros, 2009).

Durante 2008 se implementó el sistema en la escuela judicial. En el momento de publicación del artículo de Casanovas y otros (2009) la ontología estaba en proceso de refinamiento para mejorar los procesos de eficiencia y efectividad, a la vez que se hacían los ajustes necesarios en la aplicación para adaptarse mejor a las necesidades de los usuarios.

#### **2.7.2.9. *Lame's French Codes***

La ontología de derecho francés de Lame (2000; 2001; 2002; 2005) tenía como propósito principal ayudar a la búsqueda y recuperación de información legal. La construcción de la ontología se basó en un análisis terminológico de un corpus de 57 códigos de la ley francesa, una metodología de abajo a arriba inspirada en el modelo de Aussenac-Gilles y otros (2000a); Moens (2001). La metodología siguió los siguientes pasos: (1) extracción de los términos del dominio (inicialmente 118.000 términos), (2) identificación de las relaciones léxicas entre términos, (3) modelado de la jerarquía de conocimiento del dominio, (4) modelado de el conocimiento estructural del dominio, y (5) modelado del conocimiento de las funciones de las normas. Cabe señalar que el mismo Lame (2002) asumió tanto el papel de ingeniero del conocimiento como de experto del dominio.

Para la extracción de términos se utilizaron varios extractores de términos como *Syntex*. La lista inicial de 118.000 términos legales se redujo a una lista de 16.681 términos legales fundamentales. La ontología se implementó en un sistema para apoyar el refinamiento de cuestiones para su evaluación (Lame, 2002). Lame y Desprès (2005) destacan que se desarrolló una ontología más amplia de un nuevo código (Código de Investigación) con la misma metodología anteriormente mencionada, que se alineó a la ontología anterior de los códigos franceses utilizando técnicas ordinarias y la participación de expertos legales. Tanto la ontología original como la de 2005 estaban

formalizadas en archivos *XML*. El proceso de actualización de la ontología general consistió en incluir los elementos descritos con los marcadores *XML* apropiados (Lame y Desprès, 2005).

### 2.7.2.10. Italian Crime Ontology

La ontología penal italiana de Asaro y otros (2003) ofreció una conceptualización del dominio de conceptos del derecho penal italiano para los siguientes fines: (i) conseguir una estructura conceptual homogénea en varios proyectos (proyecto Minerva, proyecto Itaca y P.M. Daedalus); (ii) añadir conocimiento del dominio a las herramientas de apoyo; (iii) gestionar los documentos a través de metadatos; (iv) identificar y sugerir hipótesis penales al juez, por ejemplo, al escanear documentos a través de un analizador sintáctico profundo; y (v) marcar de forma semántica las leyes penales con un lenguaje *XML*.

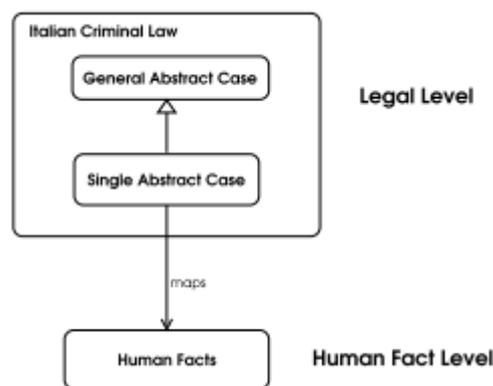


Figura 2.65. Capas de la ontología penal italiana (Asaro y otros, 2003)

La conceptualización incluyó también los conceptos relacionados con el delito no cometido y con la tentativa de delito. En una primera etapa se utilizaron diagramas de clase en *UML (Unified Modelling Language)* para la formalización. En posteriores formalizaciones se utilizaría el lenguaje de representación *DOLCE*. La ontología se construyó por el análisis de los artículos 42-49 del código penal por los expertos (un magistrado en el juzgado de Lucca (C. Asaro) tomó parte en el desarrollo de la ontología) (Asaro y otros, 2003).



representación normativa más exigente, resultó tener mucho éxito. La ontología de dominio se construyó sin la reutilización de otras representaciones de conocimiento externas (Boer y otros, 2001) e incorporó una pequeña y abstracta ontología superior que distinguía cosas como artefactos, sustancias, agentes y funciones (Winkels y otros, 2002).

En cuanto al lenguaje de formalización, la ontología de dominio de *CLIME* utiliza un lenguaje restringido e inexpressivo con un número fijo de relaciones binarias entre los términos. El modelo de datos de la ontología se puede considerar como un gráfico etiquetado. Este modelo se puede traducir superficialmente al formato *RDF*, que depende de un modelo de datos muy similar. El formato de representación de *CLIME* no tiene equivalente para la noción de cosificación en *RDF*, y no permite el tratamiento de etiquetas de relaciones como nodos de términos (Boer y otros, 2001). Boer y otros (2001) y Winkels y otros (2002) señalan que la ontología consistía en 3377 conceptos y 11897 relaciones (incluyendo las relaciones *IS-A*). La tabla 2.5 muestra la lista de los 10 conceptos más frecuentes y sus relaciones en la ontología *CLIME*.

Concept	Nr of Relations
ACTION	60
SYSTEM	48
STATE	40
ENTITY	36
EXAMINATION	34
METRIC	31
SHIP	30
DOCUMENTATION	29
ADDITIONAL-CLASS-NOTATION	28
SURVEY	26

Tabla 2.5. Los 10 conceptos más frecuentes en *CLIME* y sus relaciones (Winkels y otros, 2000).

Para el uso de *CLIME* en el proyecto *KDE* se tradujo la ontología a una base de conocimiento en *Protégé-2000*. En este proyecto la ontología se utilizó para mejorar la búsqueda en documentos técnicos arbitrarios y documentos legales utilizados en tareas de negocios asociadas con la clasificación e inspección de barcos (Winkels y otros, 2002).

### 2.7.2.12. IPROnto

*IPROnto* es una ontología del dominio de la gestión de los derechos de autor. Se diseñó para facilitar el desarrollo de aplicaciones de comercio electrónico que necesitan conocer los derechos asociados con contenido multimedia específico (Delgado y otros, 2003).

Para el desarrollo de la ontología comenzaron con especificaciones seleccionadas. Partieron de los resultados del marco <indec> y de los proyectos *Imprimatur* (Barlas, 1995) por su cobertura en describir el dominio de los derechos de propiedad intelectual (*IPR*), por lo que las definiciones de los términos y su estructura se adoptaron y se formalizaron con herramientas de ontologías web (Delgado y otros, 2003). Este trabajo se complementó con el trabajo del grupo de investigación *DMAG* (*Distributed Multimedia Applications Group*<sup>61</sup>).

Para dotar a la ontología de una base ontológica robusta utilizaron la ontología de alto nivel *SUMO*. Los aspectos legales de la ontología los tomaron de *WIPO* (*World Intellectual Property Organisation*<sup>62</sup>) ya que definía el marco legal común de los derechos de propiedad intelectual. En la figura 2.67 se muestran las fuentes de conocimiento de *IPROnto*. Inicialmente se formalizó en *DAML-OIL* y en una edición posterior se formalizó en *OWL-DL* (Delgado y otros, 2003).

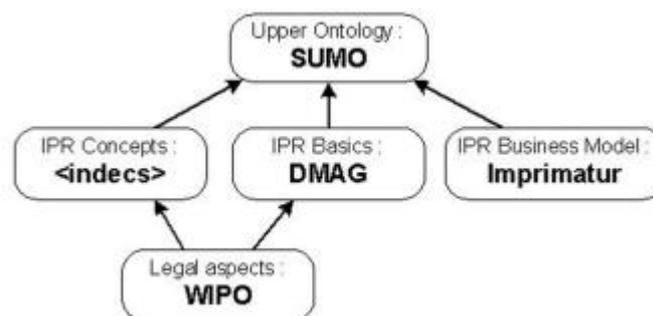


Figura 2.67. Las fuentes de conocimiento de *IPROnto* (Delgado y otros, 2003).

---

<sup>61</sup> <http://dmag.ac.upc.edu/>

<sup>62</sup> <http://www.wipo.int/portal/index.html.en>

### 2.7.2.13. *Laymen*

La ontología *Laymen* es una ontología con términos del sentido común para casos legales enmarcada dentro del proyecto *BEST*<sup>63</sup>. Representa el conocimiento de los no expertos en casos de responsabilidad civil. Junto con una herramienta de anotación semiautomática, permite a los usuarios relacionar la descripción de su caso con los términos en la ontología para no expertos. Se trata, en realidad, de dos ontologías diferentes. La primera difiere significativamente de la segunda, no sólo en que la terminología usada por no expertos y expertos es diferente, sino también en la representación requerida. Mientras que una ontología para crear descripciones estructuradas de casos necesita ofrecer la base para describir configuraciones de situaciones distintas, la ontología de la legislación se centrará en los conceptos legales o situaciones. De este modo estas tareas requieren de representaciones conceptuales diferentes. Para rellenar el hueco existente entre ambas ontologías utilizaron razonamientos lógicos y así determinar automáticamente los conceptos legales relevantes para determinar la responsabilidad civil de las partes en un caso basado en esta descripción. Como resultado se obtuvo una descripción en términos de responsabilidad civil anotados con conceptos legales (Klein y otros, 2006).

### 2.7.2.14. *Ontology of Dutch Criminal Law (OCLN.LN)*

La ontología se desarrolló dentro del proyecto *e-Court*<sup>64</sup>, diseñado para la gestión semiautomática de la información de las transcripciones de las vistas de juicios penales (Breuker y otros, 2002). La ontología estaba destinada a ser soporte de recuperación de la información de los documentos de las vistas y también se pretendía que sirviera de punto de referencia para el desarrollo de ontologías similares del derecho penal italiano y polaco. Sin embargo, como señala Breuker (2004), esta parte del proyecto no se pudo desarrollar debido a los problemas de traducción y a un presupuesto limitado.

El núcleo de *OCLN.LN* está compuesto de acciones (delitos y penas). El papel de las ontologías de indización de los documentos de las vistas del *e-Court* tenía tres objetivos: (1) ofrecer un vocabulario estructurado para las descripciones de metadatos y mantener un uso consistente y distinciones semánticas, (2) el navegador de la ontología

---

<sup>63</sup> <http://www.best-project.nl/>

<sup>64</sup> <http://www.intrasoft-intl.com/e-court> no está disponible.

era compatible con el etiquetado manual de la vistas y (3) los conceptos de la ontología se utilizaron para indizar documentos (Breuker y otros, 2005).

La ontología se formalizó inicialmente en *RDF/OWL (DAML+OIL)* en *Protégé* para ser utilizada en el sistema *e-Court*, que estaría disponible en la Web (Breuker y otros, 2002). Esta decisión con respecto a la formalización estaba motivada por el hecho de que en ese momento *LRI-Core* se iba a reutilizar como ontología superior y estaba formalizada en *DAML+OIL*, por lo que los conceptos de la ontología *OCL.NL* se “anclaron” en los concepto de *LRI-Core*. La figura 2.68 describe el “anclado” de los conceptos de *OCL.NL* a los conceptos agentes de *LRI-Core* (los conceptos en negrita).

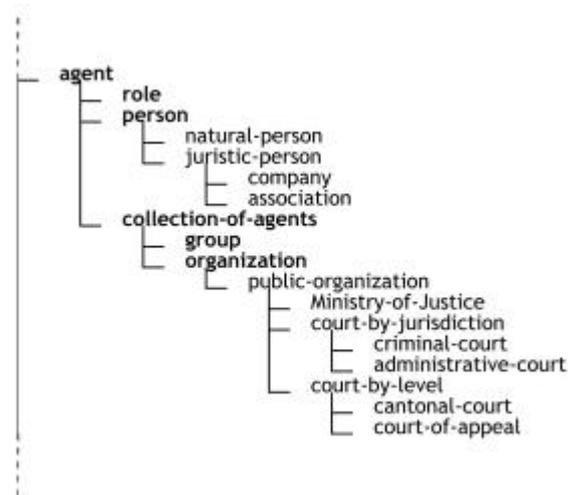


Figura 2.68. El “anclado” de conceptos en *OCL.NL/LRI-Core* (Breuker y otros, 2002).

### **2.7.2.15. DALOS**

La ontología *DALOS* significa *DrAfting Legislation with Ontology-based Support* (Francesconi y otros, 2007). El proyecto *DALOS* surgió dentro del marco de "participación electrónica" (*eParticipation*), iniciativa de la Comisión de la UE destinada a promover el desarrollo y uso de las tecnologías de la información y la comunicación en los procesos legislativos de toma de decisiones, con el fin de fomentar la calidad de la producción legislativa, para mejorar la accesibilidad y la adaptación de la legislación a nivel europeo, así como para promover la concienciación y la participación democrática de los ciudadanos en el proceso legislativo.

En particular, *DALOS* pretende garantizar que los redactores de textos jurídicos y los que toman las decisiones asuman el control sobre el lenguaje jurídico a nivel nacional y europeo, proporcionando a los legisladores herramientas de gestión y los conocimientos lingüísticos que se utilizan en los procesos legislativos, en particular en la fase de la redacción de leyes. En la base de la caracterización ontológica del lenguaje legal, *DALOS* pretende proporcionar a los legisladores las herramientas de gestión lingüística y de conocimiento para apoyar la redacción de leyes en un entorno multilingüe.

*DALOS* aborda el proceso de redacción legislativa, concretamente un proceso que crea las normas en los dominios específicos a ser regulados. Es necesario, por tanto, un apoyo lingüístico y de conocimiento que dé una descripción de los conceptos, así como sus manifestaciones léxicas en diferentes lenguas en los ámbitos específicos antes de ser regulados.

En cuanto a los estándares, el estándar de conversión de *RDF / OWL* de *WordNet* aprobado por los estándares del *W3C* se utiliza para la lingüística de recursos, garantizando así tanto la interoperabilidad como la escalabilidad de la solución.

En cuanto a lo que se refiere al sistema de organización del conocimiento, *KOS* (*Knowledge Organization System*), el recurso de *DALOS* se organizaba en dos niveles de abstracción (véase la figura 2.69):

- La capa ontológica que contiene el modelado conceptual en un nivel independiente del nivel de lenguaje;
- La capa de léxico que contiene las manifestaciones léxicas en diferentes lenguas de los conceptos en la capa ontológica.

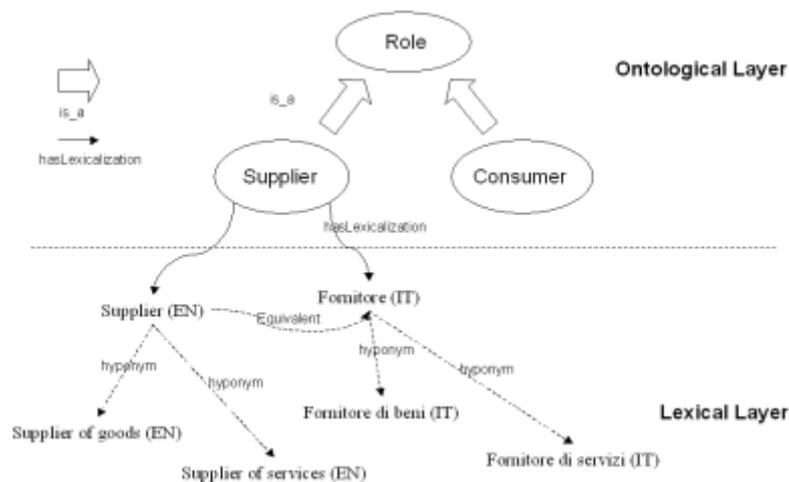


Figura.2.69. El sistema de organización del conocimiento (*KOS*) del recurso *DALOS* (Francesconi y otros, 2007: 109).

Sobre al tipo de lenguaje utilizado por las ontologías legales, se trata del fruto de la evolución de los lenguajes de representación del conocimiento. Como señalan Gómez y otros (2008), en un primer momento estaban basados en lógica de primer orden (*KIF*, Genesereth y Fikes, 1992), en marcos combinados con dicha lógica (*Ontolingua*, Farquhar y otros, 1997; Gruber, 1992). Con el auge de internet surgieron lenguajes de implementación de la web semántica como *HTML* (Raggett y otros, 1999), *XML* (Bray y otros, 2000), *RDF* (Lassila y Swick, 2000), *RDF Schema* (Brickley y Guha, 2004), *OIL* (Horrocks y otros, 2000), *DAML+OIL* (van Harmelen y otros, 2001) y *OWL* (Dean y Schreiber, 2004). De todos ellos, los más frecuentemente utilizados son *DAML+OIL*, *RDF* y *OWL*. Los demás están ya en desuso. *RDF* fue desarrollado por el *W3C* (*World Wide Web Consortium*) para describir recursos web. *RDF Schema* es una extensión de *RDF*. Ambos lenguajes sólo pueden representar conceptos, taxonomías de conceptos y relaciones binarias. Por su parte *OWL* (*Web Ontology Language*) amplía la expresividad de *RDF*. Hay tres tipos de lenguaje *OWL* (*Lite*, *DL* y *Full*) según los niveles de representación e inferencia en la ontología.

En cuanto a su función, podemos clasificar las ontologías legales en cinco tipos (Valente y otros, 2005):

- i) *comprensión de un dominio con una aplicación más general en derecho*
  - Representación del conocimiento legal (*LLD*) de McCarty (1989). Utiliza el lenguaje general para analizar el discurso legal. Base de conocimiento con una estructura profunda.
  - La ontología *FOLaw* de Valente y Breuker (1994). Base de conocimiento con una estructura profunda que tiene una arquitectura general basada en *Ontolingua*.
  - Ontología de marcos (*FBO*) de Van Kralingen (1995) y Visser (1995). Representación del conocimiento legal a través del lenguaje general basada en *Ontolingua*.
  - La base de conocimiento legal de Mommers (2001). Base de conocimiento con una estructura profunda que utiliza el lenguaje general para expresar el conocimiento legal.

- Ontología *LRI-Core* de Breuker y Hoekstra (2004b). Compatible con la adquisición del conocimiento para las ontologías legales. Base de conocimiento en *DAML+OIL/RDF* utilizando *Protégé* (convertida a *OWL*).
- Ontología *LKIF-Core* de Hoekstra y otros (2007; 2009). Compatible con la adquisición del conocimiento para las ontologías legales. Base de conocimiento en *OWL* con estructura profunda.
- Ontología de causalidad legal de Lehman, Breuker y Brower (2004). Base de conocimiento con estructura superficial que expresa la causalidad legal.
- Ontología de no expertos de Klein (2006). Ontología que representa el conocimiento no experto en casos de responsabilidad civil.
- Ontología *Micro-ontology* de Desprès y Szulman (2007). Ontología que representa conceptos en las normativas europeas. Desarrollada en *OWL* y *PLN* por el método *Terminae*.

*ii) indización y búsqueda semántica*

- Ontología de la profesión judicial *OPLK* de Benjamins y otros (2003). Base de conocimiento altamente estructurada en *Protegé* con un sistema inteligente de preguntas frecuentes para jueces. Ontología convertida en *OWL* con semántica de acceso a *Iuriservice*, sistema inteligente de preguntas frecuentes de la práctica jurídica para jueces noveles.
- Ontología de la profesión judicial *OPJK* de Casellas y otros (2004). Versión mejorada del sistema inteligente de preguntas frecuentes basadas en *Iuriservice*. Versión mejorada de *OPLK*. Base de conocimiento altamente estructurada en *Protegé*. Última versión en *OWL*.
- Ontología del código francés de Lame (2002). Base de conocimiento léxica del derecho francés. Base de conocimiento con una estructura superficial orientada al procesamiento del lenguaje natural (léxica).
- Ontología de fraudes fiscales de Leary, Vandenberghe y Zeleznikow (2003). Base de conocimiento en *UML (Unified Modeling Language)*.

*iii) organizan y estructuran información*

- *Jur-Wordnet* de Gangemi, Sagri y Tiscornia (2003). Aplicación de *Wordnet* al dominio legal. Base de conocimiento léxico en *DOLCE* con una estructura superficial.

- Ontología penal italiana de Asaro y otros (2003). Base de conocimiento con una estructura superficial que establece en *UML* un esquema de clasificación de los delitos del sistema penal italiano.
- Ontología de textos legales de Saias y Quaresma (2003). Base de conocimiento léxica en *OWL (Web Ontology Language)* con un proceso semi-automático de creación de la ontología a partir de los textos.
- Ontología *NM-L* de Shaheed y otros (2005). Ontología de representación de conceptos de alto nivel basada en la ontología nuclear *NM*.

*iv) con motor de razonamiento y resolución de problemas*

- Ontología *CLIME* de Boer, Hoekstra y Winkels (2001). Base de conocimiento del derecho marítimo en *Protegé* y *RDF*.
- *ArgumentDeveloper* de Zelenikow y Stranieri (2001). Representación del conocimiento que opera con varias bases de conocimiento jurídicas.

*v) otras*

- Ontología *IPROnto* de Delgado y otros (2001). Integra esquemas *XML DTDs* y esquemas que definen los lenguajes de expresión de los derechos de autor y los diccionarios de datos de los derechos de autor. La primera versión de la base de conocimiento estaba en *DAML+OIL* (2001). La versión actual (2008) está en *OWL*. Su función es la interoperabilidad entre los sistemas de gestión de derechos digitales (*DRM*).
- La ontología *e-Sentencias* de Casanovas y otros (2009). Es una ontología procedimental para los sistemas multimedia en los juicios. Ontología que representa las etapas del procedimiento de las vistas civiles en España.
- La ontología *OCL.NL* de Breuker y otros (2002). Ontología del derecho penal holandés en protocolo *OKBC (Open Knowledge Base Connectivity)*.
- La ontología *DALOS* de Francesconi y otros (2007). Se trata de la redacción de documentos legales en *OWL* en un entorno multilingüe (inglés, español, italiano y holandés).

Las ontologías legales han aumentado considerablemente en los últimos años. En la tabla 6 resumimos las veintitrés ontologías legales clasificadas por Breuker y otros (2009) de acuerdo a su aplicación, tipo, función, carácter (general o específica), modo de construcción (manual, semi-automática o automática) y lenguaje (monolingüe,

bilingüe o multilingüe) a las que hemos añadido las categoría de lenguaje de representación.

Ontología o proyecto	Aplicación	Tipo	Función	Lenguaje de representación	Carácter	Construcción	Lengua
LLD de McCarty	Lje. Gral. que expresa conocim. legal	Repres. del conocim.	Comp. de un dominio	-	General	Manual	Inglés
FOLaw de Valente y Breuker	Arquitect. gral. para la resoluc. de prob. legales	Base de conocim.	Comp. de un dominio	Ontolingua	General	Manual	Inglés
FBO de Van Kralingen y Visser	Lje. Gral. que expresa conocim. legal	Repres. del conocim.	Comp. de un dominio	Ontolingua basada en marcos	General	Manual	Inglés
Base de conocim. Legal de Mommers	Lje. Gral. que expresa conocim. legal	Base de conocim.	Comp. de un dominio	-	General	Manual	Inglés
LRI-Core de Breuker y Hoekstra	Adquis. del lje. para ontologías del dominio legal	Base de conocim.	Comp. de un dominio	DAML+ OIL / RDF utiliza Protégé para convertir a OWL	General	Manual	Inglés
LKIF-Core de Hoekstra y Breuker	Adquis. del lje. para ontologías del dominio legal	Base de conocim.	Comp. de un dominio	OWL-DL	General	Manual	Inglés
JurWordNet de Gangemi, Sagri y Tiscornia	Extensión de WordNet al dominio legal	Base de conocim. léxica	Organiza y estruct. la informac.	DOLCE (DAML)	General	Manual	Inglés
OPLK de Benjamins y otros	Sistema inteligente de preguntas frecuentes para jueces (Iuriservice)	Base de conocim.	Indizac. y búsq. semántica	Protégé (convertida a OWL)	Dominio	Semi-automática	Español
OPJK de Casellas y otros	Sistema inteligente de preguntas frecuentes para jueces Iuriservice (2ª versión)	Base de conocim.	Indizac. y búsq. semántica	Protégé	Dominio	Manual	Español

Tabla 2.6. Las ontologías legales.

<b>Ontología o proyecto</b>	<b>Aplicación</b>	<b>Tipo</b>	<b>Función</b>	<b>Lenguaje de representación</b>	<b>Carácter</b>	<b>Construcción</b>	<b>Lengua</b>
Ontología de códigos franceses de Lame	Recuperación de la información legal	Base de conocim. orientada al PLN	Indizac. y búsq. semántica	XML	Dominio	Automática	Francés
Ontología de fraudes financieros de Leary y otros	Recuperación de la información legal	Base de conocim.	Indizac. y búsq. semántica	UML	Dominio	Manual	Inglés
Ontología del derecho penal de Asaro y otros	Esquema de repres. de los delitos en el derecho penal italiano	Base de conocim.	Organiza y estructura la información	XML	Dominio	Manual	Italiano
CLIME de Boer y otros	Sistema de asesoramiento legal para el derecho mercantil	Base de conocim.	Razonam. y resolución de problemas	Protégé y RDF	Dominio	Manual	Inglés
LCO (Legal Causation Ontology) de Lehman y otros	Representación de la causalidad en el dominio legal	Base de conocim.	Comp. de un dominio	-	Dominio	Manual	Inglés
IPROnto (Intellectual Property Rights) de Delgado y otros	Definir lenguajes de expresión de derechos y diccionarios de datos de derechos	Base de conocim.	Interop. entre sistemas digitales de gestión de derechos de autor	DAML +OIL (2001) OWL-DL (2008)	Dominio	Manual	Inglés
e-Sentencias de Teodoro y otros	Representación de etapas procesales en vistas civiles	Conocim. procesal	Clasif. de grabaciones oficiales (imagen y video)	RDF	Dominio	Manual	Español

Tabla 2.7. Las ontologías legales (cont.).

Ontología o proyecto	Aplicación	Tipo	Función	Lenguaje de representación	Carácter	Construcción	Lengua
Ontología de la abogacía portuguesa de Saias y Quaresma	Enriquecimiento o semántico de los textos legales	Lógica de programac.	Organiza y estructura la información	OWL	Dominio	Automática	Portugués
Ontología para no expertos de M. Klein	Representac. del conocimiento no experto en casos de responsabilidad civil	Base de conocim. en lenguaje natural	Comp. de un dominio	OWL	Dominio	Semi-automática	Holandés
OCL.NL de Breuker y Elhag	Derecho penal holandés	OKBC	Estruct, principal del derecho penal holandés	-	Dominio / General	Manual	Holandés / Inglés
DALOS de Tiscornia y otros	Edición de textos legales	-	Soporte edición multiling. de textos legales e indización de docs.	OWL	Dominio	Manual	Inglés
Micro-ontology de Despres y Szulman	Representación de conceptos en la normativa europea	PLN (Método TERMINA E)	Comp. de un dominio	OWL	Dominio	Semi-automática	Francés / Inglés
Ontología UCC de Shaheed y otros	Representación de conceptos de alto nivel	Ontología de alto nivel NML basada en NM	Organiza y estructura la informac.	-	Dominio (alto nivel)	Manual	Inglés
CLO de Schweighofer y Liebwald	Gestión de la información	Representación de marcos	-	Protégé	General	Manual con soporte para ontologías legales nucleares	Inglés?
IRC de Melz y Valente	Ontología del Código de Hacienda (EEUU)	-	Razonador de casos de impuestos	OWL	Dominio	Manual	Inglés

Tabla 2.8. Las ontologías legales (cont.).

## **2.8. Conclusiones**

En este capítulo hemos presentado, por un lado, la problemática que presenta la definición de ontología y su clasificación, ya que se llama “ontología” a elaboraciones muy diversas y no se distingue entre ontologías propiamente dichas, representaciones de conocimiento, bases de conocimiento y taxonomías léxicas. Esto hace que los autores también difieran en el modo de clasificarlas, por lo que se hacía imprescindible señalar desde el principio la definición de “ontología” a adoptar en esta tesis. Además, dada la diversidad de clasificaciones sobre la amalgama de ontologías existentes hasta el momento, motivada por los distintos enfoques adoptados por los autores de la materia, hemos optado por elaborar nuestra propia clasificación de ontologías donde hemos analizado las más significativas basándonos en la literatura existente en ingeniería ontológica. Para complementar este marco, también hemos examinado las metodologías, lenguajes y herramientas de diseño de ontologías, como etapa previa al análisis del estado de la cuestión de las ontologías legales.

Por otro lado, hemos visto que hay ontologías legales que versan sobre la jurisprudencia y otras sobre conocimiento y terminología legal. En el caso de nuestra ontología del crimen organizado y el terrorismo, los términos son definidos de forma conceptual en inglés, por lo que son independientes de cada lengua. Esto significa que, como se trata de una ontología legal en inglés, español e italiano, la ontología será común a las tres lenguas, pero habrá un lexicón específico para los términos legales ingleses, otro para los españoles y otro para los italianos. Como el derecho es una ciencia social, el lenguaje general utiliza términos del derecho en su uso cotidiano, como por ejemplo “abogado”, “delito”, “juez”, etc. que son conceptos básicos de la ontología nuclear en *FunGramKB*, lo cual constituye un ejemplo del sentido común incorporado a la ontología. De este modo se reutiliza la ontología general en el diseño de la ontología satélite a la vez que se minimiza la redundancia informativa.

Nuestra ontología del crimen organizado y el terrorismo está basada en la semántica profunda, ya que tiene un enfoque conceptualista en lugar de lexicista. De este modo se evitan los problemas presentados por el léxico específico de cada lengua ya que los conceptos son unidades abstractas independientes de la lengua que se trate.

En cuanto a su aportación, nuestra ontología del crimen organizado y el terrorismo ofrece una nueva perspectiva en el ámbito de las ontologías legales, ya que pretende estructurar la información para una fácil recuperación en contextos profesionales y la resolución de tareas en situaciones reales. Mediante un corpus multilingüe se pretende ofrecer información conceptual que servirá para facilitar la comunicación a la vez que se reducirá la confusión a nivel internacional (cf. Ureña, Alameda y Felices, 2011).

A continuación, en el capítulo 3, describimos en detalle la estructura de *FunGramKB* como base de conocimiento cuya ontología nuclear es el eje angular en la que se asienta nuestra ontología satélite del crimen organizado y el terrorismo.



## Capítulo 3

# FUNGRAMKB

### 3.1. Introducción

En este capítulo presentamos la base de conocimiento *FunGramKB* y el lenguaje de representación COREL (COncceptual REpresentation LAnguage) que da cohesión a todo el módulo conceptual de la base de conocimiento.

De acuerdo a la psicología cognitiva del modelo de memoria a largo plazo de Tulving (1985), el conocimiento del sentido común que conforma el razonamiento humano consta de 3 partes (Periñán y Mairal, 2010b: 13):

- La memoria semántica, que representa un tesoro mental ya que almacena información cognitiva sobre las palabras.
- La memoria procedimental, que almacena información de cómo percibimos los eventos que realizamos en la vida real, por ejemplo cómo conducir. Es como manual de situaciones cotidianas.
- La memoria episódica, que almacena información de eventos biográficos, por ejemplo una fecha de aniversario. Es como un álbum de recortes personal.

Como podemos ver en la figura 3.1, estos tres tipos de conocimiento aparecen reflejados en el módulo conceptual de *FunGramKB*:

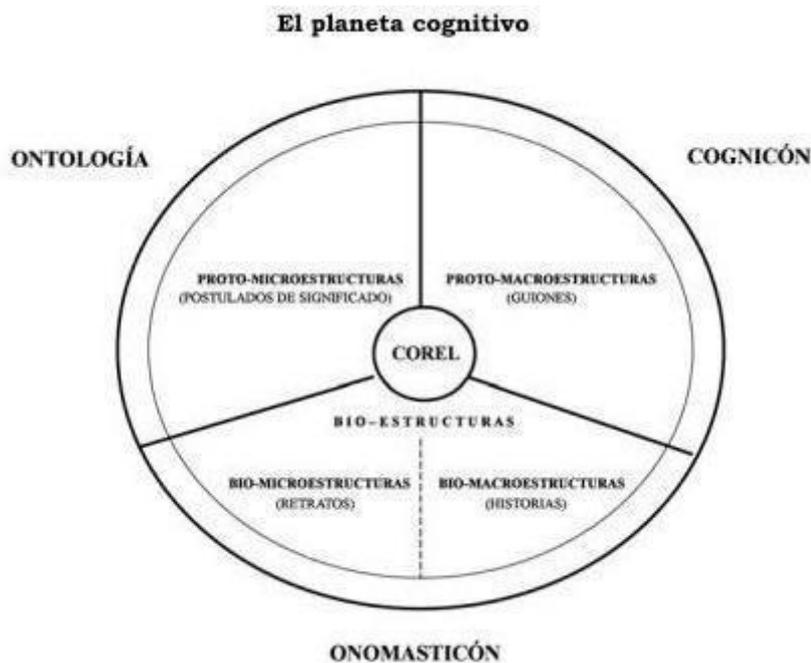


Figura 3.1. El planeta cognitivo (Periñán y Mairal, 2010b: 15)

Así, la Ontología incluye el conocimiento semántico a través de proto-microestructuras (los postulados de significado); el Cognición recoge el conocimiento procedimental en un conjunto de proto-macroestructuras (guiones); y el Onomasticón incluye conocimiento episódico en forma de dos tipos de bio-estructuras: los retratos y las historias.

*FunGramKB*<sup>65</sup> es una base de conocimiento léxico-conceptual multipropósito que ha sido diseñada para la realización de tareas de procesamiento del lenguaje natural. Es multipropósito, multifuncional y multilingüe porque puede ser reutilizada en diversas tareas (recuperación y extracción de información, sistemas de diálogo, traducción automática, etc) y en diversas lenguas<sup>66</sup>. Con respecto a la multilingüedad, Aguado y otros (2007) presentaron un detallado análisis de las estrategias más comunes en varios sistemas de conocimiento. Entre todos los modelos revisados, *FunGramKB* coincide con el modelo de representación del conocimiento que establece enlaces entre la ontología y las distintas lenguas que la componen. *FunGramKB* consta de un lexicon para cada lengua pero de una ontología única en la que las relaciones entre unidades

<sup>65</sup> <http://www.fungramkb.com>

<sup>66</sup> *FunGramKB* ha sido diseñada para trabajar con siete lenguas: alemán, búlgaro, catalán, español, francés, inglés e italiano.

léxicas de distintas lenguas se establecen a nivel conceptual. Además, *FunGramKB* está formada por tres niveles principales de conocimiento, léxico, gramatical y conceptual que tienen módulos independientes pero que están interrelacionados entre sí (cf. Perrián y Arcas, 2011):

Nivel léxico:

- El Lexicón almacena información morfosintáctica, pragmática y de colocaciones de las unidades léxicas.
- El Morficón asiste al analizador y generalizador en el tratamiento de la morfología flexiva.

Nivel gramatical:

- El Gramaticón almacena los esquemas constructivos que pueden ser utilizados por el algoritmo de enlace sintáctico-semántico de la Gramática del Papel y la Referencia (GPR) (Van Valin y LaPolla, 1997 ; Van Valin, 2005).

Nivel conceptual:

- La Ontología presenta un catálogo jerárquico de los conceptos que un hablante tiene en mente cuando habla de situaciones cotidianas. El conocimiento semántico es expresado a través de postulados de significado. Así, la Ontología está compuesta de dos módulos: un módulo de propósito general, i.e. la Ontología Nuclear, y varios módulos terminológicos de dominios específicos, las ontologías satélites (medicina, derecho, ...etc).
- El Cognicón almacena conocimiento procedimental a través de guiones, es decir, esquemas conceptuales de eventos estereotípicos basado en el modelo temporal de Allen (1983). Los guiones nos permiten describir, por ejemplo, cómo conducir un coche.
- El Onomasticón almacena conocimiento enciclopédico, sobre instancias de entidades y eventos como Manuel Fraga o el 11-M. Aquí se almacena el conocimiento a través de dos tipos de esquemas (retratos e historias) ya que las instancias pueden describirse de manera sincrónica o diacrónica.

Como vemos en la Figura 3.2, en *FunGramKB*, cada lengua tiene su propio módulo léxico y gramatical, mientras que el módulo conceptual es compartido por todas las lenguas. Por tanto, el ingeniero de conocimiento construirá un léxico para cada lengua, pero una única Ontología, Cognición y Onomasticón. Por tanto, *FunGramKB* tiene carácter conceptualista ya que la Ontología ejerce de módulo angular sobre el que se asienta toda la estructura de la base de conocimiento. A diferencia de otras bases léxicas en las que el significado se expresa sobre la base de relaciones superficiales entre unidades léxicas (*FrameNet* o *MultiWordnet*), *FunGramKB* se basa en la conocida “semántica profunda”, que muestra un mayor potencial expresivo [Periñán y Arcas, 2007; Felices y Marín, (en prensa)].

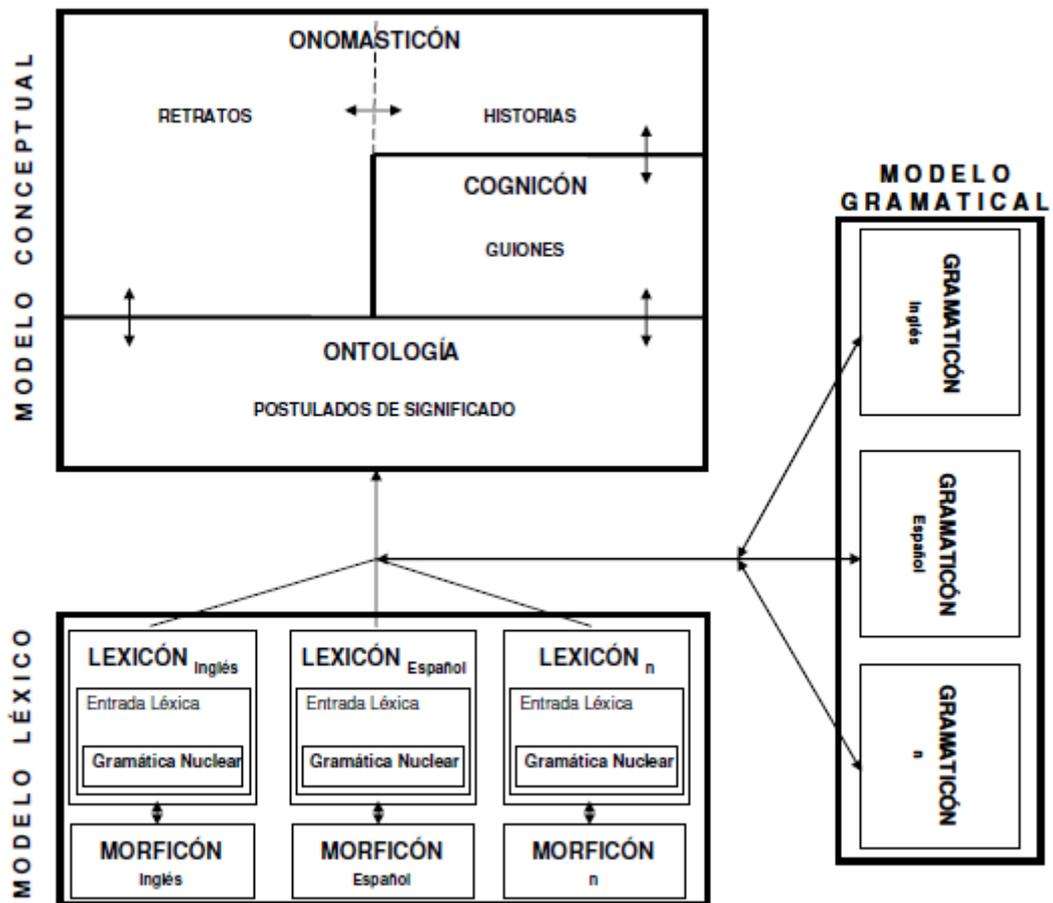


Figura 3.2. La arquitectura modular de *FunGramKB* (Periñán y Arcas, 2011: 3)

Como *FunGramKB* tiene un propósito general y no es de un dominio específico puede ser ampliada y enriquecida con conocimiento experto. Como éste proviene del lenguaje

general, la ontología nuclear se puede dotar de conocimiento experto a través de enlaces con ontologías satélites, como se muestra en la Figura 3.3.

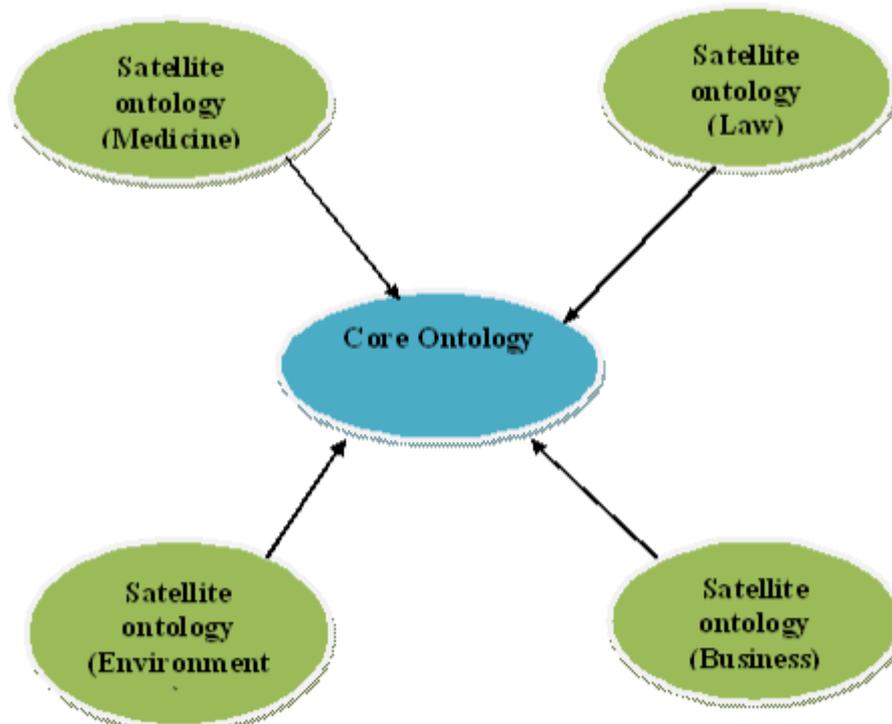


Figura 3.3. La ontología nuclear y las ontologías satélites (Mairal, Perrián y Samaniego 2011: 5).

Como la ontología nuclear es de conocimiento general y no de un dominio específico, es posible dotar a *FunGramKB* de conocimiento experto a través de ontologías satélite vinculadas a la nuclear. Así, en esta tesis mostramos cómo el modelo multinivel de *FunGramKB* se puede exportar a un modelo subontológico terminológico, i.e. a una ontología satélite. Como se mostrará en la sección 4.2., mediante el modelado de los denominados “concepto espejo” se reduce la redundancia informativa a la vez que se reutiliza la información conceptual de la ontología nuclear.

### 3.2. Las unidades conceptuales

Como señalan Perrián y Arcas (2010b), la estructura de la Ontología de *FunGramKB* se caracteriza por estar compuesta por tres tipos de unidades conceptuales:

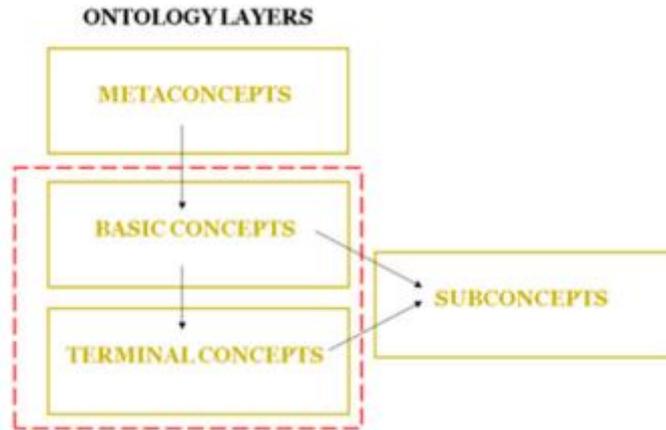


Figura 3.4. Jerarquía conceptual en *FunGramKB* (Jiménez, Luzondo y Cabello, 2011b: 19).

- a) Los metaconceptos. Es el nivel superior y está formado por 42 metaconceptos que representan dimensiones cognitivas en mayúsculas precedidas por el signo [#] como resultado del análisis de las ontologías lingüísticas más relevantes como SUMO (Niles y Pease, 2001a, 2001b), DOLCE (Gangemi y otros., 2002; Masolo et al., 2003). Como ejemplos de metaconceptos, de los eventos **#COMMUNICATION**, **#MATERIAL**, **#PSYCHOLOGICAL**,...etc; de las entidades **#ABSTRACT**, **#PHYSICAL**, **#PROCESS**,...etc; y de las cualidades **#PHYSICAL**, **#PSYCHOLOGICAL**, **#SOCIAL**,...etc, según observamos en las figuras 3.5, 3.6 y 3.7 respectivamente.

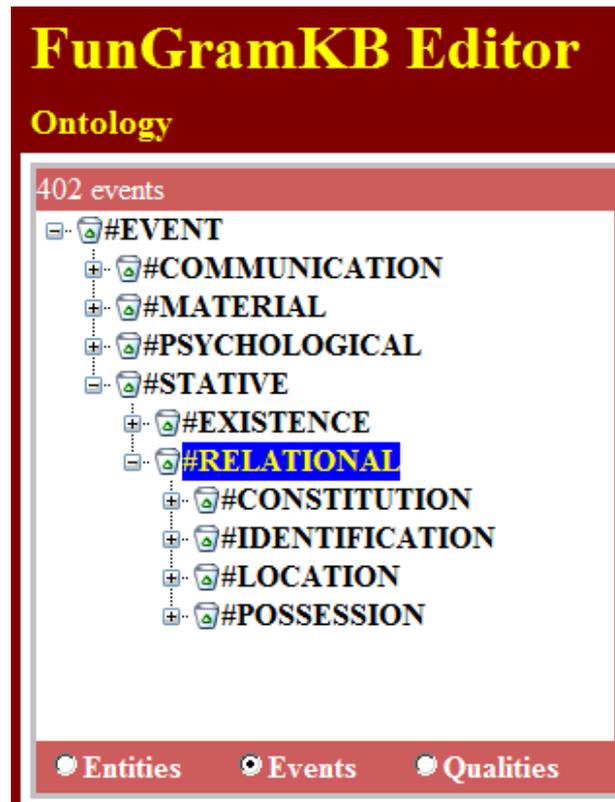


Figura 3.5. Representación parcial de los metaconceptos de los eventos

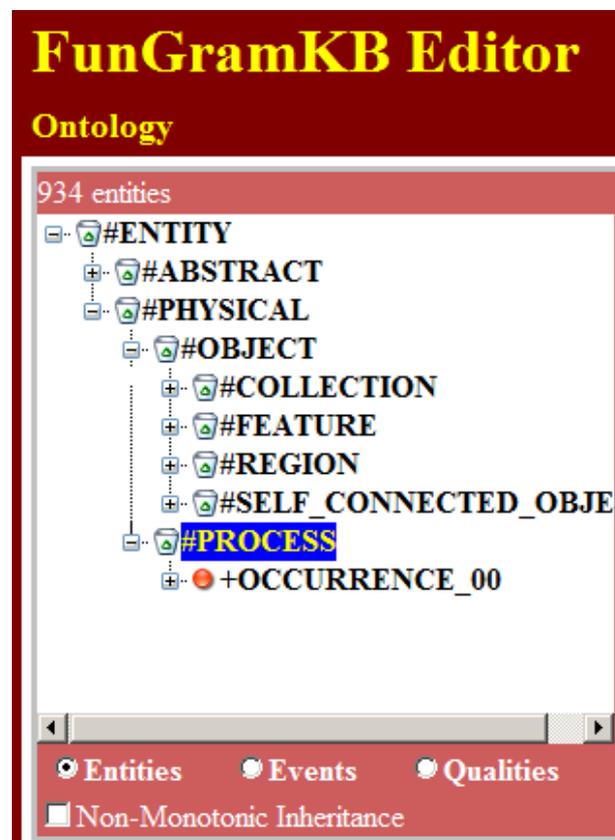


Figura 3.6. Representación parcial de los metaconceptos de las entidades

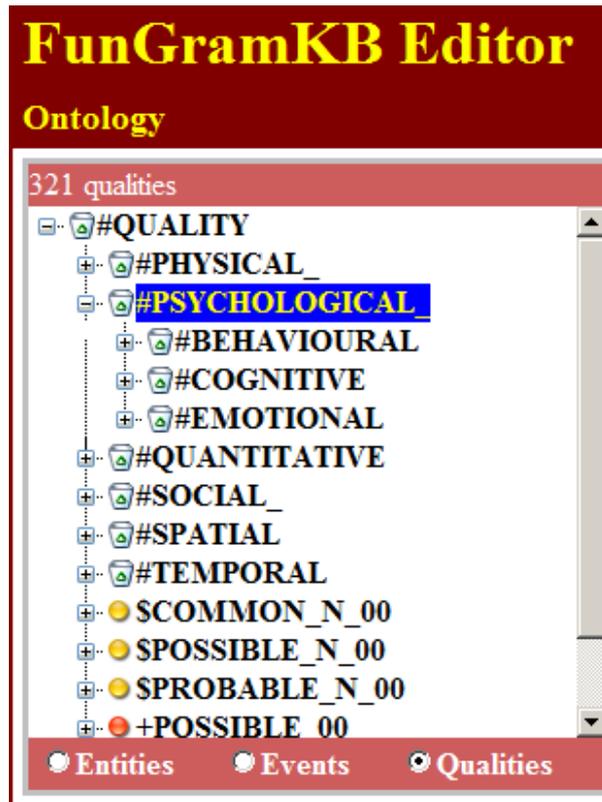


Figura 3.7. Representación parcial de los metaconceptos de las cualidades

- b) Los conceptos básicos. Inicialmente identificados a partir del vocabulario definitorio del *Longman Defining Dictionary of Contemporary English* (Procter, 1978), recurso fundamental en el procesamiento del lenguaje natural. El catálogo final de conceptos básicos es el resultado de las fases de conceptualización, jerarquización, remodelación y refinamiento descritas por Perrián y Mairal (2011). Están en un nivel intermedio y se representan con el signo [+] seguido de un guión bajo y dos ceros (por ejemplo los eventos +**THINK\_00**, +**BLAME\_00**, etc; las entidades +**GROUP\_00**, +**LAW\_00**; las cualidades +**ALONE\_00**, +**BAD\_00**, etc). Se utilizan en los postulados de significado y como preferencias de selección en los marcos temáticos. En las figuras 3.8, 3.9 y 3.10 incluimos una visión parcial de los conceptos básicos y terminales en las dimensiones conceptuales **#EVENT** > **#PSYCHOLOGICAL** > **#COGNITION**, **#ENTITY** > **#PHYSICAL** > **#OBJECT** > **#COLLECTION** y **#QUALITY** > **#SOCIAL** respectivamente:

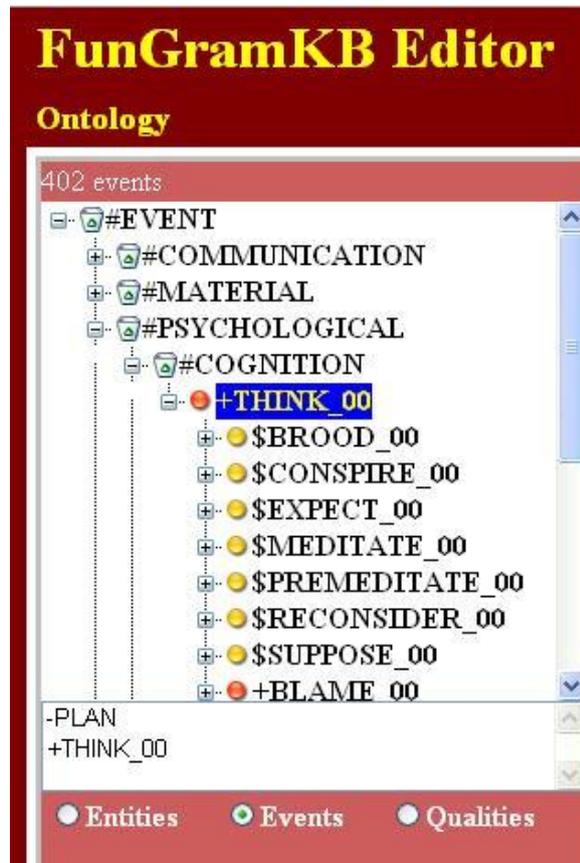


Figura 3.8. Representación parcial de los eventos

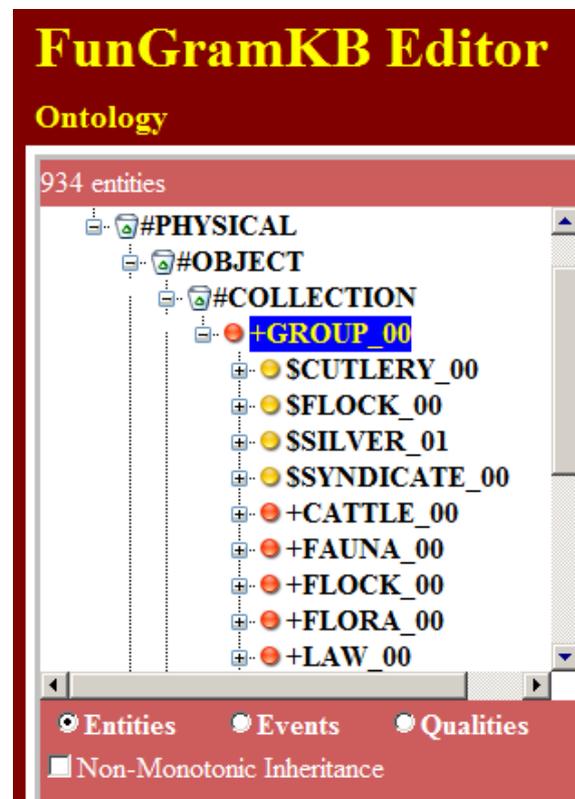


Figura 3.9. Representación parcial de las entidades

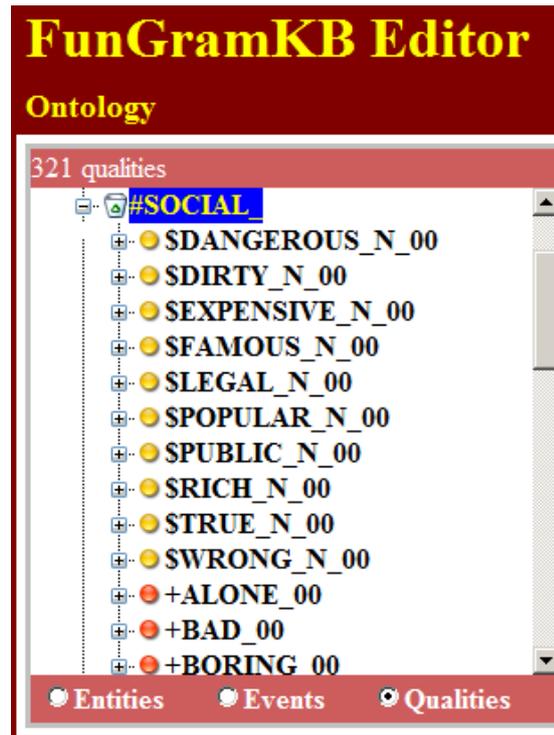


Figura 3.10. Representación parcial de las cualidades

- c) Los conceptos terminales. Representan el nodo final de la taxonomía y aparecen representados por el signo [\$] y también seguidos de un guión bajo y dos ceros (por ejemplo \$CONSPIRE\_00 en la figura 3.8, \$SYNDICATE\_00 en la figura 3.9 y \$LEGAL\_N\_00 en la figura 3.10 respectivamente).

Una gran mayoría de estos conceptos son lo que Periñán y Mairal (2010b: 17) llaman “conceptos de motivación léxica”, ya que deben estar lexicalizados en al menos una lengua de las que componen *FunGramKB*. Ésta es una característica obligatoria para los conceptos terminales pero no para los conceptos básicos, que pueden funcionar como “conceptos paraguas” que facilitan la organización de la taxonomía (véase el caso de +ORGANISED\_CRIME\_00 en las secciones 4.2. y 4.3.) Otra característica esencial de estos conceptos es que no son primitivos semánticos como los formulados por Goddard y Wierzbicka (2002). El Marco Temático (MT) y el Postulado de Significado (PS) se utilizan para matizar las propiedades semánticas de estos conceptos. Ambos tienen una base conceptual ya que se construyen con conceptos y no con unidades léxicas. Por tanto, los MT y los PS son representaciones del conocimiento semántico independientes de cada lengua. Así, el concepto en *FunGramKB* se concibe como una

“representación mental estructurada”. De hecho la ontología nuclear utiliza COREL<sup>67</sup> (COncceptual REpresentation LAnguage) como lenguaje de representación mental con una sintaxis y una semántica similar a la de las lenguas naturales para así modelar el significado de sus conceptos (Periñán-Pascual y Mairal Usón 2010: 18).

### 3.3. La gramática de COREL: el marco temático y el postulado de significado

El marco temático (en adelante MT) es un constructo conceptual que especifica los participantes prototípicos de un evento en una situación cognitiva. Al especificar los participantes del MT no lo hacemos de un verbo concreto, sino de su dimensión cognitiva. Por ejemplo, si consideramos la dimensión cognitiva #**MATERIAL**, identificamos dos papeles temáticos *Theme* (Tema) y *Referent* (Referente):

- (a) (x1) Theme -> entidad que realiza un evento voluntariamente o no;
- (b) (x2) Referent -> entidad que forma parte del evento causado por otra entidad.

Así, las dimensiones cognitivas (metaconceptos) están caracterizadas por unos papeles temáticos concretos (apéndice 1) que nos sirven para construir los MT correspondientes. Si tuviéramos que especificar las propiedades del concepto básico +**STEAL\_00** convendríamos que está subordinado al concepto básico +**OBTAIN\_00**, el cual a su vez depende del metaconcepto #**MATERIAL**. Como éste metaconcepto tiene asignado un MT prototípico (1) todos los conceptos subordinados a este metaconcepto heredarán dicho MT.

- (1) (x1) Theme (x2) Referent

Por tanto, una de las propiedades del evento +**STEAL\_00** será este MT (2), al que añadiremos una descripción detallada de su significado a través del (PS) (3).

- (2) MT: (x1:+**HUMAN\_00**)Theme (x2:+**ARTEFACT\_00**)Referent
- (3) PS: +(e1: +**OBTAIN\_00** (x1)Theme (x2)Referent (f1: +**PLACE\_00**)Origin (f2: +**CRIME\_00**)Result)

---

<sup>67</sup> Véase Periñán y Arcas (2004); Periñán y Mairal (2010b) para una descripción detallada del lenguaje COREL.

Como podemos observar, los MT se integran en los PS. Así, los papeles temáticos de cada metaconcepto ofrecen una descripción detallada de las instancias prototípicas de una dimensión cognitiva dada. Con un inventario reducido de papeles temáticos podemos marcar las características diferenciadoras de cada dimensión cognitiva. El “tema” es el papel temático central en torno al cual giran todos los demás y que adquiere una definición diferente en cada dimensión cognitiva. De hecho, el “tema” caracteriza la propia denominación utilizada, “marco temático” ya que constituye el papel conceptual central en todos los metaconceptos. Perrián y Mairal (2010b: 20) señalan que el inventario de papeles temáticos se estableció de acuerdo a los criterios de obligatoriedad y especificidad semántica de Koenig y otros (2003) por el cual un participante es argumento semántico de un verbo si y solo si cumple los criterios de obligatoriedad y especificidad semántica. Los participantes semánticos pueden ser argumentos (x) o satélites (f) (de acuerdo a la Gramática Funcional de S. Dik, 1997) dependiendo de si es necesario para dar sentido completo al verbo en cuestión o si simplemente aporta información adicional sobre las circunstancias en las que se desarrolla la acción verbal.

El inventario de papeles temáticos en *FunGramKB* no difiere sustancialmente de los roles semánticos propuestos por Halliday (1985) y Dixon (1991) como señalan Perrián y Mairal (2010b: 35).

El PS es la expresión más detallada del significado de la unidad conceptual. Un PS es un conjunto de una o más predicaciones (**e1, e2, e3,...**) unidas lógicamente. Es un constructo cognitivo que representa las características genéricas de un concepto, no de una unidad léxica. Además de recoger conocimiento semántico, el PS ha de recoger el conocimiento del sentido común, que no aparece en ningún diccionario.

Como características fundamentales de un PS están los rasgos nucleares y los ejemplares. La función de los primeros es nuclear, mientras que la de los segundos es identificativa. Así, los rasgos nucleares aparecen reflejados en la ontología a través de predicaciones estrictas [+] y los rasgos ejemplares a través de predicaciones rebatibles [\*]. Por ejemplo, “una vaca es un mamífero” es una predicación estricta porque es siempre cierta, mientras que “los gatos tienen cuatro patas” es una predicación rebatible ya que puede darse el caso de que un gato sólo tenga tres patas.

A la hora de construir los PS, el ingeniero de conocimiento debe decidir cuánto conocimiento debe almacenar sobre un concepto, es decir, ha de decidir sobre el grado de granularidad de las definiciones. El criterio seguido en *FunGramKB* es similar al de Laurence y Margolis (1999: 29) por el cual el concepto debe aportar una cantidad considerable de información pero no demasiada como para hacer del concepto algo rígido (Periñán y Mairal, 2010b: 24).

### 3.3.1. Las predicaciones

Los PS pueden estar formados por una o más predicaciones que representan cada uno de los rasgos de un concepto<sup>68</sup>. Cada una está representada por una variable **e** indizada (**e1**, **e2**, **e3**,...) seguida por dos puntos (:) y precedida por un operador de razonamiento (ver sección 3.3.2). A modo de ejemplo presentamos el PS de la entidad **+PUNISHMENT\_00** (4), del evento **+STEAL\_00** (5) y de la cualidad **+LEGAL\_00** (6):

- (4a) +(e1: +BE\_00 (x1: +PUNISHMENT\_00)Theme (x2: +OCCURRENCE\_00)Referent)  
 +(e2: +PUNISH\_00 (x3: +HUMAN\_00)Theme (x4: +HUMAN\_00 ^+ANIMAL\_00)Referent (f1: x1)Scene)
- (5a) +(e1: +OBTAIN\_00 (x1)Theme (x2)Referent (f1: +PLACE\_00)Origin (f2: +CRIME\_00)Result)
- (6a) +(e1: +BE\_01 (x1)Theme (x2: +LEGAL\_00)Attribute)  
 +(e2: +PERMIT\_00 (x3: +LAW\_00)Theme (x1)Referent (x4: +HUMAN\_00)Goal)

A continuación presentamos los equivalentes aproximados en lenguaje natural de las predicaciones anteriores.

- (4b) Un castigo es una situación.

<sup>68</sup> De hecho sólo unos pocos conceptos básicos carecen de PS por ser primitivos semánticos. En estos casos aparece *sp* (del inglés *semantic primitive*) en lugar del PS, por ej. +DO\_00 (ver Periñán y Mairal 2010b: 27).

Una persona castiga a otra o a un animal en una situación determinada.

(5b) Alguien obtiene algo de forma ilícita.

(6b) Algo que es legal es permitido por la ley.

En el caso de las entidades, las predicaciones deben hacer referencia a su *definendum*, ya sea explícitamente en la primera predicación o por coindización, ya que cada predicación expresa un rasgo de la entidad. En cuanto a las cualidades, ocurre algo similar, pero en lugar del *definendum* es la entidad la que tiene que aparecer en cada predicación.

En *FunGramKB* encontramos dos tipos de predicaciones en sus PS: libres y ligadas (Periñán y Mairal, 2010b: 29). Las predicaciones libres son aquellas conceptualmente completas, mientras que las predicaciones ligadas dependen de otras para ser una especificación conceptual completa. Se presentan entre dos paréntesis `[( ( ) ]` que sirven para marcar el ámbito de actuación del ligamiento cognitivo. Veamos el ejemplo de `$$SYNDICATE_00` (7) a continuación:

```
(7) *((e1: +BE_00 (x1: $$SYNDICATE_00)Theme (x2:
+GROUP_00)Referent)(e2: +COMPRISE_00 (x2)Theme (x3: s
+COMPANY_00)Referent))
```

Las siguientes son características de las predicaciones ligadas:

- (i) Sólo pueden ligarse predicaciones contiguas. Normalmente, se ligan dos predicaciones, aunque pueden ser más.
- (ii) Debe existir coindización hacia algunos de los participantes de la primera predicación (**x2** en el ejemplo de `$$SYNDICATE_00`) a partir de la segunda predicación ligada,. De hecho las predicaciones ligadas aportan una especificación conceptual fundamental a una entidad de la primera predicación.
- (iii) Se omite el operador de razonamiento delante de cada una de las predicaciones ligadas salvo en el paréntesis de inicio del ligamiento, que debe ir precedido por dicho operador. Veamos el ejemplo de `+BLAME_00` (8):

- (8)  $+(e1: +SAY\_00 (x1)Theme (x4: (e2: past +DO\_00 (x3)Theme (x2)Referent))Referent (x3)Goal)(e3: +BE\_01 (x2)Theme (x5: +BAD\_00)Attribute))$

Aunque aparentemente parezca que hay una predicación principal y otra “subordinada conceptualmente”, a nivel cognitivo se trata de una predicación compleja. **+BLAME\_00** no se puede definir con dos rasgos diferentes, i.e. “culpar es decir algo” y “alguien hizo algo malo”, sino con un rasgo complejo, “culpar es decir que alguien o algo hizo algo mal o es responsable de cosas malas”.

Puede ocurrir que las predicaciones funcionen como preferencias de selección a través de uno o varios conceptos, pero esto sólo es posible para las entidades y las cualidades. Como vemos en (9), el PS de **+ANSWER\_00**. En el caso de las cualidades sólo pueden realizar el papel temático de *Attribute*, *Frequency*, *Position* o *Speed*.

- (9)  $+(e1: +SAY\_00 (x1)Theme (x2)Referent (x3)Goal (f1: (e2: past +SAY\_00 (x3)Theme (x4: +QUESTION\_00)Referent (x1)Goal))Scene)$

En (9) las preferencias de selección del satélite *f1* se expresa a través de una predicación; mientras que en el argumento *x4* se ve enriquecido por la entidad **+QUESTION\_00**.

En *FunGramKB* hay tres tipos de operadores posibles en las predicaciones (Periñán y Mairal, 2010b: 31): los operadores de razonamiento, los operadores del evento y los operadores del participante, que describimos a continuación.

### 3.3.2. Los operadores de razonamiento y del evento

El ingeniero de conocimiento modela el conocimiento del mundo a través de los PS de cada concepto. Así encontramos dos tipos de operadores de razonamiento que actúan en las predicaciones en *FunGramKB*: **[+] estrictos** y **[\*] rebatibles**. Los operadores estrictos expresan rasgos que no pueden ser rebatidos bajo ninguna circunstancia (reglas sin excepciones), por ejemplo “un juez es una persona”. Los operadores rebatibles expresan información del sentido común como por ejemplo “la personas tienen la obligación de cumplir las leyes, pero pueden no hacerlo”. De hecho, el motor de razonamiento de *FunGramKB* interpreta las predicaciones estrictas como razonamiento

monotónico y las predicaciones rebatibles como razonamiento no monotónico. Veamos el PS de **+BIRD\_00** que indica que “un pájaro típicamente vuela” (10) aunque podría no ser el caso, como ocurre con **+PENGUIN\_00** (11).

(10) \*(e3: +FLY\_00 (x1)Agent (x1)Theme (x4)Origin (x5)Goal)

(11) +(e1: +BE\_00 (x1: +PENGUIN\_00)Theme (x2: +BIRD\_00)Referent)

...

(e9: n +FLY\_00 (x1)Theme)

Los operadores del evento adquieren distintos valores como se muestra en la siguiente tabla:

Atributo		Valor
Aspectualidad		ing   pro   egr
Temporalidad		rpast   past   npast   pres   nfut   fut   rfut
Modalidad	epistémico	cert   prob   pos
	no epistémico	obl   adv   perm
	Polaridad	n

Tabla 3.1. Operadores de predicación (Periñán y Mairal, 2010b: 33)

Los operadores de aspectualidad expresan las cualidades que marcan el desarrollo de un evento, esto es, su inicio, su continuidad y su finalización, etapas que expresamos a través de los operadores *ing* (ingresivo), *pro* (progresivo) y *egr* (egresivo).



Figura 3.11. Operadores de aspectualidad (Periñán y Mairal, 2010b: 34)

(12) i. María empezó a reír.

ii. María estaba riendo.

iii. María dejó de reír.

Los operadores de temporalidad tienen la función de ubicar la predicación en un eje temporal. La figura 3.12 ofrece las diferentes partes del eje temporal.

<b>pasado remoto</b>	<b>pasado</b>	<b>pasado cercano</b>	<b>presente</b>	<b>futuro cercano</b>	<b>futuro</b>	<b>futuro remoto</b>
(rpast)	(past)	(npast)	(pres)	(nfut)	(fut)	(rfut)
<b>i</b>	<b>ii</b>	<b>iii</b>	<b>iv</b>	<b>v</b>	<b>vi</b>	<b>vii</b>
Referencia temporal						

Figura 3.12. Operadores de temporalidad (Periñán y Mairal, 2010: 34)

- (13) i. María había cocinado.  
 ii. María cocinó.  
 iii. María ha cocinado.  
 iv. María está cocinando.  
 v. María está a punto de cocinar.  
 vi, vii. María cocinará.

Como ejemplo de uso del operador de temporalidad mostramos el PS de **+BLAME\_00** (14):

- (14) +((e1: +SAY\_00 (x1)Theme (x4: (e2: past +DO\_00 (x3)Theme (x2)Referent))Referent (x3)Goal)(e3: +BE\_01 (x2)Theme (x5: +BAD\_00)Attribute))

En cuanto a los operadores de modalidad, *FunGramKB* distingue entre epistémicos y no-epistémicos. Así, entre los epistémicos están *cert* (certeza), *prob* (probabilidad), y *pos* (posibilidad) para expresar el grado de certeza del hablante con respecto al contenido de la predicación. Por otro lado, los operadores no-epistémicos tienen un carácter deóntico y reflejan los siguientes valores: *obl* (obligación), *adv* (consejo) y *perm* (permiso). Por ejemplo, en el PS de **+PERMIT\_00** (15) encontramos un operador modal no-epistémico “una persona le dice a otra que *puede* hacer algo”.

- (15) +(e1: +SAY\_00 (x1)Theme (x4: (e2: perm +DO\_00 (x3)Theme (x2)Referent))Referent (x3)Goal)

Para finalizar hay que destacar el operador de polaridad, que en *FunGramKB* se utiliza para expresar la negación fuerte, por ejemplo “X es falso”. Se trata de una cuestión clave para entender la naturaleza no monotónica de los operadores de razonamiento mencionada anteriormente en esta sección, donde la falta de conocimiento no implica su falsedad. Por tanto, el operador *n* difiere de lo que se conoce como negación débil, la cual indica que “nada en esta base de conocimiento confirma que A sea cierto” (Periñán y Mairal, 2010b: 35). Por ejemplo, (16) presenta el PS de **+REFUSE\_00**, i.e. “decimos que no haremos algo”, donde el operador de negación se combina con un operador de temporalidad.

(16) +(e1: +SAY\_00 (x1)Theme (x4: (e2: fut n +DO\_00 (x1)Theme  
(x2)Referent))Referent (x3)Goal)

### 3.3.3. Los operadores de participante: los cuantificadores y los operadores lógicos

Los conceptos que actúan como preferencias de selección en un participante pueden ir acompañados de operadores que expresan cuantificación. La Tabla 3.2 presenta los operadores que el ingeniero de conocimiento puede utilizar en *FunGramKB*.

Atributo	Valor
Cuantificador Absoluto	1   2   3   4 ...
Cuantificador Relativo	m   s   p
Cuantificador de Pluralidad	i

Tabla 3.2. Los operadores de cuantificación (Periñán y Mairal, 2010b: 36)

Para ilustrar los operadores cuantificación ofrecemos los PS (17) y (18) donde **\$FLOCK\_00** (17) incluye un cuantificador relativo, por ejemplo “una bandada es un grupo de pájaros”; por otro lado **+BODY\_00** (18) ilustra tanto el cuantificador absoluto como el relativo, por ejemplo “un cuerpo tiene una cabeza, varias extremidades y un tronco”.

(17) \*((e1: +BE\_00 (x1: \$FLOCK\_00)Theme (x2: +GROUP\_00)Referent)(e2: +COMPRISE\_00 (x2)Theme (x3: m +BIRD\_00)Referent))

(18) \*(e3: +COMPRISE\_00 (x1)Theme (x4: 1 +HEAD\_00 & s +LIMB\_00 & 1 +TRUNK\_00)Referent)

Es importante señalar que el uso de los cuantificadores no es obligatorio y que sólo podemos utilizar un cuantificador para cada preferencia de selección. Si conocemos el número exacto de entidades utilizaremos el cuantificador absoluto, de lo contrario utilizaremos un cuantificador relativo que expresan *m* (muchos), *s* (algunos) y *p* (pocos). Aplicaremos el operador *i* en el caso de que la cantidad sea indeterminada.

Un caso especial ofrecen las cualidades, ya que utilizamos los cuantificadores relativos para expresar la gradación de una determinada dimensión semántica. Si queremos expresar los distintos grados de temperatura, podemos describir que una entidad está (19) muy caliente (“abrasador”), (20) un poco caliente (“cálido”), (21) muy fría (“gélido”), o (22) un poco fría (“fresco”) (cf. Perrián y Arcas, 2008: 139; Perrián y Mairal, 2010b: 37).

- (19) +(e2: +BE\_01 (x1)Theme (x3: m +HOT\_00)Attribute)
- (20) +(e2: +BE\_01 (x1)Theme (x3: p +HOT\_00)Attribute)
- (21) +(e2: +BE\_01 (x1)Theme (x3: m +COLD\_00)Attribute)
- (22) +(e2: +BE\_01 (x1)Theme (x3: p +COLD\_00)Attribute)

Para concluir, señalar brevemente que hay tres tipos de conectores lógicos en *FunGramKB* válidos en COREL, i.e. la conjunción [**&**], la disyunción [**|**] y la exclusión [**^**] (figura 3.13).

Tipo	Operador	Ámbito
Conjunción	&	Entre satélites o preferencias de selección.
		Entre predicaciones, argumentos o satélites.
Disyunción		Entre satélites o preferencias de selección.
Exclusión	^	Entre satélites o preferencias de selección.

Figura 3.13. Los conectores lógicos (Perrián y Mairal, 2010c: 13)

Los operadores lógicos operan entre satélites o preferencias de selección, no entre argumentos (participantes). En el ejemplo de **+BODY\_00** (18) veíamos un ejemplo de conjunción [**&**] entre preferencias de selección. La relación lógica entre dos argumentos

es siempre la conjunción, pero la gramática de COREL no permite formalizarla en ese caso ya que es redundante porque el razonador entiende el encadenamiento de dos predicaciones o participantes.

A continuación ofrecemos el PS de **+OFFEND\_00** (23) y el PS de **\$FLATTER\_00** (24) para ejemplificar el conector de disyunción [ | ] entre preferencias de selección y satélites respectivamente.

(23) +(e1: +DO\_00 (x1)Theme (x3)Referent (f1: (e2: n +BE\_01 (x1)Theme (x4: +POLITE\_00)Attribute (f2: x2)Goal))Manner (f3: (e3: +BE\_01 (x2)Theme (x5: +SAD\_00 | +ANGRY\_00)Attribute)))Result)

(24) +(e1:+SAY\_00 (x1)Theme (x2)Referent (x3)Goal (f1: (e2: +FEEL\_00 (x2)Agent (x3)Theme (x4: +HAPPY\_00)Attribute))Purpose | (f2: (e3: +OBTAIN\_00 (x1)Theme (x5)Referent))Purpose)

En (25) ofrecemos un ejemplo del conector de exclusión [ ^ ] entre satélites en el PS de **+HANDLE\_00**.

(25) \*(e3: +TAKE\_00 (x4: +HUMAN\_00)Theme (x1: +HANDLE\_00) Referent (f1: (e4: +MOVE\_00 (x4)Agent (x3)Theme (x5) Location (x6)Origin (x7)Goal))Purpose ^ (f2: (e5: +HOLD\_00 (x4)Theme (x3)Referent))Purpose)

### **3.4. Conclusiones**

En este capítulo hemos ofrecido una descripción detallada de la base de conocimiento *FunGramKB* y del lenguaje de notación COREL. Esta base de conocimiento multilingüe permite el desarrollo de tareas de PLN (recuperación y extracción de la información, traducción automática, diccionarios de búsquedas conceptuales,...) en varias lenguas. Mediante el lenguaje de notación COREL, que consta de una sintaxis y semántica propias que se utilizan para la construcción de MT y PS, se da cohesión al módulo conceptual de *FunGramKB*. Con este enfoque conceptualista u ontológico se ofrece una mayor expresividad en la representación del conocimiento a la vez que se permite acceso al conocimiento enciclopédico, que hasta el momento se ha probado inabarcable en lingüística.

Así, hemos elegido *FunGramKB* para el diseño y desarrollo de una ontología del crimen organizado y el terrorismo porque consta de un modelo teórico y herramientas propias reutilizables en la elaboración de cualquier ontología satélite, independientemente de cuál sea su dominio temático. Además, frente a otras ontologías basadas en la semántica superficial, en las que la representación del significado de una palabra se establece por las relaciones que tiene con otras palabras (p. ej. *Wordnet*), el modelo multinivel de la ontología nuclear de *FunGramKB* en metaconceptos, conceptos básicos y terminales permite la representación conceptual reflejo de la estructura cognitiva humana. Por todo ello consideramos que la base de conocimiento *FunGramKB* nos ofrece el marco teórico y las herramientas más idóneas para el desarrollo de nuestra ontología satélite del crimen organizado y el terrorismo.



## **Capítulo 4**

# **DESARROLLO DE UNA ONTOLOGÍA SATÉLITE EN FUNGRAMKB A PARTIR DEL GLOBAL CRIME TERM CORPUS (GCTC)**

### **4.1. Introducción**

Con el desarrollo creciente de la cooperación judicial internacional en materia penal surgen nuevas líneas de investigación relativas al lenguaje específico utilizado en el ámbito legal a la vez que se hace necesario el perfeccionamiento de herramientas de procesamiento de la información de acceso al conocimiento experto. Como hemos destacado en el capítulo 2, una de las principales herramientas para organizar y conceptualizar la información legal de manera efectiva reside en la creación de ontologías, que representan conocimiento compartido y que pueden ser reutilizadas y compartidas por varias aplicaciones de software y grupos de personas (Gómez, Fernández y Corcho, 2004: 8-9).

Para poblar una ontología, los terminólogos e ingenieros del conocimiento compilan corpus para su posterior explotación. En este capítulo presentamos las cuestiones metodológicas relativas a la compilación del Global Crime Term Corpus (GCTC) que pretende satisfacer algunas de las demandas en el dominio legal y de la construcción de ontologías legales. Dicho corpus pretende servir de base en el proceso de elaboración de una ontología satélite del crimen organizado y el terrorismo que tendrá como propósito estructurar la información para una fácil recuperación en contextos profesionales y la resolución de tareas en situaciones reales. Mediante un corpus multilingüe se pretende ofrecer información conceptual que servirá para facilitar la comunicación a la vez que se reducirá la confusión terminológica a nivel internacional (cf. Ureña, Alameda y Felices, 2011: 302).

Antes de presentar el Global Crime Term Corpus (GCTC) conviene señalar las nociones de “corpus” y “termino” adoptadas aquí. Según Meyer (2002: xi-xii), la definición de “corpus” depende de lo amplia que la queramos hacer. Así, en la evolución de la lingüística de corpus, vemos cómo la noción de “corpus” ha ido cambiando. Francis (1982: 17) y Sinclair (1991: 171) destacaban que un “corpus” ha de ser suficientemente representativo de una lengua o de una variedad de una lengua, ejemplos (1) y (2) respectivamente:

- (1) (...) a corpus is a collection of texts assumed to be representative of a given language, dialect, or other subset of a language to be used for linguistic analysis.
- (2) (...) a corpus is a collection of naturally-occurring language text chosen to characterize a state or variety of a language.

Atkins, Clear y Ostler (1992: 1) añadían en su definición que el corpus debe tener como requisito ser recogido para un propósito específico.

- (3) (a corpus is) a subset of an ELT (Electronic Text Library) built according to explicit design criteria for a specific purpose.

Según *The Expert Advisory Group on Language Engineering Standards (EAGLES)* (1996) y Tognini-Bonelli (2001: 2), en los ejemplos (4) y (5) respectivamente, el “corpus” está compuesto de un grupo de “extractos del lenguaje” o textos, que sirven de muestra representativa del lenguaje y que son analizados lingüísticamente:

- (4) Corpus: a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language.
- (5) A corpus can be defined as a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis. (...) in general there is consensus that a corpus deals with natural, authentic language.

El grupo de trabajo de córpora textuales *EAGLES* (1996) llega incluso a destacar la noción de “computer corpus” como:

- (6) A corpus which is encoded in a standardised and homogenous way for open-ended retrieval tasks. Its constituent pieces of language are documented as to their origins and provenance.

Por su parte, Tognini-Bonelli (2001: 55) propone la siguiente definición, que es la que adoptamos aquí:

- (7) A corpus is taken to be a computerized collection of authentic texts, amenable to automatic or semiautomatic processing or analysis. The texts are selected according to explicit criteria in order to capture the regularities of a language, a language variety or a sublanguage.

Por tanto entendemos que un corpus es un conjunto de textos almacenados informáticamente para su posterior procesamiento automático o análisis. Los textos son seleccionados de acuerdo a criterios explícitos, para captar las regularidades de una lengua, una variedad de la lengua o un sublenguaje (entendiendo sublenguaje como lenguaje específico).

En el caso de nuestro corpus se han recogido textos específicos de la cooperación internacional en materia penal, y, más en concreto, documentación relacionada con el crimen organizado y el terrorismo en organizaciones de prestigio internacional tales como la Unión Europea, el Consejo de Europa o Naciones Unidas, entre otras.

Una característica fundamental del corpus es el propósito para el que es recogido, es decir, qué líneas de investigación pretende cubrir. Así, el corpus debe ser representativo del lenguaje objeto de investigación (Reppen, 2011: 31). De hecho, el material recogido debe corresponder al lenguaje objeto de estudio (Biber, 1993: 243). En otras palabras, si pretendemos investigar el lenguaje utilizado en artículos científicos de literatura inglesa, recopilar artículos periodísticos en inglés no nos servirá de nada. En cuanto al tamaño del corpus, como señala Flowerdew (2004: 18), no hay un tamaño ideal; todo depende de lo que el corpus contenga y de lo que se esté investigando. Así, la consideración principal a tener en cuenta en el diseño de un corpus reside en que debe estar diseñado de modo que sea adecuado al propósito de la investigación. Mientras que muchos grandes còrpora se recopilaron para investigar cuestiones lingüísticas generales, los

córpore especializados a menudo se diseñan para responder a cuestiones específicas de la investigación (Koester, 2011: 66).

Este autor también señala que, con respecto al grado de especialización, un corpus puede ser más o menos especializado, y puede ser especializado de distintas maneras, dependiendo del propósito de la investigación (ver Hunston 2002: 14). Flowerdew (2004: 21) enumera una serie de parámetros según los cuales un corpus puede ser especializado.

- Propósito concreto para su compilación, por ejemplo, para investigar una unidad gramatical o léxica determinada.
- Contextualización: situación particular, participantes y propósito comunicativo.
- Género, por ejemplo, para ascensos laborales (propuesta de beca, cartas de ventas),
- Tipo de texto / discurso, por ejemplo, textos de biología, conversación cotidiana.
- Tema o materia, por ejemplo, economía.
- Variedad del inglés, por ejemplo, el inglés como lengua extranjera.

En el caso de GCTC, el propósito del corpus es la extracción de terminología del dominio del crimen organizado y el terrorismo para elaborar una ontología satélite vinculada a la ontología nuclear de la base de conocimiento de *FunGramKB*. El motivo de elegir el subdominio legal del crimen organizado y el terrorismo fue por su relevancia internacional actual y por la probada falta de referencias para la población de ontologías. Así, los términos seleccionados como ganadores poblarán una ontología satélite específica del crimen organizado y el terrorismo que estará vinculada a la ontología nuclear en la base de conocimiento *FunGramKB*. Al corpus se le ha llamado Global Crime Term Corpus (GCTC) debido a que trata terminología del dominio legal. De acuerdo a las líneas marcadas por Felices, Ureña y Alameda (2011: 66-67), partimos de la hipótesis de que el modelo multinivel de la ontología nuclear de *FunGramKB* puede ser trasladado a ontologías satélite vinculadas a ésta. De este modo, al igual que ocurre en la ontología nuclear, se minimiza la redundancia informativa y se maximiza el conocimiento (Mairal, Perrián y Samaniego 2011: 268; Felices y Marín-R [en prensa]).

Así, mediante la vinculación de ontologías satélite a la ontología nuclear de *FunGramKB* hacemos que la base de conocimiento sea aún más robusta.

En cuanto a la noción de “terminología”, aunque no hay consenso entre los terminólogos en su definición, las propuestas por Cabré y Sager a continuación nos parecen muy acertadas. Según Cabré (1993: 169):

El conjunto de las palabras especializadas de una determinada disciplina (y también de un ámbito de actividad específica) constituye, como se sabe, la terminología propia de esa especialidad. (...) la peculiaridad más notable de la terminología, en contraste con el léxico común, consiste en que sirve para designar los conceptos propios de las disciplinas y actividades de especialidad.

Para Sager (1993: 42-43): “Los elementos caracterizados por una referencia especial dentro de una disciplina son los términos de esa disciplina, y colectivamente forman su terminología”. Como actualmente tampoco hay consenso general en cuanto a la definición de “término” adoptamos aquí la propuesta por Felices, Ureña y Alameda (2001: 67), basada en Cabré (1999) y Temmerman (2000), entre otros, según la cual un “término” es una unidad léxico-conceptual perteneciente al discurso especializado de un dominio de conocimiento concreto, como por ejemplo el de la medicina, el derecho o la tecnología y que no es característico de la lengua general.

Como hemos mencionado anteriormente, una cuestión fundamental en el desarrollo de un corpus representativo es la selección de las fuentes, esto es, los textos que componen el repositorio de documentos del corpus. Para la construcción de GCTC se consideraron como fuentes principales las instituciones internacionales y trabajos académicos. Para ello se seleccionó una serie de instituciones de las que obtener los documentos. Debido a las características del lenguaje objeto de estudio, las fuentes tenían que ser instituciones no sólo con relación a aspectos legales generales, sino también particularmente relacionadas con aspectos del terrorismo y el crimen organizado a nivel internacional. Además, tenían que ser entidades con páginas web oficiales con acceso libre para poder obtener los documentos en formato digital y así poder facilitar su procesamiento computacional. Después de una búsqueda exhaustiva, finalmente se

eligieron, entre otras, las siguientes instituciones: UN (United Nations), The Criminal Court of Justice, Europol (European Police Office), Eurojust (European Judicial Cooperation Unit) y OSCE (Organization for Security and Co-operation in Europe). Se ha probado que estas organizaciones son ampliamente representativas en la lucha contra el crimen organizado y el terrorismo, ya que ofrecen una rica representación de los aspectos técnicos y del vocabulario especializado que se utiliza oficialmente, tanto para expedir documentos oficiales en el cumplimiento de la ley como para combatir actos criminales y terroristas en un entorno global y democrático. Como consecuencia, estas instituciones podrían beneficiarse de los resultados de este trabajo de investigación. También se utilizaron otras fuentes, como obras de referencia académicas y artículos de investigación, debido a la gran concentración de términos especializados en ellos.

La gran cantidad de textos recogidos llevó a la creación de una base de datos que recogiera toda la información relevante relacionada con los textos. Así, la base de datos sirve como sistema de almacenamiento para una rápida identificación y acceso a los textos cuando sea necesario. Cada entrada incluye información sobre la lengua en la que está escrito el documento, el tipo de texto, la fuente de la que procede, una breve descripción del contenido y un código numérico de identificación final. De este modo cada documento tiene un código de etiquetado que incluye como información básica la lengua, el tema, el tipo de texto y el contenido. Debido a las limitaciones de procesamiento del extractor, en la base de datos hay (espacios incluidos) una limitación de 100 caracteres para la “brief description” y de 45 caracteres para “title”. En la figura 4.1 se puede ver un extracto de dicha base de datos.

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
3	English	Fight against organised crime	EOAct (joint) evaluating international undertaking	Organised Crime	Joint Action	Eur-Lex
4	Spanish	Fight against organised crime	SOAct (joint) evaluating international undertaking	Organised Crime	Joint Action	Eur-Lex
5	English	Combating organised crime	EODec (council) persons training cooperation	Organised Crime	Proposal Council Decision	Eur-Lex
6	Spanish	Combating organised crime	SODec (council) persons training cooperation	Organised Crime	Proposal Council Decision	Eur-Lex
7	English	Fight against transnational organised crime	EODec (council) combating persons trafficking	Organised Crime	Proposal Council Decision	Eur-Lex
8	Spanish	Fight against transnational organised crime	SODec (council) combating persons trafficking	Organised Crime	Proposal Council Decision	Eur-Lex
9	English	Prevention and fight against organised crime	ECom financial sector prevention	Organised Crime	Communication	Eur-Lex
10	Spanish	Prevention and fight against organised crime	SCom financial sector prevention	Organised Crime	Communication	Eur-Lex
11	English	Fight against organised crime	EODec (council framework) explanatory memorandum	Organised Crime	Proposal Council Framework	Eur-Lex
12	Spanish	Fight against terrorism financing	STCom terrorism financing	Terrorism	Communication	Eur-Lex

Figura 4.1. Base de datos.

Sirva a modo de ejemplo el siguiente código, “EBAct (plan) criminal matters cooperation”, donde se resume brevemente el contenido del texto. Así “E” significa “English”, “B” se utiliza para denominar “both”, es decir, tanto crimen organizado como terrorismo y “Act (plan)” significa “action plan”. A continuación, en la descripción, se detalla que se trata del plan de acción de Viena. Otros ejemplos serían los siguientes:

- (1) STCom terrorism financing
- (2) EBAct (plan) criminal matters cooperation
- (3) EODec (council) combating persons trafficking

Donde la primera letra indica la lengua en la que está escrito en texto, ya sea “S” (*Spanish*), “E” (*English*) o “I” (*Italian*); la segunda letra indica el subdominio al que pertenece el texto, en este caso “T” (*terrorism*), “B” (*both*) o “O” (*organized crime*); las tres primeras letras de la última palabra, en este caso, “Com” (*Communication*), “Act” (*Action plan*), “Dec” (*Decision*). La descripción del contenido del archivo aparece a continuación: “*terrorism financing*”, “*criminal matters cooperation*” y “*combating persons trafficking*” respectivamente. En el caso de que haya varios archivos con el mismo subtítulo, se añade un índice numérico para diferenciarlos, como por ejemplo: “*terrorism financing1*”, “*terrorism financing2*”, y así sucesivamente.

Una vez elaborada la selección y descarga de los documentos más relevantes, fue preciso realizar una serie de tareas de edición semiautomática para filtrar los errores tipográficos resultantes del formateo de los archivos de su formato original en “pdf” a texto plano. Este pre-procesamiento de los textos es necesario debido a las características del extractor de términos, parte de la suite de *FunGramKB*, que sólo funciona con texto plano. Esta fase es clave ya que, a mayor corrección del documento mayores posibilidades de extracción de términos en fases posteriores de la extracción terminológica. Hay un número indefinido de posibles errores ortográficos y tipográficos que pueden ocurrir durante el proceso de formateo a texto sencillo. A continuación se muestran algunas de las tareas de edición más frecuentes (a-i) (extracto de Felices, Ureña y Alameda (2011: 70):

- (a) Rediseño de textos de dos columnas. Convertir textos de dos columnas a una sola;

- (b) Corrección de errores ortográficos. Con frecuencia surgidos con “f”, “fl” o “ff”;
- (c) Corrección de intercalados erróneos entre el cuerpo del texto y los encabezados, membretes o títulos;
- (d) Corrección de intercalados erróneos de pies de páginas con el cuerpo del texto;
- (e) Corrección de sílabas o letras de palabras separadas por espacios (por ejemplo, “f i e l d”);
- (f) Corrección de sílabas o letras erróneamente separadas y unidas a otras palabras (por ejemplo, “crimi nalactivity” en vez de “criminal activity”);
- (g) Corrección de palabras divididas silábicamente mediante un guión al final de línea;
- (h) Eliminación de enlaces web o direcciones “url”;
- (i) Eliminación de direcciones de correo electrónico.

Para abordar esta tarea de manera efectiva se pueden utilizar programas como *Notepad++*<sup>69</sup>.

La estructura de este capítulo es la siguiente. Primero describimos la extracción terminológica de los términos (4.2.). Seguidamente mostramos la conceptualización de varios términos en lenguaje natural y COREL (4.3.), para finalmente proponer la taxonomía resultante de dichos términos (4.4.) y ofrecer unas conclusiones del capítulo (4.5.). Cabe destacar que los ejemplos presentados son sólo una muestra de la ontología satélite del crimen organizado y el terrorismo.

---

<sup>69</sup> *Notepad++* es una herramienta de código abierto y disponible gratuitamente en <http://notepad-plus-plus.org/>.

## 4.2. La extracción terminológica

Tras la compilación del corpus y la edición de los textos, la siguiente fase es la extracción semiautomática de los términos. La figura 4.2 muestra el proceso de compilación y filtrado en *FunGramKB*.

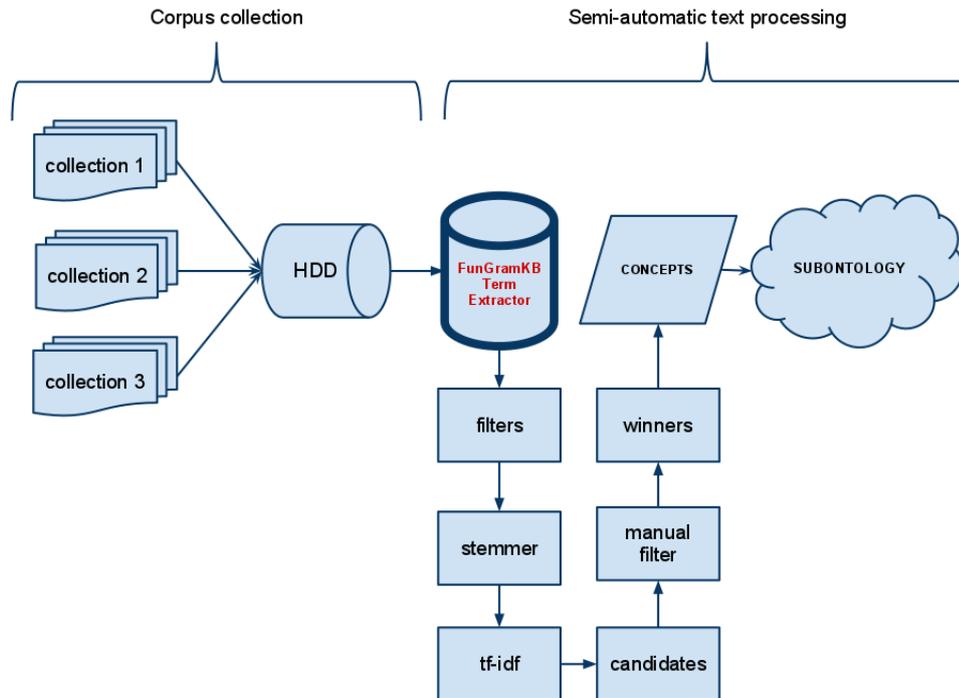


Figura 4.2. Proceso de compilación y filtrado en *FunGramKB*.

*FunGramKB Term extractor* es una herramienta semiautomática de extracción de términos que tiene un enfoque estadístico basado en el cálculo de la denominada *term frequency-inverse document frequency (tf-idf)*, que mide la importancia estadística o peso específico de cada término en un corpus dado. Este cálculo, expresado mediante un índice numérico, es el resultado de ponderar o “normalizar” la aparición de una palabra en un documento (*tf*) con la aparición de esta misma palabra en los distintos documentos de los que se compone el corpus (*idf*). Así, el *tf-idf* prioriza aquellos términos que tienen una frecuencia absoluta elevada pero un *idf* bajo, mientras que descarta aquellos que ocurren en muchos (o en la mayoría) de los documentos del corpus. De esta forma, cuanto mayor sea el índice *tf-idf* de un término, mayor la probabilidad de que sea un término especializado. Este índice matemático guiará al terminólogo en la elaboración de una lista de términos especializados (Felices, Ureña y Alameda, 2011: 68).

En la Figura 4.3 vemos los distintos elementos presentes en el extractor. Las herramientas “*pre-processing*”, “*processing (indexing)*” y “*processing (statistics)*” forman parte del proceso de subida de textos, donde el indizador “*processing (indexing)*” permite subir los textos de forma estructurada para su posterior procesamiento; el procesador estadístico “*processing (statistics)*” realiza el cálculo matemático de los índices *tf-idf* y presenta una lista de términos candidatos; finalmente la herramienta “*view*” permite al terminólogo filtrar los falsos candidatos por medio de una serie de opciones de eliminación (“*remove*” y “*nesting*”).



Figura 4.3. Pantalla principal del extractor de *FunGramKB*.

Como herramientas secundarias del extractor están “*search*”, el motor de búsqueda de unidades léxicas que facilita su localización en el corpus de textos, y la herramienta “*corpus*”, que es especialmente útil ya que nos permite consultar los documentos que conforman nuestro corpus especializado a la vez que presenta las estadísticas descriptivas básicas de los textos indizados además de ofrecer el número de palabras (*tokens*) incluidos. Como podemos ver en la figura 4.4, en la pantalla de visualización del “*corpus*” también se ofrece la lista de falsos candidatos desechados durante el proceso de filtrado.

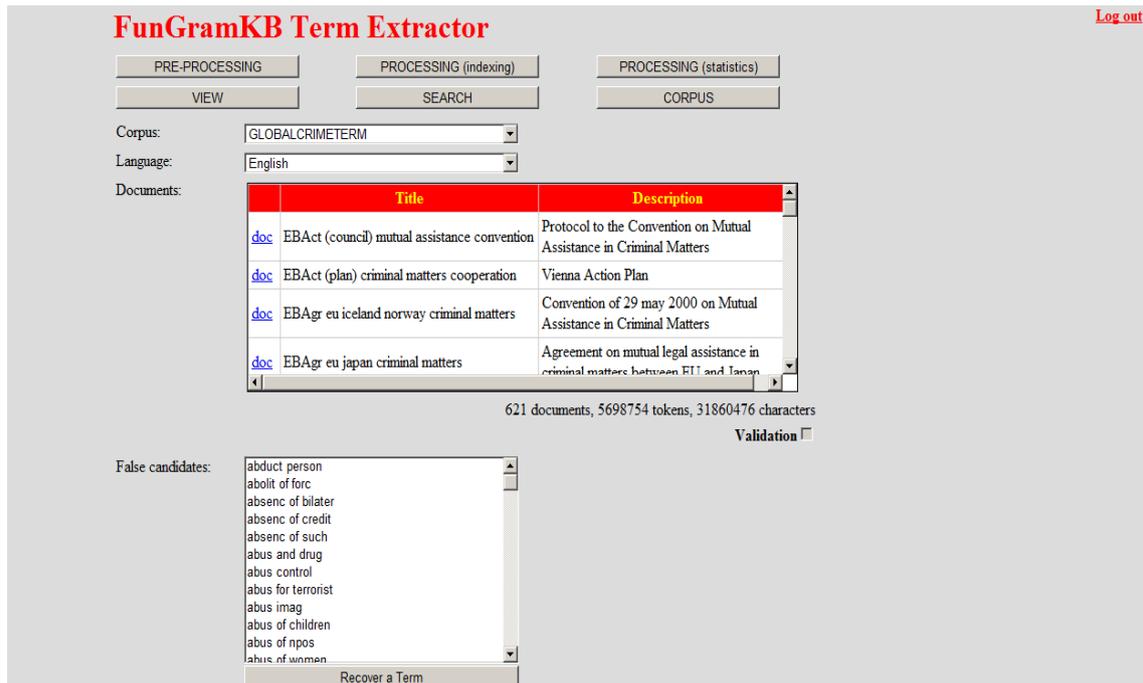


Figura 4.4. Corpus y papelera de reciclaje terminológica.

En la Figura 4.5 podemos ver ejemplos del corpus del crimen organizado y el terrorismo GCTC. Sólo para el inglés se han subido al extractor 621 textos, lo cual ha dado un resultado de más de 5 millones y medio de palabras (*tokens*).

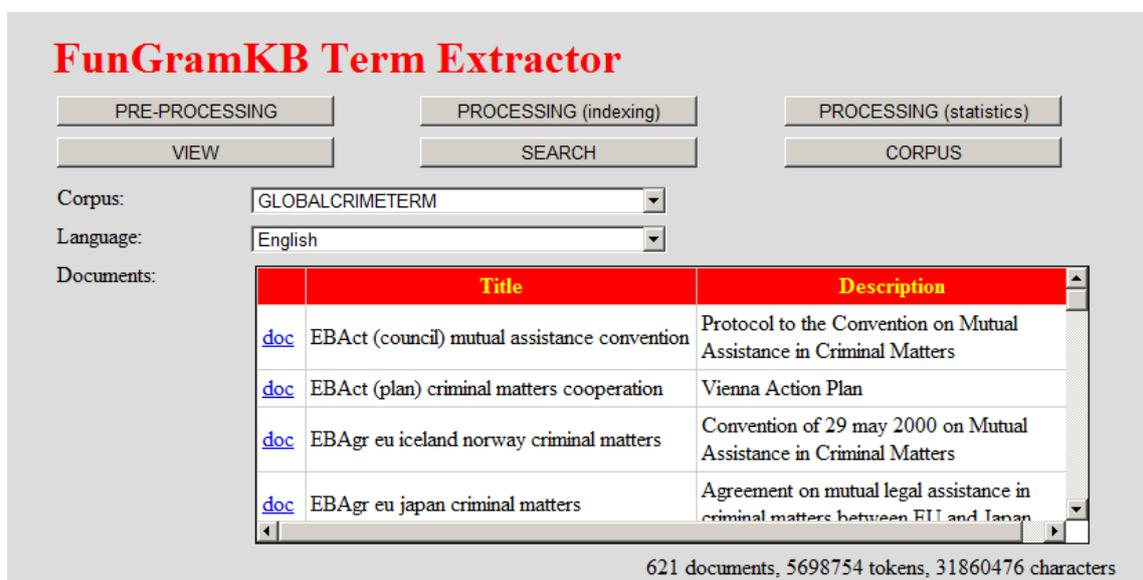


Figura 4.5. Pantalla de visualización del corpus.

Conviene señalar que la carga del corpus la realiza el terminólogo principal mediante la herramienta “*corpus (indexing)*” del extractor. El proceso de carga se divide en dos fases: la “*precarga*” y la “*carga*”. En la primera de estas fases, el terminólogo introduce cuatro parámetros en el extractor: (i) el documento concreto que va a cargar (archivo),

(ii) el título del texto (el del archivo *txt* que contiene el texto), (iii) una breve descripción del contenido principal del texto, y (iv) una etiqueta que identifique el subdominio al que pertenece el texto (por ejemplo, la geología, la medicina, o, en el caso del GCTC, el derecho), (Figura 4.6 y ver también Felices, Ureña y Alameda, 2011: 73-74).

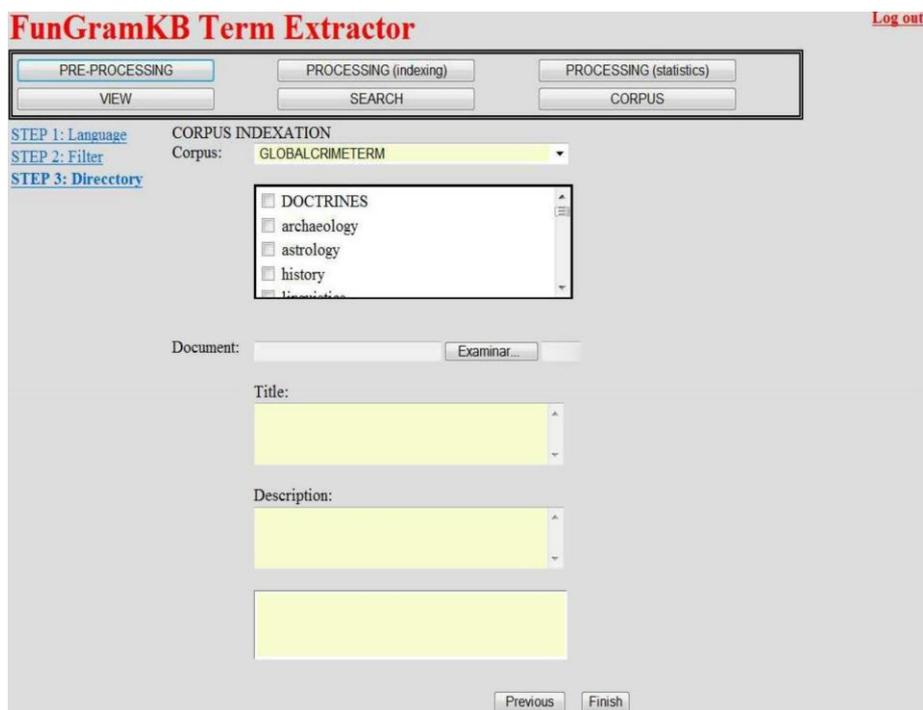


Figura 4.6. Pantalla de precarga del extractor de *FunGramKB*.

Debido a las limitaciones del extractor, tanto el título como la descripción no podrán superar los 40 y 100 caracteres respectivamente. Una vez seleccionados estos parámetros, el texto aparecerá precargado y entonces el terminólogo podrá acceder a la fase de carga, en la que podrá comprobar si el archivo, el título y la descripción son susceptibles de validarse de forma definitiva o si, por el contrario, es necesario hacer modificaciones. Para facilitar dicha comprobación, la ventana de carga muestra un resumen que incluye el título del texto precargado y su descripción correspondiente. En caso de que sea necesario hacer modificaciones, el terminólogo puede regresar a la fase de precarga para corregir errores. Si la precarga ha sido correcta, el terminólogo la valida y el texto se carga finalmente en el extractor. Los textos cargados aparecen entonces reflejados de instantáneamente en la pestaña “corpus” (Figura 4.7.).

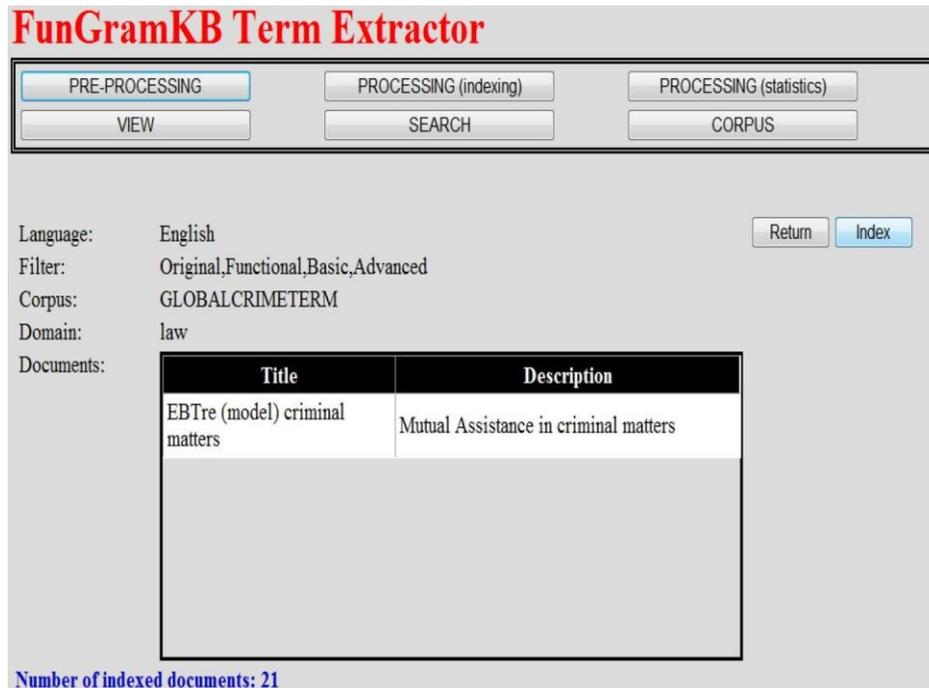


Figura 4.7. Pantalla de carga del extractor de *FunGramKB*.

Una gran ventaja que ofrece el extractor es la carga simultánea de varios documentos, hasta un máximo aproximado de diez, dependiendo del tamaño global de los archivos. Cabe señalar que no es posible eliminar ni modificar la información del título y descripción de un documento una vez cargado en el extractor, por lo que es conveniente asegurarse de comprobar convenientemente los datos en la etapa de precarga de los textos.

Como hemos mencionado anteriormente, en un futuro próximo GCTC estará compuesto de tres componentes diferentes representando las lenguas inglesa, española e italiana, de las cuales estas dos últimas no se han desarrollado todavía.

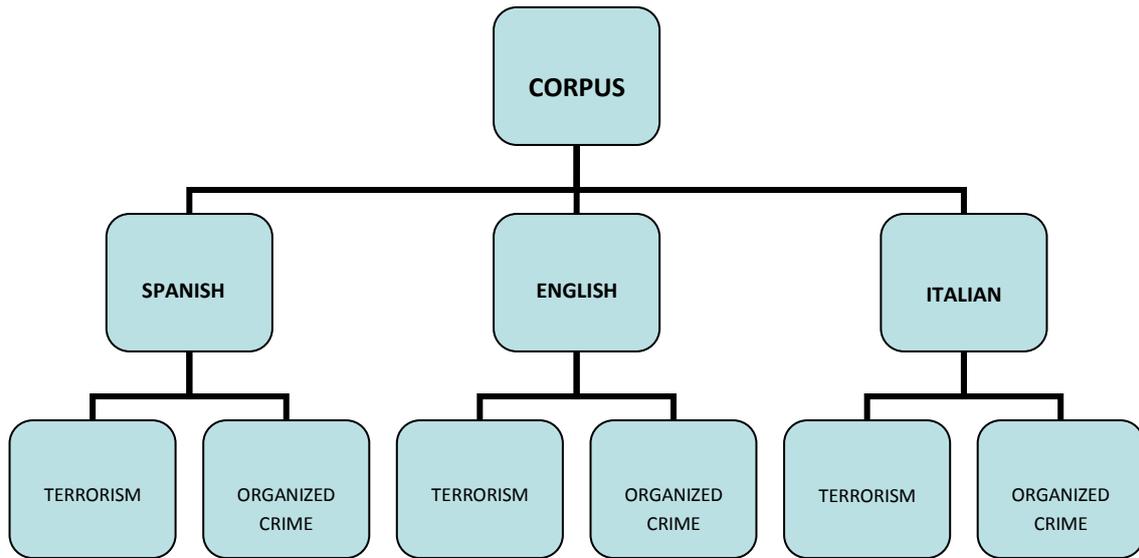


Figura 4.8. Componentes de GCTC.

Los pasos iniciales en el proceso de compilación del GCTC produjeron resultados significativos y satisfactorios que se describen a continuación. El trabajo realizado hasta ahora en el componente inglés presenta un corpus con textos del crimen organizado y el terrorismo que está representado por una gran variedad de tipos de texto, desde informes técnicos a declaraciones o reglamentos. Sólo el componente inglés consta de más de 5 millones y medio de palabras distribuidas en 621 documentos.

Como hemos señalado anteriormente, la representatividad es un criterio fundamental en la compilación de corpus. Así, se siguieron dos criterios fundamentales en la fase de compilación del GCTC: equilibrio y representatividad. Con equilibrio nos referimos a que el corpus contiene una representación justa entre todas las partes que lo componen. A este respecto, de los 621 documentos un 49% tratan del crimen organizado, 34% del terrorismo y el 17% restante son textos sobre ambas materias. La representatividad se refiere al hecho de que un corpus debe, idealmente, ser un claro ejemplo del lenguaje que pretende mostrar. A este respecto, la estructura robusta de GCTC reside principalmente en la gran variedad de textos que contiene y en la gran cantidad de palabras relacionadas con el dominio legal. La gran variedad de tipos de texto incluye, entre otros, informes, acuerdos, declaraciones, normativa, tratados y resoluciones que en total suman 45 tipos de texto (cf. Ureña, Alameda y Felices 2011: 304).

Antes de abordar la fase de extracción terminológica conviene matizar algunas nociones fundamentales en este proceso. En primer lugar destacamos la noción de “n-grama”. Un “n-grama” es una unidad de descripción básica dentro del ámbito del procesamiento computacional del lenguaje, que suele coincidir con lo que comúnmente denominamos “palabra” en su dimensión ortográfica. Así, hemos considerado tres tipos básicos de n-gramas en *FunGramKB Term Extractor*: los “unigramas”, que son palabras individuales (por ejemplo, *terrorism*), los “bigramas” o secuencias de dos palabras (por ejemplo, *organised crime*) y los “trigramas”, que son formaciones unificadas de tres palabras (por ejemplo, *transnational organised crime*). Así, el extractor clasifica los términos del corpus en unigramas, bigramas y trigramas, por lo que, en la fase de filtrado, el terminólogo trabajará con una lista de unigramas, bigramas y trigramas candidatos a término.

De acuerdo al contenido semántico y relevancia terminológica de los “n-gramas”, podemos clasificar cada unidad en “término candidato”, “término ganador” y “término falso”. Así, será “término candidato” aquel que tras un primer filtrado se posiciona como unidad terminológica potencialmente relevante en un dominio especializado concreto. Será “término ganador” un “n-grama” candidato propuesto por el extractor y validado por el terminólogo como término propio de un área de conocimiento concreto. Por último, será “término falso” aquel “n-grama” candidato desechado por el terminólogo por pertenecer al lenguaje general. En este caso son de gran utilidad los *Learners’ Dictionaries* ya que ayudan al terminólogo a discriminar las unidades terminológicas de las que no lo son. Otros casos de falsos candidatos son las combinaciones que no están conectadas, no tienen sentido o son meras colocaciones. Sólo los “términos ganadores” serán definidos en lenguaje natural y formalizados en COREL.

Tras concluir el proceso de carga de documentos, el extractor realiza el cálculo estadístico, índice *tf-idf* (ver más arriba), de todas las unidades del corpus mediante el cual extrae la lista de términos candidatos, que ha dado como resultado una lista de unigramas, bigramas y trigramas candidatos a término que podemos ver a través de la pestaña “view”, en las figuras 4.9, 4.10 y 4.11 respectivamente.

**FunGramKB Term Extractor** [Log out](#)

PRE-PROCESSING    PROCESSING (indexing)    PROCESSING (statistics)

VIEW    SEARCH    CORPUS

Corpus: GLOBALCRIMETERM    Terms

English

Unigram     Bigram     Trigram

Functional  
 Basic  
 Advanced

			Term	f	tf-idf		
view	edit	<input type="checkbox"/>	fatf	7299	202.37	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	money	12181	181.15	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	launder	10893	179.26	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	europol	7037	177.68	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	state	26978	171.83	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	financ	16540	169.74	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	offenc	10846	164.16	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	oc	1434	158.03	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	bank	6725	154.68	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	terrorist	9333	150.36	<input type="checkbox"/>	remove
view	edit	<input type="checkbox"/>	member	16020	148.79	<input type="checkbox"/>	remove

Terms: 12512    context: 50    fragments: 2

Figura 4.9. Tabla de unigramas candidatos.

**FunGramKB Term Extractor** [Log out](#)

PRE-PROCESSING    PROCESSING (indexing)    PROCESSING (statistics)

VIEW    SEARCH    CORPUS

Corpus: GLOBALCRIMETERM    Terms

English

Unigram     Bigram     Trigram

Functional  
 Basic  
 Advanced

			Term	f	tf-idf		
view	edit	<input type="checkbox"/>	state parti	4575	251.46	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	money launder	8820	202.7	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	member state	9980	165.82	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	organ crime	2299	157.66	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	financi institut	3626	156.26	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	organis crime	3332	140.51	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	request state	1524	139.8	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	te sat	623	136.92	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	terrorist financ	2456	136.24	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	oc group	796	133.26	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	framework decis	1602	124.63	<input type="checkbox"/>	remove nesting

Terms: 19224    context: 50    fragments: 2

Figura 4.10. Tabla de bigramas candidatos.

**FunGramKB Term Extractor** Log out

PRE-PROCESSING    PROCESSING (indexing)    PROCESSING (statistics)

VIEW    SEARCH    CORPUS

Corpus: GLOBALCRIMETERM    Terms

English

Unigram     Bigram     Trigram

© Functional  
 Basic  
 Advanced

			Term	f	tf-idf		
view	edit	<input type="checkbox"/>	anti money launder	1340	114.48	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	european arrest warrant	518	107.5	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	traffick in human	876	104.16	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	joint supervisor bodi	442	102.31	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	victim of traffick	505	99.68	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	mutual legal assist	954	98.75	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	prevent detent order	263	98.03	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	proceed of crime	771	87.6	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	transnat organ crime	442	85.44	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	financ of terror	713	84.24	<input type="checkbox"/>	remove nesting
view	edit	<input type="checkbox"/>	prosecutor s offic	277	79.76	<input type="checkbox"/>	remove nesting

Terms: 4340    context: 50    fragments: 2

Figura 4.11. Tabla de trigramas candidatos.

Previo al cálculo estadístico, el extractor aplica a los términos un filtro denominado *stemming* (“truncamiento”) que los reduce a sus raíces léxicas. Así, los “n-gramas” que visualizamos en las listas de “unigramas”, “bigramas” y “trigramas” son raíces léxicas, por lo que, para ver su forma completa, necesitaremos acceder a los textos mediante la pestaña “view” a la izquierda de cada “n-grama”. Por ejemplo, el “unigrama” candidato “smuggl” es una raíz que agrupa las unidades léxicas *smuggle* o *smuggling*. Gracias al *stemming* se consigue agrupar un mayor número de términos y un ajuste estadístico más preciso de los mismos (cf. Felices, Ureña y Alameda 2011: 77).

Los resultados preliminares obtenidos en el extractor de *FunGramKB* destacan su utilidad en la extracción terminológica. Tras subir al extractor el componente de textos en inglés del *GCTC*, que contiene más de cinco millones y medio de palabras (*tokens*), y tras aplicar los primeros filtros y el procesador estadístico, el número inicial se redujo aproximadamente a 5.700 términos candidatos, la cual es una cantidad comparativamente mucho menor de términos candidatos. Si comparamos el extractor con otras aproximaciones como el análisis manual de concordancias o colocaciones, cabe señalar que el número de términos candidatos se redujo en un periodo de tiempo bastante corto.

Tras el análisis estadístico, los términos candidatos aparecen clasificados de acuerdo a su carga semántica. Por tanto, los que aparecen más arriba de la lista son términos especializados estadísticamente más relevantes, mientras que los términos que presentan un índice *tf-idf* menor de 3 no son potencialmente especializados. Es importante señalar que el proceso de extracción terminológica en *FunGramKB* es semiautomático y que la decisión última de clasificación de un término en especializado o no recae en el criterio del terminólogo.

Una de las características más destacadas del extractor de *FunGramKB* es su gran potencial en el filtrado de falsos candidatos. El modo “View” contiene para cada candidato a término la opción de borrado simple (*remove*), por la que si el terminólogo elige esta opción un trigramma dado como *abolition of force* será automáticamente enviado a la papelera de reciclaje terminológica. Además, el extractor puede incluso realizar el borrado anidado de bigramas y trigramas (*nesting*). Por ejemplo, mediante el borrado anidado de *abolition of force* borramos tanto el trigramma *abolition of force* como cada uno de sus componentes individuales, esto es, *abolition*, *of* y *force*, y también borramos las combinaciones entre ellos *abolition of* y *of force* (cf. Ureña, Felices y Alameda 2011: 305). Es importante también destacar que es posible eliminar varios términos a la vez simplemente con seleccionar la casilla a la izquierda de *remove* de cada término candidato y al hacer clic sobre el mismo. En la figura 4.12 a continuación vemos un ejemplo de términos desechados en el proceso de filtrado.

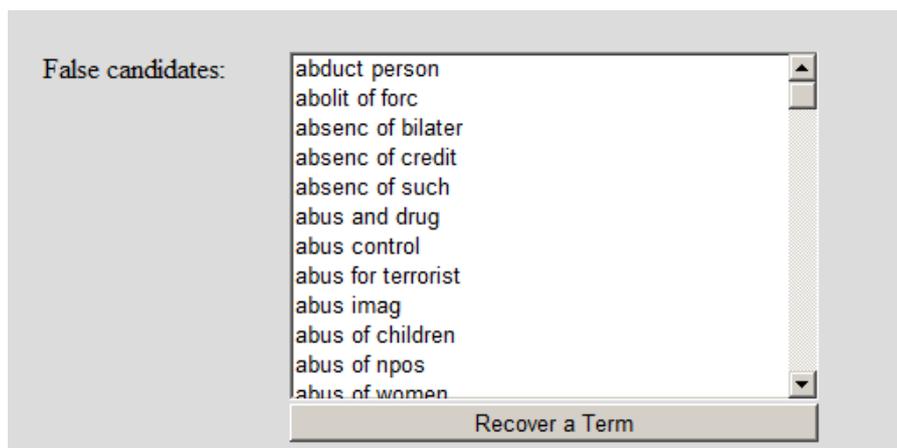


Figura 4.12. Papelera de reciclaje terminológica.

En la Tabla 4.1 ofrecemos el resumen algorítmico propuesto por Felices, Ureña y Alameda (2011: 79-80) de todos los casos que pueden ocurrir durante el filtrado terminológico, así como las decisiones que se han de adoptar en cada caso (el símbolo ‘>’ indica “aplíquese”, mientras que un asterisco ‘\*’, indica “término eliminado”, y ‘N/A’ indica que no es un caso posible en *FunGramKB*):

(i) Trigramas XYZ (léxico + léxico + léxico):

> *remove* = \*XYZ

> *nesting* = \*XYZ | \*X | \*Y | \*Z | \*XY | \*YZ

Ejemplo:

Trigramas “international global crime”

> *remove* = \*international global crime

> *nesting* = \*international global crime |

\*international | \*global | \*crime |

\*international global | \*global crime

(ii) Trigramas XYZ (léxico + funcional + léxico)

> *remove* = \*XYZ

> *nesting* = \*XYZ | \*X | \*Z

Ejemplo:

Trigramas “crimin or terrorist” (p. ej., “criminal or terrorist”)

> *remove* = \*crimin or terrorist

> *nesting* = \*crimin or terrorist | \*crimin | \*terrorista

(iii) Trigramas XYZ (léxico + léxico + funcional)

> *remove* = \*XYZ

> *nesting* = \*XYZ | \*X | \*Z

Ejemplo:

Trigramas “access to such”

> *remove* = \*access to such”

> *nesting* = \*access to such | \*access | \*such

(iv) Trigramas XYZ (funcional + léxico + léxico)

N/A

(v) Bigramas XY (léxico + léxico)

> *remove* = \*XY

> *nesting* = \*XY | \*X | \*Z

Ejemplo:

Bigramas “avoid transact” (p. ej., “avoid transaction”)

> *remove* = \*avoid transact

> *nesting* = \*avoid transact | \*avoid | \*transact

(vi) Bigramas XY (léxico + funcional)

N/A

(vii) Bigramas XY (funcional + funcional)

N/A

(viii) Unigramas X (léxico)

> *remove* = \*X

> repercusión: al borrar el unigrama X, X **no** desaparece de las formaciones bigramáticas o trigramáticas que contengan X.

Tabla 4.1. Algoritmo para el filtrado terminológico (resumen de Felices, Ureña y Alameda, 2011: 79-80).

Felices, Ureña y Alameda (2011: 80) también señalan que, para realizar un filtrado óptimo, es necesario comenzar de “arriba abajo”, es decir por el filtrado de los “trigramas” candidatos, para luego filtrar los “bigramas” y finalmente los “unigramas”. De este modo se optimiza el tiempo total empleado en el filtrado ya que, por las características del borrado anidado, se puede borrar simultáneamente el trigramas candidato y todas sus combinaciones, lo cual dinamiza sustancialmente el proceso. En

la etapa de filtrado de “unigramas” candidatos hay tres filtros posibles. El avanzado (*advanced*), el básico (*basic*) y el funcional (*functional*). Aquí es muy recomendable comenzar el filtrado con el filtro “avanzado” activado, ya que de este modo desaparecen miles de palabras de uso frecuente en los diccionarios no especializados. A continuación, el terminólogo debe continuar por seleccionar el filtro básico porque así desaparecen cientos de palabras no especializadas de uso muy frecuente. Finalmente, al aplicar el filtro funcional desaparecen de la lista de “unigramas” candidatos unidades funcionales como los pronombres, los verbos auxiliares o las preposiciones, que por sí solas no pueden ser término. Es importante señalar que la eliminación de “unigramas” en cualquiera de estos niveles no supone la eliminación de los “bigramas” y “trigramas” formados por éstos, es decir, el borrado anidado es posible para los “bigramas” y “trigramas”, pero no para los “unigramas”. También hay que destacar que los términos falsos quedan temporalmente almacenados en la papelera terminológica, a la cual se puede acceder a través de la pestaña “corpus” si es preciso consultar y/o recuperar alguno de ellos, hasta que se valide la lista definitiva de términos ganadores, momento en el que el extractor borrará automáticamente todos sus componentes.

Como veremos en la sección 4.3., se hace necesario establecer unos criterios para discernir qué términos candidatos son finalmente términos ganadores. Uno de estos criterios será el uso de diccionarios especializados, necesarios para distinguir las unidades léxicas (“unigramas”, “bigramas” y “trigramas”) que son términos especializados. Pero la labor del terminólogo se complica sustancialmente cuando trata de discriminar los “bigramas” y “trigramas” especializados de los que no lo son, ya que un gran número de ellos no aparecen en los diccionarios. En esos casos se hace necesaria e imprescindible la consulta del corpus de textos especializados subidos al extractor.

Para desechar un término candidato, además de analizar el índice de frecuencia *tf-idf* y de ver si el término se encuentra en los *Learner’s Dictionaries*, tendremos en cuenta los siguientes criterios (extracto adaptado de Carrión [en prensa]):

- a) Analizaremos si se trata de un término que no tiene significado especializado por sí solo, como es el caso de “rea” y “reus”, que adquieren significado en los “bigramas” *mens rea* [Latin: a guilty mind] y *actus reus* [Latin: a guilty act].

- b) En el caso de las siglas y acrónimos, estudiaremos si tienen proyección internacional o reconocimiento enciclopédico a través de su consulta en *Google*, *Wikipedia* y en los diccionarios y enciclopedias especializados. Ej: *OC* (*Organised Crime*), *WMD* (*Weapons of Mass Destruction*). Como denotan entidades y eventos las incluiremos en un archivo aparte ya que se integrarán en el Onomasticón y no en la ontología satélite.
- c) Para decidir si estamos ante un “bigrama” o “trigrama” propiamente dicho, nos aseguraremos de que su significado no sea una extensión de cada una de sus partes, como ocurre en *hostage taking*, sino que se trate de un término con un significado completamente nuevo, como es el caso de *carousel fraud*:

(1) *Hostage taking*: Taking hostages is today considered a crime or an act of terrorism; the use of the word in this sense of abductee became current only in the 1970s. The criminal activity is known as kidnapping. An acute situation where hostages are kept in a building or a vehicle that has been taken over by armed terrorists or common criminals is often called a hostage crisis.

**Hostage taking** is still often politically motivated or intended to raise a ransom or to enforce an exchange against other hostages or even condemned convicts. (<http://en.wikipedia.org/wiki/Hostage>).

(2) *Carousel fraud*: **Missing trader fraud** (also called **Missing Trader Intra-Community, MTIC**, or **carousel fraud**) is the theft of Value Added Tax (VAT) from a government by organised crime gangs who exploit the way VAT is treated within multi-jurisdictional trading where the movement of goods between jurisdictions is VAT-free. ([http://en.wikipedia.org/wiki/Carousel\\_fraud](http://en.wikipedia.org/wiki/Carousel_fraud)).

***El fraude carrusel es una amenaza particular para el sistema del IVA***

Un modelo de fraude que parece haber crecido durante los últimos años y que plantea ahora una preocupación importante para los Estados miembros en el comercio intracomunitario es el denominado **fraude carrusel**. Este mecanismo de fraude es un ejemplo claro de abuso de la exención del IVA

en el comercio intracomunitario, que podría muy bien aprovecharse de la próxima ampliación y extenderse a los nuevos Estados miembros. ([http://ec.europa.eu/translation/bulletins/puntoycoma/88/pyc884\\_es.htm](http://ec.europa.eu/translation/bulletins/puntoycoma/88/pyc884_es.htm)).

Como se trata de un término relativamente reciente *carousel fraud* no aparece aún en los diccionarios. Por otro lado, puede suceder que el término aparezca tanto en un *Learner's dictionary* como en los diccionarios especializados. Es por ejemplo el caso de *White-collar crime*:

- (3) *White-collar crime*: crime in which an office worker or someone in business illegally takes money from their employer or the people they deal with in their business. (*Cambridge Advanced Learner's Dictionary*).

*White-collar crime*. (1940) A nonviolent crime usu. involving cheating or dishonesty in commercial matters. Examples include fraud, embezzlement, bribery, and insider trading. (*Black's Law Dictionary*).

En este caso el uso especializado se ha generalizado y se ha hecho del uso común, por lo que, en principio, podríamos pensar que no deberíamos incluirlo en la ontología satélite del crimen organizado y el terrorismo, pero dada su estrecha conexión con el crimen organizado sí podría incluirse. En la fase de jerarquización de conceptos habría que determinar si hay fundamento o no para que esté sólo en la ontología nuclear o también en la ontología satélite como concepto espejo.

En el proceso de filtrado nos encontramos con casos especiales como *launder*, que, al consultar *Learner's Dictionaries* como el *Cambridge* o el *Longman*, vemos que es ampliamente significativo que ya lo asocian a *money*. Esto nos va a indicar que se trata de un término propiamente dicho, como corroboran los diccionarios especializados. En el caso del “bigrama” *money laundering*, se trata de un significado metafórico que el razonador de *FunGramKB* no podría inferir, ya que sólo puede unir el significado de dos conceptos por separado, como es el caso de *hostage taking* (mencionado más arriba), donde sí podría definir el “bigrama” a partir de *hostage* y *take*, que se trata de una mera colocación.

Una vez seleccionados los términos ganadores formalizaremos su conceptualización. Para realizar esa tarea, el extractor dispone de una herramienta de edición de términos (Figura 4.13.).

The screenshot shows a software interface for editing terms. At the top, there are three main sections: 'Senses:', 'CONCEPT:', and 'METACONCEPT:'. The 'Senses:' section is currently empty. The 'CONCEPT:' section contains a 'DESCRIPTION:' field and a 'Save' button. The 'METACONCEPT:' section has three radio buttons labeled 'Entities', 'Events', and 'Qualities'. Below these sections, there are several buttons: 'Rename', 'Delete', 'Done', 'Duplicate', and 'absorb'. The 'absorb' button is highlighted in blue. The bottom part of the interface is organized into a grid of language-specific sections: English, Spanish, Italian, French, German, Bulgarian, and Catalan. Each language section has a text input field and two buttons labeled 'Y' and 'N'.

Figura 4.13. Herramienta de edición de términos ganadores.

En esta sección hemos presentado las pautas seguidas en la compilación del GCTC. Hemos señalado que se establecieron una serie de criterios para la compilación de textos relevantes para nuestro corpus. A este respecto, los textos elegidos para la sección de la lengua inglesa se recogieron de una serie de organizaciones internacionales de reconocido prestigio cuyos documentos recogen términos especializados relevantes para nuestro corpus. También hemos destacado que el extractor es una herramienta de la suite de *FunGramKB* que asiste al terminólogo en una serie de tareas que incluyen desde el filtrado de falsos candidatos mediante el borrado simple o complejo, pasando por el filtrado de términos candidatos hasta llegar al cálculo de la carga semántica de los términos.

En la siguiente sección abordamos la conceptualización de los términos como etapa final de la adquisición terminológica, desde las definiciones en los diccionarios a su formalización en lenguaje natural y luego en COREL.

### 4.3. La conceptualización de los términos

Para la conceptualización de los términos de una ontología satélite de un dominio terminológico concreto, proponemos aquí seguir la metodología COHERENT, propuesta por Perrián y Mairal (2011), que sirvió de base metodológica del nivel conceptual básico de la ontología nuclear de *FunGramKB*.

Con los ejemplos mostrados en esta sección nos proponemos demostrar que la metodología COHERENT es válida en el modelado ontológico de cualquier ontología satélite. De acuerdo a esta metodología, podemos definir cualquier concepto de cualquiera de las lenguas implementadas en la base de conocimiento de *FunGramKB*<sup>70</sup> con un inventario cerrado de conceptos básicos. Así, mostraremos que se trata de una metodología productiva que sirve para enriquecer la ontología nuclear porque, como veremos a continuación, en el proceso de modelado de la ontología satélite ésta se nutrirá tanto de conceptos especializados, como de conceptos con conocimiento compartido ya existentes en la ontología nuclear (los denominados conceptos espejo).

A continuación mostramos la metodología a seguir en la construcción del nivel conceptual básico de cualquier ontología satélite partiendo de un corpus de textos para un dominio concreto. Para tal propósito utilizamos ejemplos del dominio del crimen organizado y el terrorismo en inglés que forman parte del proyecto de investigación en el que se enmarca este trabajo.

En una ontología satélite dada utilizaremos un corpus de textos especializados que, con ayuda del extractor de *FunGramKB* (presentado y analizado en la sección anterior), nos ayudará a determinar los términos ganadores candidatos a conceptos que conformarán la ontología satélite. Así, en la selección de términos ganadores no centraremos nuestra

---

<sup>70</sup> El inglés y el español están totalmente implementados en la versión actual de *FunGramKB*, si bien se ha empezado ya a trabajar con otras lenguas como el alemán, el italiano, el búlgaro y el catalán.

decisión en el índice de frecuencia ni en su uso, porque se trata de criterios estadísticos que están lejos de cumplir nuestro propósito, sino en las definiciones de los términos en los diccionarios especializados. En otras palabras, basaremos nuestra decisión en cómo están definidos los términos en los diccionarios especializados. Por tanto, utilizaremos el mencionado índice de frecuencia como una mera referencia que tendrá que ser contrastada con el uso de los términos en los textos especializados del corpus en cuestión, que nos ayudará no sólo a decidir si se trata de un término o no, sino también a elaborar su definición en lenguaje natural. La consulta del corpus de textos constituye un proceso realmente útil en el caso de los “bigramas” y “trigramas”, ya que en muchos casos no aparecen en los diccionarios especializados, como podemos ver en los siguientes ejemplos.

- (1) *Dirty money*. Criminal activity which is linked to organised crime also continues to be responsible for a large proportion of the *dirty money* flowing through financial channels. (GCTC).
- (2) *Self laundering*. As many criminals launder their own proceeds in these offences, the incrimination of *self laundering* in those countries involved in trafficking in human beings and illegal migration (where *self laundering* is not currently criminalised) should be actively considered. (GCTC).
- (3) *Bank secrecy laws*. Historically, *bank secrecy laws* have become notorious for providing cover for war operatives to funnel and secure money procured by them through looting the property of war prisoners. (GCTC).

En el caso de los “bigramas” y “trigramas” altamente especializados se ha probado realmente útil su consulta en los recursos *onlook*, *thefreedictionary*, y *Wikipedia*. Los dos primeros nos remiten a diccionarios tanto generales como especializados y *Wikipedia* aporta la contextualización del término.

### ***Dirty money***

- a. *thefreedictionary.com*: Goods or money obtained illegally.
- b. *Wiktionary* a través de *onlook.com*: (*idiomatic*) Money that is illegally gained, illegally transferred or illegally utilized. Especially money gained through forgery, bribery, or thievery.

Por otro lado, podemos encontrarnos con casos en los que el término aparece tanto en los *Learner's dictionaries* (*Longman Dictionary of Contemporary English*, *Cambridge Advanced Learner's Dictionary* y/o *Collins English Dictionary*, en adelante *LDCE*, *CALD* y *CED* respectivamente) como en los diccionarios especializados *Black's Law Dictionary*, *Oxford Dictionary of Law*, *Diccionario de términos jurídicos* de Alcaraz y Hughes y el *Diccionario bilingüe de terminología jurídica* de Bossini y Gleeson (en adelante *BLD*, *ODL*, *DTJ* y *DBTJ* respectivamente). Este es el caso de *money laundering*:

### **Money laundering**

- a. *LDCE*: when money that has been obtained illegally is put into legal businesses or bank accounts in different countries, so that it is difficult for people to discover where it came from.
- b. *CALD*: the crime of moving money that has been obtained illegally through banks and other businesses to make it seem as if the money has been obtained legally.
- c. *BLD*: n. (1974) The act of transferring illegally obtained money through legitimate people or accounts so that its original source cannot be traced. • Money-laundering is a federal crime. 18 USCA § 1956. It is also addressed by state governments, e.g. through the Uniform Money Services Act. Because some money-laundering is conducted across national borders, enforcement of money-laundering laws often requires international cooperation, fostered by organizations such as Interpol.
- d. *ODL*: Legitimizing money from organized or other crime by paying it through normal business channels. EU measures exist to control, on an EU-wide basis, the laundering of money, especially that resulting from organized crime.

Vemos que, como ocurría con *white-collar crime*, el uso especializado de *money laundering* se ha generalizado y se ha hecho del uso común, por lo que podríamos pensar que no deberíamos incluirlo en la ontología satélite del crimen organizado y el terrorismo, pero, dada su estrecha conexión con el crimen organizado, habría que incluirlo. En la fase de jerarquización de conceptos habrá que determinar si hay fundamento o no para que esté en la ontología nuclear, en la ontología satélite, o en ambas (concepto espejo). De este modo el concepto de la ontología nuclear expresaría conocimiento del sentido común y el mismo

concepto en la ontología satélite denotaría conocimiento experto, como ocurre por ejemplo con *crime*, *mastermind* y *syndicate*.

En el proceso de elaboración de definiciones en lenguaje natural, al consultar los diccionarios, encontramos términos que aparecen tanto en los *Learner's Dictionaries* como en los diccionarios especializados. Como el derecho es una disciplina social, consta de mucho conocimiento compartido (conocimiento del sentido común). Es difícil, por tanto, delimitar el conocimiento experto del que no lo es ya que en los diccionarios denominados *Learner's* hay términos que son ya de uso común, que incluso han llegado a fosilizarse, como es el caso de *crime* y *offence* que aportamos en la sección 4.3.1.

Como vamos a ejemplificar conceptos del crimen organizado y el terrorismo, los diccionarios especializados en terminología jurídica nos van a ayudar a elaborar las definiciones. Lo interesante es ver cómo se presentan dichas definiciones en los diccionarios especializados. De este modo una ontología satélite del crimen organizado y el terrorismo aportará conocimiento experto a través de sus conceptos, lo cual hará que la base de conocimiento sea más robusta.

En la ontología satélite modelamos conocimiento experto, por ello describimos el proceso cognitivo mediante el modelado de conceptos a través de los PS. Así evitamos la circularidad en las definiciones tan frecuente en los *Learner's Dictionaries* como en los diccionarios especializados. Como veremos a continuación, *offence* y *crime* son términos sinónimos según los diccionarios especializados. Tomamos como punto de partida los fenómenos del crimen organizado (*organized crime*) y el terrorismo (*terrorism*) para luego ejemplificar varios casos de terminología específica de las actividades delictivas asociadas a ellos.

#### **4.3.1. De los diccionarios a la elaboración de definiciones en lenguaje natural y COREL: Casos de estudio**

En este capítulo ofrecemos una descripción detallada del protocolo a seguir en el proceso de elaboración de definiciones. Los ejemplos que presentamos pertenecen al dominio del crimen organizado y el terrorismo, fruto de la labor de investigación realizada en el proyecto en el que se enmarca este trabajo. En el análisis de conceptos como casos de estudio, hemos seleccionado una muestra significativa que ejemplifican los resultados obtenidos a partir del GCTC.

De acuerdo con Perinián y Arcas (2004, 2005) los diccionarios sirven de guía lexicográfica para el ingeniero de conocimiento. Según uno de los compromisos ontológicos establecidos por Perinián y Arcas (2010a: 2669), el proceso de estructuración de la ontología se basa en un proceso de negociación de los ingenieros del conocimiento. Así se evitan subjetividades en la toma de decisiones en cuanto a los conceptos a incluir. Por ello, en la elaboración de definiciones de dichos conceptos, se hace imprescindible utilizar tanto diccionarios monolingües en inglés como bilingües en español.

Como partimos de una perspectiva conceptualista, el ingeniero del conocimiento ha de decidir si debe crear un concepto básico o terminal según el caso. En este punto cabe señalar que se crearán conceptos terminales sólo cuándo haya un *differentiae* bien definido que lo distinga significativamente de su inmediato concepto superordinado. Además, para que un concepto terminal sea productivo, tiene que tener al menos formalización en una de las lenguas objeto de estudio en el proyecto en el que se enmarca el presente trabajo y que están integradas en *FunGramKB* (inglés, español o italiano). Así, en el caso de existir laguna léxica de una lengua, ésta quedaría cubierta por otra(s). De este modo, una ontología satélite del crimen organizado y el terrorismo relacionaría palabras en inglés, español e italiano con el propósito de modelar el conocimiento de forma universal.

Una vez que el ingeniero de conocimiento decide qué conceptos básicos y qué terminales va a incluir en la taxonomía, el significado de cada una de las unidades léxicas del dominio cognitivo se pasará a lenguaje de notación COREL para así conformar las unidades conceptuales. En ese proceso, mediante PS, el ingeniero del

conocimiento elaborará las definiciones de los conceptos terminales con un conjunto limitado de conceptos básicos.

Partimos de los llamados *Learner's Dictionaries* (*LDCE*, *CALD* y *CED*) y contrastamos su definiciones con las de los diccionarios especializados (*BLD*, *ODL*, *DTJ* y *DBTJ*) para elaborar las definiciones en lenguaje natural que luego pasaremos a COREL. En los casos de estudio analizamos si se trata de un “concepto espejo” (el mismo concepto aparece tanto en la ontología nuclear como en la satélite, pero con PS distintos. Mientras en la ontología nuclear el concepto refleja el conocimiento compartido, en la satélite representa el conocimiento experto). En dicho análisis utilizamos minúsculas en cursiva para representar las unidades léxicas y mayúsculas para representar los conceptos, estos últimos precedidos de [+] o [ \$ ], ya sea un concepto básico o terminal, y seguidos del índice **\_00**, **\_01**, etc. Al lado del concepto aparece la definición en lenguaje natural y a continuación su representación en COREL a través de su PS y una lista de las unidades léxicas asociadas al concepto.

#### **4.3.1.1. Conceptos espejo**

Partimos del análisis de *crime* y *offence* en los diccionarios tanto de uso general como especializado por tratarse de unidades léxicas claves en el fenómeno de la criminalidad organizada. Al contrastar la definición de *crime* en los *Learner's Dictionaries* con los diccionarios especializados vemos que *crime* sería un “concepto espejo”. El mismo concepto aparecerá reflejado tanto en la ontología nuclear como en la satélite, si bien habrá una clara distinción entre ambas. El conocimiento del sentido común (compartido con la ontología satélite) está reflejado en la ontología nuclear y el conocimiento experto en la satélite.

##### Conceptualización en la ontología nuclear:

+CRIME\_00: an act punishable by law; usually considered an evil act.

+(e1: +BE\_00 (x1: +CRIME\_00)Theme (x2: +OCCURRENCE\_00)Referent)

+(e2: n +BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)

Unidades léxicas asociadas: *crime*, *law-breaking*, *delito*, *ilegalidad*, *infracción*, *quebrantamiento*, *quebranto*, *transgresión*, *vulneración*.

Conceptualización en la ontología satélite:

+CRIME\_00: a violation of the law punishable by the state in criminal proceedings.

+(e1: +BE\_00 (x1: +CRIME\_00)Theme (x2: +OCCURRENCE\_00)Referent)

+(e2: n +BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)

+(e3: +BE\_00 (x1)Theme (x4: +FELONY\_00 ^ +MISDEMEANOR\_00)Referent)

\*(e4: +PUNISH\_00 (x5: +COUNTRY\_00)Theme (x6: +CRIMINAL\_00)Referent)

\*(e5: +BE\_02 (x6)Theme (x7: +CRIMINAL\_PROCEEDING\_00)Location

(f1: +IN\_00)Position)

Unidades léxicas asociadas: *crime, criminal wrong, offence, delito, ofensa.*

Al comparar las definiciones de *crime* y *offence* vemos que, según el *ODL*, hoy en día se pueden llegar a utilizar indistintamente, si bien el *BLD* nos dice que *offence* es un *minor crime*. En la conceptualización de estos dos términos tendríamos que decidir si entonces *offence* sería simplemente una unidad léxica asociada a la entidad +CRIME\_00 o si sería el concepto terminal \$OFFENCE\_00, en el que especificaríamos que *offence* en un delito menor.

**Crime**

- a. *LDCE*: 1. Illegal activities in general. 2. An illegal action, which can be punished by law.
- b. *CALD*: 1. Illegal activities. 2. An illegal act.
- c. *BLD*: An act that the law makes punishable; the breach of a legal duty treated as the subject-matter of a criminal proceeding. - Also termed criminal wrong. See OFFENSE (1).
- d. *ODL*: 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.
- e. *DTJ*: delito.

f. *DBTJ*: crimen, delito, violación de la ley penal.

### ***Offence***

a. *LDCE*: An illegal action or a crime.

b. *CALD*: 2. A crime.

c. *BLD*: 1. A violation of the law; a crime, often a minor one. See CRIME. - Also termed criminal offense. [Cases: Criminal Law.]

"The terms 'crime,' 'offense,' and 'criminal offense' are all said to be synonymous, and ordinarily used interchangeably. 'Offense' may comprehend every crime and misdemeanor, or may be used in a specific sense as synonymous with 'felony' or with 'misdemeanor,' as the case may be, or as signifying a crime of lesser grade, or an act not indictable, but punishable summarily or by the forfeiture of a penalty." 22 CJ,S. Criminal Law § 3, at 4 (1989).

d. *ODL*: A \*crime. The modern tendency is to refer to crimes as offences. Offences may be classified as \*indictable or \*summary and as \*arrestable or nonarrestable.

e. *DTJ*: delito, violación, acto punible, ofensa.

f. *DBTJ*: infracción, delito, acto punible, ilícito, ofensa, violación.

Tras analizar en detalle las definiciones de *crime* y *offence* vemos que *offence* no tendría un PS distinto al de *crime* por lo que hemos optado por aglutinar *offence* como palabra asociada al concepto básico +CRIME\_00 ya que la especificidad de *offence* puede marcarse a través de su correspondiente lexicón.

Si comparamos las definiciones de los diccionarios especializados con las definiciones en la ontología satélite, podemos plantearnos el nivel de granularidad que debemos aportar en las mismas (cf. Perrián y Mairal, 2010b: 24), es decir ¿cuánto conocimiento semántico debemos incluir en el PS de los conceptos especializados? Los diccionarios jurídicos ofrecen una gran cantidad de información en sus definiciones, de carácter enciclopédico en la mayoría de los casos, por lo que, si comparamos las definiciones de la ontología satélite con las de las fuentes lexicográficas, el nivel de granularidad de la

ontología satélite será más grueso, pero lo suficientemente fino para ofrecer el conocimiento experto. De todos modos, la cantidad de conocimiento experto a incluir en las definiciones será fruto del consenso obtenido entre los ingenieros del conocimiento y los expertos del dominio en cuestión. Por otra parte, el conocimiento enciclopédico aparece recogido en el Onomasticón y el conocimiento procedimental en el Cognición, que junto al conocimiento semántico de la ontología componen el módulo conceptual de *FunGramKB*.

A continuación ofrecemos una muestra significativa de términos candidatos a conceptos. Comenzamos clasificando el fenómeno del crimen organizado como actividad criminal colectiva y después clasificamos algunas de las actividades criminales llevadas a cabo por las bandas organizadas. De acuerdo a Sánchez (2008: 47-57), las siguientes son actividades típicas del crimen organizado: tráfico de drogas y estupefacientes, tráfico y explotación laboral y sexual de seres humanos, tráfico ilegal de armas, falsificaciones, robo y tráfico ilegal de vehículos, blanqueo de dinero, estafa, fraude fiscal y de subvenciones, corrupción, contrabando y criminalidad informática. Siguiendo la clasificación de Sánchez, ofrecemos un análisis detallado de la definición de varias de estas actividades del crimen organizado.

En el análisis de *organized crime* en los *Learner's Dictionaries* vemos que utilizan el mismo bigrama para referirse tanto a la organización criminal (“banda organizada”) como a sus actividades delictivas. Sin embargo, el diccionario especializado *BLD* si deja clara la distinción.

### ***Organized crime***

- a. *CALD*: criminal organizations which plan and commit crime, or the crimes which are committed by such organizations.
- b. *Wiktionary*: 1. A set of large criminal organizations (often competing for markets and territories) that deal in illegal goods and services. 2. The crimes perpetrated by those organizations.
- c. *BLD*: Widespread criminal activities that are coordinated and controlled through a central syndicate. See RACKETEERING. 2. Persons involved in these

criminal activities; a syndicate of criminals who rely on their unlawful activities for income. See SYNDICATE.

d. *DBTJ*: crimen organizado.

Nos encontramos, por tanto, ante un nuevo caso de “concepto espejo”, es decir, el mismo concepto aparece tanto en la ontología nuclear como en la satélite. Así, el concepto básico +ORGANIZED\_CRIME\_00 expresa el conocimiento del sentido común en la ontología nuclear (“organización criminal y sus actividades delictivas”) y los conceptos básicos +ORGANIZED\_CRIME\_00 y +SYNDICATE\_00 denotan conocimiento experto en la ontología satélite (“actividades delictivas propias de las bandas organizadas” y “banda organizada (sindicato del crimen)” respectivamente). +ORGANIZED\_CRIME\_00 a su vez desempeña la función de concepto paraguas en la ontología satélite ya que nos sirve para definir las actividades típicas del crimen organizado (*arms trafficking, corruption, counterfeiting, drug trafficking...*).

Además, en el análisis de *gang, mafia, mob* y *ring* como “banda organizada”, en los *Learner’s Dictionaries* vemos que en sus definiciones no contienen características diferenciadoras suficientes como para crear un concepto nuevo, por lo que las podemos aglutinar como unidades léxicas bajo el mismo concepto +ORGANIZED\_CRIME\_00. Dado que los lexicones de *FunGramKB* se “pueblan” a partir de los conceptos a los que están asociados, lo que significa que “la construcción de entradas léxicas implica el modelado ontológico previo de sus correspondientes conceptos” (Mairal y Periñán 2009a: 221), la tarea del lexicógrafo computacional consiste en especificar la información morfosintáctica, pragmática y de colocaciones de aquellas palabras aglutinadas como unidades léxicas en la ontología satélite. Por tanto, aunque *criminal organisation, gang, mafia, mob, organised criminal group, organised crime, ring, banda criminal* y *crimen organizado* dependen del mismo concepto sus peculiaridades se especificarán en sus correspondientes lexicones.

### ***Gang***

- a. *LDCE*: a group of criminals who work together.
- b. *CALD*: a group of criminals who work together.

- c. *BLD*: A group of persons who go about together or act in concert, esp. for antisocial or criminal purposes. - Many gangs have common identifying signs and symbols, such as hand Signals and distinctive colors. Also termed street gang.
- d. *DTJ*: **gang of crooks** (PENAL banda criminal).
- e. *DBTJ*: Pandilla, banda.

### ***Mafia***

- a. *LDCE*: **the Mafia** a large organized group of criminals who control many illegal activities, especially in Italy and the US.
- b. *CALD*: a large group of organized criminals.
- c. *DBTJ*: mafia.

### ***Mob***

- a. *LDCE*: **informal** a group of people of the same type [= gang]. The Mob the MAFIA (=a powerful organization of criminals).
- b. *CALD*: a large group of people that is often violent or not organized.
- c. *DTJ*: en los Estados Unidos *The Mob* es la <<mafia>>.
- d. *DBTJ*: Tumulto, banda, masa, populacho.

### ***Ring***

- a. *LDCE*: a group of people who illegally control a business or criminal activity.
- b. *CALD*: a group of people who are involved in an illegal activity together.
- c. *DTJ*: camarilla, banda, pandilla, red; cártel, sindicato.

Según esto, la conceptualización de +ORGANIZED\_CRIME\_00 en la ontología nuclear es la siguiente:

+ORGANIZED\_CRIME\_00: criminal organization or the crimes committed by such organization.

+(e1: +BE\_00 (x1: +ORGANIZED\_CRIME\_00)Theme (x2: +CRIME\_00)Referent)

+(e2: +BE\_00 (x1)Theme (x3: +ORGANIZATION\_00)Referent)

+(e3: n +BE\_01 (x3) Theme (x4: +LEGAL\_00)Attribute)

Unidades léxicas asociadas: *criminal organization, gang, mafia, mob, organized crime, organized criminal group, ring, banda criminal, crimen organizado.*

Como hemos señalado anteriormente, el conocimiento experto se formaliza en la ontología satélite. Mientras que en la ontología nuclear aparecen los dos significados (“banda criminal” y “actividades delictivas de las bandas criminales”) bajo el mismo concepto (+ORGANIZED\_CRIME\_00) en la ontología satélite distinguimos dos “conceptos espejo”, por un lado el concepto +ORGANIZED\_CRIME\_00 y por otro +SYNDICATE\_00.

Conceptualización en la ontología satélite:

+ORGANIZED\_CRIME\_00: criminal activities coordinated and controlled through a central syndicate.

+(e1: +BE\_00 (x1: +ORGANIZED\_CRIME\_00)Theme (x2: +CRIME\_00)Referent)

+(e2: +CONTROL\_00 (x3: +SYNDICATE\_00)Theme (x1)Referent)

Unidades léxicas asociadas: *organized crime, crimen organizado.*

Como veíamos en la primera acepción de la definición de *organized crime* en el *BLD*, *racketeering* es una actividad propia de las bandas organizadas, que, al consultar los *Learner’s Dictionaries* vemos que presenta conocimiento del sentido común frente al conocimiento especializado del *BLD*. Por tanto, como mostramos a continuación, se trata también de un “concepto espejo”.

### ***Racketeering***

- a. *LDCE*: when someone earns money through crime and illegal activities.
- b. *CALD*: making money from a dishonest or illegal activity.
- c. *BLD*: 1. A system of organized crime traditionally involving the extortion of money from businesses by intimidation, violence, or other illegal methods.  
2. A pattern of illegal activity (such as bribery, extortion, fraud, and murder) carried out as part of an enterprise (such as a crime syndicate) that is owned or controlled by those engaged in the illegal activity. • The modern sense (sense 2) derives from the federal RICO statute, which greatly broadened the term's original sense to include such activities as mail fraud, securities fraud, and the collection of illegal gambling debts.
- d. *DTJ*: bandolerismo, extorsión de chantaje e intimidación
- e. *DBTJ*: extorsión, chantaje, intimidación o amenazas para obtener ingresos, latrocinio.

La conceptualización en la ontología nuclear es:

\$RACKETEERING\_00: making money from a dishonest or illegal activity.

+(e1: +BE\_00 (x1: \$RACKETEERING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +DECEIVE\_00 (x3) Theme (x4) Referent (f1: (e3 + OBTAIN\_00 (x3) Theme (x5: +MONEY\_00)Purpose)

Unidades léxicas asociadas: *extortion, racketeering, extorsión*.

Y la conceptualización en la ontología satélite:

\$RACKETEERING\_00: an organized crime activity that involves extortion of money from businesses by intimidation or violence.

+(e1: +BE\_00 (x1: \$RACKETEERING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +DECEIVE\_00 (x3: +SYNDICATE\_00) Theme (x4: +BUSINESS\_00)  
Referent (f1: (e3 + OBTAIN\_00 (x3) Theme (x5: +MONEY\_00)Purpose)

+(e4: +USE\_00 (x3)Theme (x6: +VIOLENCE\_00)Referent (f1: x6)Instrument)

Unidades léxicas asociadas: *extortion, racketeering, extorsión, latrocinio.*

En cuanto a *syndicate*, al examinar su caso, vemos que los diccionarios generales no distinguen el significado especializado del término, que sí aparece en la segunda acepción del *BLD*. Nos encontramos, por tanto, ante otro caso de “concepto espejo”. El mismo concepto aparece recogido en las dos ontologías, donde el conocimiento compartido (“sindicato”) aparece reflejado en el concepto terminal de la ontología nuclear \$SYNDICATE\_00 y el significado experto (“sindicato del crimen”) en el concepto básico de la satélite +SYNDICATE\_00. Este es un caso de promoción del llamado modelo en espiral, que permite la promoción de conceptos terminales a básicos y la degradación de los conceptos básicos a terminales (cf. Perrián Pascual y Arcas Túnez, 2010b). Dada la alta productividad de su significado especializado, el concepto terminal de la ontología nuclear \$SYNDICATE\_00 pasa a ser el concepto básico +SYNDICATE\_00 en la ontología satélite del crimen organizado y terrorismo, por ser necesario para la formalización de los PS de, entre otros, +ORGANISED\_CRIME\_00, \$RACKETEERING\_00 y \$SMUGGLING\_00.

### ***Syndicate***

- a. *LDCE*: a group of people or companies who join together in order to achieve a particular aim.
- b. *CALD*: a group of people or companies who join together in order to achieve something.
- c. *BLD*: A group organized for a common purpose; esp., an association formed to promote a common interest, carry out a particular business transaction, or (in a negative sense) organize criminal enterprises. See ORGANIZED CRIME. - syndicate (sin-di-kayt), vb. - syndication (sin-di-kay-shan), n. - syndicator (sin-di-kay-t<lr), n.
- d. *DTJ*: *US argot* (PENAL banda criminal; v. *ring*).

e. *DBTJ: crime* Banda criminal, grupo de criminales organizados.

Así, la ontología nuclear recoge su significado más general con el siguiente PS:

\$SYNDICATE\_00: a group of people or companies who join together in order to achieve a particular aim.

\*((e1: +BE\_00 (x1: \$SYNDICATE\_00)Theme (x2: +GROUP\_00)Referent)(e2: +COMPRISE\_00 (x2)Theme (x3: s +COMPANY\_00)Referent))

Unidades léxicas asociadas: *corporation, syndicate, corporación, sindicato*.

Por su parte, el “concepto espejo” \$SYNDICATE\_00 recogerá el conocimiento experto en la ontología satélite:

\$SYNDICATE\_00: a group that organizes criminal activities.

+(((e1: +BE\_00 (x1: +SYNDICATE\_00)Theme (x2: +GROUP\_00)Referent)(e2: +COMPRISE\_00 (x2)Theme (x3: s +CRIMINAL\_00)Referent) (e3: +ORGANIZE\_00 (x1)Theme (x4: +CRIME)Referent)))

Unidades léxicas asociadas: *criminal organization, organized criminal group, banda criminal, sindicato (del crimen)*.

A continuación presentamos varios ejemplos de “conceptos espejo” relacionados con el terrorismo. El primer caso es el de *agression* que, como vemos más abajo, en su sentido general denota el delito de agresiones (conocimiento del sentido común que formalizamos en la ontología nuclear) y en su sentido experto expresa el ataque armado que un país ejerce sobre otro sin haber sido atacado previamente (conocimiento experto que conceptualizamos en la ontología satélite). Como paso previo a la conceptualización, y como hemos hecho anteriormente, ofrecemos las entradas de los diccionarios, tanto generales como especializados.

### ***Aggression***

- a. *LDCE*: 1. angry or threatening behaviour or feelings that often result in fighting. 2. the act of attacking a country, especially when that country has not attacked first.
- b. *CALD*: spoken or physical behaviour which is threatening or involves harm to someone or something.
- c. *BLD*: Int'l law. A grave breach of international law by a nation. • The prohibition of aggression is a peremptory rule (*jus cogens*). Aggressors are guilty of an international crime. But there is no generally accepted definition of what constitutes aggression despite many attempts over the years to devise one.
- d. *OLD*: (in international law) According to the General Assembly Resolution (3314) on the Definition of Aggression 1975, the use of armed force by one state against the sovereignty, territorial integrity, or political independence of another state or in any way inconsistent with the Charter of the United Nations. The Resolution lists examples of aggression, which include the following. (1) Invasion, attack, military occupation, or annexation of the territory of any state by the armed forces of another state. (2) Bombardment or the use of any weapons by a state against another state's territory. (3) Armed blockade by a state of another state's ports or coasts. (4) The use of a state's armed forces in another state in breach of the terms of the agreement on which they were allowed into that state. (5) Allowing one's territory to be placed at the disposal of another state, to be used by that state for committing an act of aggression against a third state. (6) Sending armed bands or guerrillas to carry out armed raids on another state that are grave enough to amount to any of the above acts.
- d. *DTJ*: ataque, asalto, agresión, acometimiento.
- e. *DBTJ*: agresión.

La conceptualización en la ontología nuclear es:

+AGGRESSION\_00: spoken or physical behaviour which is threatening or involves harm to someone or something.

+(e1: +BE\_00 (x1: +AGGRESSION\_00)Theme (x2: +CRIME\_00)Referent)

+(e1: +ATTACK\_00 (x3)Theme (x4)Referent (f1: +VOICE\_00 ^  
+HAND\_00)Instrument)

(f1:(e2: +DO\_00 (x5: +HARM\_00)Theme (x4)Referent))Purpose)

Unidades léxicas asociadas: *aggression, agresión.*

Y la conceptualización en la ontología satélite:

+AGGRESSION\_00: the breach of international law by a country.

+(e1: +BE\_00 (x1: +AGGRESSION\_00)Theme (x2: +CRIME\_00)Referent)

+(e1: +ATTACK\_00 (x3: +COUNTRY\_00)Theme (x4:  
+COUNTRY\_00)Referent (f1: +ARMY\_00)Instrument)

Unidades léxicas asociadas: *aggression, ataque, agresión.*

En segundo lugar, la consulta de *safe haven* en los diccionarios de uso general nos lo presenta como “un refugio para alguien en peligro”, mientras que en su sentido especializado puede significar, o “un refugio con protección militar para las minorías étnicas en peligro” (*OLD*); o “un refugio para criminales o terroristas” (*GCTC*).

### ***Safe haven***

- a. *LDCE*: A place where someone can go in order to escape from possible danger or attack.
- b. *CALD*: A place where someone is safe from danger.
- c. *OLD*: A zone of territory within a sovereign state demarcated by the United Nations or other international organization) as a refuge to which a persecuted \*ethnic minority can choose to retire. While within such a zone the ethnic minority is afforded military protection by the body that established the

zone. The international community set up safe havens in Iraq and the former Yugoslavia in response to acts of systematic persecution carried out by the government of the sovereign state concerned against part of its own population. See also HUMANITARIAN INTERVENTION.

- d. *GCTC*: there must be no **safe havens** for those who organize, aid or abet, profit from, or participate in the smuggling of or trafficking in human beings. Jurisdictions become **safe havens** for persons and funds related to corrupt activities unless there is a legal framework and mechanisms to detect, investigate and prosecute corruption offences and related money laundering, and to recover stolen assets.
- e. *GCTC* Countries should also take all possible measures to ensure that they do not provide **safe havens** for individuals charged with the financing of terrorism, terrorist acts or terrorist organizations (...).The fundamental role of overt warfare in counterterrorism is to eliminate the regimes that provide **safe haven** to terrorist groups.
- f. *Collins translator*: refugio.

Así, la conceptualización de \$SAFE\_HAVEN\_00 en la ontología nuclear es la siguiente:

\$SAFE\_HAVEN\_00: a refuge for people in danger.

+(e1: +BE\_00 (x1: \$SAFE\_HAVEN\_00)Theme (x2: +PLACE\_00)Referent)

+(e2: +BE\_02 (x1)Theme (x3: +REFUGE\_00)Location)  
(f1:+HUMAN\_00)Beneficiary)

Unidades léxicas asociadas: *safe haven, refugio*.

Mientras que en la ontología satélite su formalización es:

\$SAFE\_HAVEN\_00: a refuge with military protection for a persecuted ethnic minority; a refuge for criminals or terrorists.

+(e1: +BE\_00 (x1: \$SAFE\_HAVEN\_00)Theme (x2: +PLACE\_00)Referent)

\*((e2: +BE\_02 (x1)Theme (x3: +REFUGE\_00)Location)  
(e3:+ETHNIC\_MINORITY\_00)Beneficiary)(f1: +ARMY\_00)Instrument))

\*((e4: +BE\_02 (x1)Theme (x3: +REFUGE\_00)Location) (f2:+CRIMINAL\_00  
^+TERRORIST\_00)Beneficiary))

Unidades léxicas asociadas: *safe haven, refugio*.

El caso de *terrorist* es bastante significativo porque se trata de un concepto que se ha fosilizado, como bien muestran los diccionarios de uso general. Al consultar los diccionarios especializados vemos que no ofrecen una entrada propia, por lo que hemos tenido que recurrir al *GCTC* para poder formalizar el conocimiento experto del concepto.

### ***Terrorist***

- a. *LDCE*: someone who uses violence such as bombing, shooting etc to obtain political demands.
- b. *CALD*: someone who is involved in terrorism.
- c. *Thefreedictionary (Thesaurus)*: **terrorist** - a radical who employs terror as a political weapon; usually organizes with other terrorists in small cells; often uses religion as a cover for terrorist activities.
- d. *GCTC*: b) The term **terrorist** refers to any natural person who: (i) commits, or attempts to commit, terrorist act by any means, directly or indirectly, unlawfully and wilfully; (ii) participates as an accomplice in terrorist act ; (iii) organises or directs others to commit terrorist act ; or (iv) contributes to the commission of terrorist act by a group of persons acting with a common purpose where the contribution is made intentionally and with the aim of furthering the terrorist act or with the knowledge of the intention of the group to commit a terrorist act .
- e. *DTJ*: terrorista.

Así, la conceptualización en la ontología nuclear es la siguiente:

+TERRORIST\_00: someone who uses terror as a political weapon.

+(e1: +BE\_00 (x1: +TERRORIST\_00)Theme (x2: +HUMAN\_00) Referent)

+(e2: +USE\_00 (x1)Theme (x3: +TERROR\_00)Referent (f1: +WEAPON\_00)Instrument)

+(e3: +BE\_01 (x3)Theme (x4: +POLITICAL\_00)Attribute)

Unidades léxicas asociadas: *terrorist, terrorista*.

Por otro lado, la conceptualización en la ontología satélite sería:

+TERRORIST\_00: someone who commits a terrorist act; participates as an accomplice in terrorist act; or organizes others to commit a terrorist act.

+(e1: +BE\_00 (x1: +TERRORIST\_00)Theme (x2: +HUMAN\_00) Referent)

+(e2: +USE\_00 (x1)Theme (x3: +TERROR\_00)Referent (f1: +WEAPON\_00)Instrument)

+(e3: +BE\_01 (x3)Theme (x4: +POLITICAL\_00)Attribute)

+(e4: +ORGANIZE\_00 ^ +PARTICIPATE\_00 (x1)Theme (x5: +TERRORIST\_CELL ^ +TERRORIST\_ACT\_00)Referent)

Unidades léxicas asociadas: *terrorist, terrorista*.

Como último ejemplo de “concepto espejo” mostramos el caso de *torture*, que en su sentido general (conocimiento del sentido común) denota “el acto de infligir dolor extremo para obtener información”; mientras que el conocimiento especializado radica en que “el dolor extremo puede ser causado en el cuerpo o la mente para conseguir información o una confesión, castigar a la víctima u obtener placer sádico”.

### ***Torture***

- a. *LDCE*: n. an act of deliberately hurting someone in order to force them to tell you something, to punish them, or to be cruel. V. to deliberately hurt someone in order to force them to give you information, to punish them, or to be cruel.

- b. *CALD*: n. to cause someone severe pain, often in order to make them tell you something. V. to cause someone severe pain, often in order to make them tell you something.
- c. *BLD*: n. The infliction of intense pain to the body or mind to punish, to extract a confession or information, or to obtain sadistic pleasure. - *torture*, vb.
- d. *OLD*: n. Under section 134 of the Criminal Justice Act 1988, the offence committed by a public official (or someone with the official's acquiescence) of intentionally inflicting severe physical or mental suffering on any person anywhere in the world.
- e. *DBTJ*: *tortura*.

Por un lado, la conceptualización en la ontología nuclear es la siguiente:

+TORTURE\_00: the cause of severe pain to get information.  
+(e1: +BE\_00 (x1: +TORTURE\_01)Theme (x2: +CRIME\_00)Referent)  
+(e2: +DO\_00 (x3)Theme (x4: m+PAIN\_00)Referent)  
(f1: (e3: +OBTAIN\_00 (x3)Theme (x5: +INFORMATION\_00)Referent))Purpose)

Unidades léxicas asociadas: *torture*, *tortura*.

Y, por otro, la conceptualización en la ontología satélite es:

+TORTURE\_00: the cause of severe pain to the body or mind to punish, get a confession or information or to obtain sadistic pleasure (to be cruel).  
+(e1: +BE\_00 (x1: +TORTURE\_01)Theme (x2: +CRIME\_00)Referent)  
+(e2: +DO\_00 (x3: +HUMAN\_00)Theme (x4: m+PAIN\_00)Referent)  
+(e3: +BE\_02 (x4)Theme (x5: +BODY\_AREA\_00 ^ +MIND\_00)Location  
(f1: +IN\_00)Position)  
(f1: (e4: +PUNISH\_00 (x3)Theme (x6: +VICTIM\_00)Referent)Purpose)

(f2: (e5: +OBTAIN\_00 (x3)Theme (x7: +CONFESSION ^  
+INFORMATION\_00)Referent))Purpose)

(f3: (e6: +BE\_01 (x3)Theme (x8: +CRUEL\_00)Attribute)Purpose))

Unidades léxicas asociadas: *torture, tortura*.

En el análisis de los “conceptos espejo” presentados en esta sección hemos visto casos de “conceptos espejo claros” (donde el conocimiento especializado está bastante claro), como por ejemplo +AGGRESSION\_00 y +SYNDICATE\_00, y “conceptos espejo oscuros” (donde el conocimiento especializado no está suficientemente claro, por lo que se hace necesaria la consulta al GCTC complementaria a la consulta de los diccionarios especializados), como por ejemplo + ORGANIZED\_CRIME\_00 y +TERRORIST\_00.

Como hemos señalado al comienzo de la sección 4.3.1, aquí hemos presentado una muestra de varios “conceptos espejo”, extracto del apéndice 1, donde ofrecemos los PS de 20 “conceptos espejo” de nuestra ontología satélite del crimen organizado y el terrorismo. En dicho apéndice se pueden consultar más casos de “conceptos espejo claros” (p. ej. +FREEZE\_00, +SEIZE....) y de “conceptos espejo oscuros” (+BLACK\_MARKET\_00, \$PHISHING\_00 , +TAX\_HAVEN\_00,...).

#### 4.3.1.2. Conceptos especializados

Los conceptos especializados son más fáciles de detectar que los “conceptos espejo” porque, mientras que en estos últimos confluye el conocimiento del sentido común con el conocimiento experto, en los primeros el conocimiento que expresan es inminentemente experto, es decir, reflejan conocimiento prototípico del dominio en cuestión, si bien en el ámbito legal confluye una gran cantidad de conocimiento compartido.

Un claro ejemplo de esto lo encontramos en *forgery* porque, mientras que los *Learner's Dictionaries* mezclan los significados de *counterfeiting* y *forgery* (“the crime of copying official documents, money etc”) en la misma unidad léxica (*forgery*), los diccionarios especializados sí que hacen la distinción entre *forgery* (“the act of fraudulently making a false document or altering a real one to be used as if genuine”) y *counterfeiting*

(“unlawful making of false money”). Por eso hemos optado por considerar a +FORGERY\_00 “concepto espejo” (cf. Apéndice 1) y a \$COUNTERFEITING\_00 concepto especializado.

Para ver clara esta distinción ofrecemos más abajo la entrada en las fuentes lexicográficas tanto de *counterfeit* y *forge* como de *counterfeiting* y *forgery*.

### ***Counterfeit***

- a. *LDCE*: To copy something exactly in order to deceive people [= fake].
- b. *CALD*: To make something to look like the original, usually for dishonest or illegal purposes.
- c. *Onelook*: To fraudulently make money, a document, art, or other item which is forged or created to look real, and intended to pass for real.
- d. *Thefreedictionary*: To criminally forge or print a false copy of money, bonds, or other valuable documents, intending to profit from the falsity.
- e. *BLD*: "Literally a counterfeit is an imitation intended to pass for an original. Hence it is spurious or false, and to counterfeit is to make false. For this reason the verbs counterfeit and forge are often employed as synonyms and the same is true to some extent of the corresponding nouns. No error is involved in this usage but it is important to distinguish between the words as far as possible when used as the labels of criminal offenses, In the most restricted sense, [c]ounterfeiting is the unlawful making of false money in the similitude of the genuine. At one time under English statutes it was made treason. Under modern statutes it is a felony.
- f. *DTJ*: *falsificar*.

### ***Forge***

- a. *LDCE*: to illegally copy something, especially something printed or written, to make people think that it is real [↔ [counterfeit](#)].
- b. *CALD*: to make an illegal copy of something in order to deceive people.

- c. *DTJ*: falsificar, falsear.

Si buscamos *counterfeiting* y *forgery* en los *Learner's Dictionaries* no encontramos la entrada del primero, aunque sí la del segundo, lo cual nos indica que *counterfeiting* se trata de un término propiamente dicho.

### ***Counterfeiting***

- a. *DBTJ*: falsificación, cuando se copia o imita algo para que crean que es el original. Este engaño es una estafa, ya que se hace con ánimo de defraudar a la gente, al aceptar como original y genuino lo que realmente es una imitación.

### ***Forgery***

- a. *LDCE*: the crime of copying official documents, money etc .
- b. *CALD*: the crime of making an illegal copy of something.
- c. *BLD*: the act of fraudulently making a false document or altering a real one to be used as if genuine <the contract was void because of the seller's forgery>. • Though forgery was a misdemeanor at common law, modern statutes typically make it a felony. Also termed *false making*.
- d. *OLD*: The offence of making a "false instrument" in order that it may be accepted as genuine, thereby causing harm to others.
- e. *DTJ*: falsificación.
- f. *DBTJ*: falsificación, alteración de un documento con ánimo de cometer fraude.

De acuerdo a lo señalado anteriormente, la conceptualización en COREL de \$COUNTERFEITING\_00 es la siguiente:

\$COUNTERFEITING\_00: unlawful making of false money.

+(e1: +BE\_00 (x1: \$COUNTERFEITING\_00)Theme (x2: ORGANISED\_CRIME\_00)Referent)

(f1: (e2: +CREATE\_00 (x3: +CRIMINAL\_00)Theme (x4:  
s+MONEY\_00)Referent) (f2)Instrument (e3: n+BE\_01 (x5)Theme (x4:  
+TRUE\_00)Attribute))Result

+(e4: +CHANGE\_00 (x3)Theme (x4)Referent (f3: (e5: +BECOME\_00  
(x4)Theme (x5: +REAL\_00)Attribute))Result

Unidades léxicas asociadas: *counterfeiting, falsificación de moneda.*

Otro ejemplo de conocimiento experto es *layering*. Como no hemos encontrado esta unidad léxica en los diccionarios especializados, hemos optado por buscarlo en *Google* y nos hemos encontrado con varias páginas web de ámbito legal (*AUSTRAC, FraudAid, USLegal,...*). Se trata de una de las etapas en el proceso de blanqueo de dinero, por la cual el dinero negro es transferido a cuentas bancarias en el extranjero, que suelen ser paraísos fiscales, donde entra y sale el dinero.

### ***Layering***

- a. *AUSTRAC (Australian Transaction Reports and Analysis Centre)*<sup>71</sup>: Layering is the second stage of the money laundering process, in which illegal funds or assets are moved, dispersed and disguised to conceal their origin.
- b. *Wikipedia*: Money laundering often occurs in three steps: first, cash is introduced into the financial system by some means ("placement"), the second involves carrying out complex financial transactions in order to camouflage the illegal source ("layering"), and the final step entails acquiring wealth generated from the transactions of the illicit funds ("integration").
- c. *GCTC*: Of the three phases of ML (placement, **layering** and integration), placement is undoubtedly the most risky (...).

En base a las definiciones de arriba, la formalización de \$LAYERING\_00 en COREL es la siguiente:

\$LAYERING\_00: the second stage of money laundering by which the dirty money is transferred to offshore accounts where it is moved in and out.

---

<sup>71</sup> [http://www.austrac.gov.au/elearning/pdf/intro\\_amlctf\\_layering\\_techniques.pdf](http://www.austrac.gov.au/elearning/pdf/intro_amlctf_layering_techniques.pdf)

+(e1: +BE\_00 (x1: \$LAYERING\_00)Theme (x2: +MONEY\_LAUNDERING)Referent)

+(e2: +TRANSFER\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x5)Origin (x6)Goal (x7: +TAX\_HAVEN\_00)Location (f1: +IN\_00 & +OUT\_00)Position)

Unidades léxicas asociadas: *layering*.

Como último ejemplo de unigrama especializado presentamos *waterboarding*. Se trata de una forma de tortura en la que la víctima sufre ahogamiento.

### **Waterboarding**

- a. *BLD*: 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. See TORTURE.
- b. *Wikipedia*: **Waterboarding** is a form of torture in which water is poured over the face of an immobilized captive, causing the individual to experience the sensation of drowning.

Como *waterboarding* es un tipo de tortura será concepto terminal de +TORTURE\_00.

\$WATERBOARDING\_00: a form of torture with water.

+(e1: +BE\_00 (x1: \$WATERBOARDING \_00)Theme (x2: +TORTURE\_00)Referent)

+(e2: +USE\_00 (x3: +TERRORIST\_00)Theme (x4 +WATER\_00) Referent)

+(e3: +FEEL\_00 (x3)Agent (x5: +VICTIM\_00)Theme (x3: +AFRAID\_00)Attribute)

(f1:(e4: ing+DROWN\_00 (x3)Theme (x5: +VICTIM\_00)Referent)Result)

Unidades léxicas asociadas: *waterboarding*.

Como señalábamos al comienzo de la sección 4.3., la labor de conceptualización se complica sustancialmente en el caso de los bigramas y trigramas porque no aparecen en los diccionarios especializados. Para poder elaborar las definiciones se hace necesaria la consulta de páginas web de organismos oficiales, el GCTC y *Wikipedia*. A continuación presentamos a modo de ejemplo las modalidades de fraude fiscal *carousel fraud*, *cuckoo smurfing* y *self-laundering*.

### *Carousel fraud*

- a. *CED*: the practice of importing goods from a country where they are not subject to VAT, selling them with VAT added, then deliberately not paying the VAT to the government.
- b. *Wikipedia*: **Missing trader fraud** (also called **Missing Trader Intra-Community, MTIC**, or **carousel fraud**) is the theft of Value Added Tax (VAT) from a government by organised crime gangs who exploit the way VAT is treated within multi-jurisdictional trading where the movement of goods between jurisdictions is VAT-free. This allows the fraudster (person who commits fraud) to charge VAT on the sale of goods, and then instead of paying this over to the government's collection authority, simply absconds, taking the VAT with him. The term "missing trader" refers to the fact that the trader goes missing with the VAT. "Carousel" refers to a more complex type of fraud in which VAT and goods are passed around between companies and jurisdictions, similar to how a carousel goes round and round. In the UK the fraud is investigated by HM Revenue & Customs and prosecuted by the Crown Prosecution Service and the Crown Office and Procurator Fiscal Service.
- c. *European Commission*<sup>72</sup>: Un modelo de fraude que parece haber crecido durante los últimos años y que plantea ahora una preocupación importante para los Estados miembros en el comercio intracomunitario es el denominado **fraude carrusel**. Este mecanismo de fraude es un ejemplo claro

---

<sup>72</sup> [http://ec.europa.eu/translation/bulletins/puntoycoma/88/pyc884\\_es.htm](http://ec.europa.eu/translation/bulletins/puntoycoma/88/pyc884_es.htm)

de abuso de la exención del IVA en el comercio intracomunitario, que podría muy bien aprovecharse de la próxima ampliación y extenderse a los nuevos Estados miembros.

Como *carousel fraud* es un tipo de fraude, será concepto terminal de +FRAUD\_00.

\$SCAROUSEL\_FRAUD\_00: the crime of importing goods from a tax free country. The goods are sold with taxes included but without paying them to the government.

+(e1: +BE\_00 (x1: \$SCAROUSEL\_FRAUD\_00)Theme (x2: +FRAUD\_00)Referent)

+(e2: +SELL\_00 (x3: +CRIMINAL\_00)Agent (x4: s+ALCOHOL^s+TOBACCO\_00)Theme (x5)Origin (x6)Goal (f1: (e3 n + PAY\_00 (x3)Agent (7: +TAX\_00)Theme (x5) Origin (x8: +GOVERNMENT\_00)Goal))Condition) (f2: (e3: +SELL\_00 (x3)Agent (x4)Theme (x9)Origin (x10)Goal) (f3: (e4: +PAY\_00 (x10)Agent (x11: +TAX\_00) (x9)Origin (x3)Goal))Condition)

Unidades léxicas asociadas: *carousel fraud, fraude carrusel.*

Los siguientes dos bigramas son dos tipos de blanqueo de dinero. El primero es *cuckoo smurfing*, blanqueo en el que el dinero se blanquea en la cuenta bancaria de un tercero; y el segundo *self laundering*, blanqueo en el que el delincuente blanquea dinero en su propia cuenta bancaria.

### ***Cuckoo smurfing***

- a. *AUSTRAC*<sup>73</sup>: the perpetrators of this money laundering typology seek to transfer wealth through the bank accounts of innocent third parties.
- b. *GCTC*: **Cuckoo smurfing** can be effectively disrupted by the consistent use of identification and know-your-customer procedures for the depositors of cash into third party bank accounts. **Cuckoo smurfing** allows a controller and an ARS to oversee payments made in another country without the risk of being detected.

---

<sup>73</sup> [http://www.austrac.gov.au/files/typ\\_2008\\_mlm.pdf](http://www.austrac.gov.au/files/typ_2008_mlm.pdf)

Así, la formalización de \$CUCKOO\_SMURFING\_00 en COREL es la siguiente:

\$CUCKOO\_SMURFING\_00: the crime of putting illegal money into the third parties bank accounts.

+(e1: +BE\_00 (x1: \$CUCKOO\_SMURFING\_00)Theme (x2: +MONEY\_LAUNDERING\_00)Referent)

+(e2: +PUT\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x7: +BANK\_ACCOUNT\_00)Location (x5)Origin (x6)Goal

(f1: (e3+BE\_02 (x7)Theme (x8)Location) (f2: +IN\_00)Condition))Result

Unidades léxicas asociadas: *cuckoo smurfing*.

### ***Self laundering***

- a. *GCTC*: As many criminals launder their own proceeds in these offences, the incrimination of **self laundering** in those countries involved in trafficking in human beings and illegal migration (where **self laundering** is not currently criminalised) should be actively considered.

Por su parte, la formalización en COREL de \$SELF\_LAUNDERING\_00 es así:

\$SELF\_LAUNDERING\_00: the crime of putting illegal money into the criminal's own bank account.

+(e1: +BE\_00 (x1: \$SELF\_LAUNDERING\_00)Theme (x2: +MONEY\_LAUNDERING\_00)Referent)

+(e2: +PUT\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x7: +BANK\_ACCOUNT\_00)Location (x5)Origin (x6)Goal

(f1: (e3+BE\_02 (x7)Theme (x6)Location) (f2: +IN\_00)Condition))Result

Unidades léxicas asociadas: *self laundering, autoblanqueo*.

Como hemos señalado anteriormente, aquí hemos presentado una muestra de varios conceptos especializados, extracto del apéndice 2, donde ofrecemos los PS de 10 conceptos especializados de nuestra ontología satélite del crimen organizado y el terrorismo. En dicho apéndice se pueden consultar más casos, como por ejemplo \$LONE\_WOLF\_00, \$PHARMING\_00 y \$VISHING\_00.

#### **4.3.1.3. Conclusiones**

En primer lugar señalar el gran número de conceptos seleccionados por el extractor pertenecientes al lenguaje jurídico general. Dado que estamos trabajando en una ontología satélite del crimen organizado y el terrorismo, deberían ir en la ontología nuclear, ya que la ontología satélite abarca conceptos mucho más específicos.

De acuerdo a lo observado en el proceso de elaboración de definiciones podemos establecer los siguientes criterios ontológicos:

1. Productividad del concepto, es decir, si es posible incluir varias unidades léxicas, si forma bigramas y trigramas, etc. Como señalan Mairal y Periñán (2009a: 222-223), debemos crear un nuevo concepto cuando nos encontramos al menos con una unidad léxica a la que no se le puede aplicar ninguno de los postulados de significado ya existentes en la ontología satélite, siempre y cuando las cualidades ontológicas de ese nuevo concepto sean compartidas por todas las unidades léxicas asociadas a él.
2. Evitar la circularidad de las definiciones agrupando los sinónimos y cuasi-sinónimos en un mismo concepto (*Confiscate – Seize; Crime – (Criminal) Offence/Wrong; Gang – Organised crime – Mob – Ring*). Tanto los diccionarios de uso general como los especializados nos servirán de referencia para la elaboración de definiciones en lenguaje natural que serán la base de las predicaciones que elaboraremos en COREL a través de los postulados de significado. Al analizar las entradas léxicas en los diccionarios especializados encontramos también circularidad en las definiciones de términos sinónimos (*confiscate - confiscation y seize - seizure*). Según el *Black's Law Dictionary*:

**confiscate** (kon-fa-skayt), **vb.** 1. To appropriate (property) as forfeited to the government. 2. To seize (property) by authority of law.

**confiscation** (kon-fi-skay-shdn), **n.** (16c) 1. Seizure of property for the public treasury. 2. Seizure of property by actual or supposed authority. - confiscatory (k;m-fis-b-tor-ee), adj. - confiscator (kon-f<:l-skay-t<:lr), n.

**seize, vb.** (Be) 1. To forcibly take possession (of a person or property). 2. To place (someone) in possession. 3. To be in possession (of property). 4. To be informed of or aware of (something). See SEISIN; SEIZURE.

**seizure, n.** (15c) The act or an instance of taking possession of a person or property by legal right or process; esp., in constitutional law, a confiscation or arrest that may interfere with a person's reasonable expectation of privacy.

3. Definir el concepto negativo como negación del positivo. Ej: *legal – illegal*. Así *ilegal* significa “no legal”. De esta forma se reutiliza la información del concepto positivo.

+LEGAL\_00

+(e1: +BE\_01 (x1)Theme (x2: +LEGAL\_00)Attribute)

+(e2: +PERMIT\_00 (x3: +LAW\_00)Theme (x1)Referent (x4: +HUMAN\_00)Goal)

\$LEGAL\_N\_00

\*(e1: +BE\_01 (x1)Theme (x2: \$LEGAL\_N\_00)Attribute)

+(e2: n +BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)

4. Destacar el conocimiento experto sobre el conocimiento compartido mediante los PS. En los casos de “conceptos espejo”, el concepto existente en la ontología nuclear se referencia en la ontología satélite que aporta la definición especializada con su correspondiente postulado de significado (+AGGRESSION\_00, +FREEZE\_01,...).

Con la aplicación de estos criterios ontológicos se demuestra que la metodología COHERENT, seguida en el desarrollo de la ontología nuclear, es totalmente válida en la creación de ontologías satélite de dominios especializados, ya que al analizar un dominio especializado aparece, por un lado, conocimiento compartido que se materializa en “conceptos espejo” (+CRIME\_00, +SEIZE\_00,...) que aparecerán tanto en la ontología nuclear como en la satélite pero con significado general en la primera y significado especializado en la segunda; y, por otro, el conocimiento experto (+CAROUSEL\_FRAUD\_00, +WATERBOARDING\_00,...) prototípico de los conceptos de la ontología satélite.

#### **4.4. La jerarquización cognitiva**

La definición en COREL está formada por el GENUS + DIFFERENTIAE. El *genus* es el que nos va a delimitar la jerarquía de los conceptos. El concepto hiperónimo será el concepto básico, *genus* compartido por los conceptos terminales que son sus hipónimos.

Al tratarse de un dominio específico, organizamos primero los conceptos en grupos según si pertenecen al crimen organizado o al terrorismo. Después, de acuerdo a las subontologías existentes en FunGramKB, clasificamos las entidades en #ENTITY y las acciones en #EVENT que permiten la organización interna de sustantivos y verbos propiamente dichos. Así, los conceptos definidos en COREL en el capítulo anterior se distribuyen en esas dos subontologías y se organizan jerárquicamente de acuerdo a la relación IS-A donde la subsunción es la única relación posible entre los conceptos de la taxonomía. Los conceptos aparecen unidos por relaciones de herencia e inferencia por la cual los conceptos terminales heredan las propiedades de su hiperónimo más inmediato (concepto básico). Como los metaconceptos tienen asignados un esquema temático prototípico, en este caso #TRANSFORMATION, todos los conceptos subordinados a este metaconcepto comparten dicho esquema en forma de MT. (Periñán y Mairal, 2010). En los eventos el MT aparece reflejado en el PS, por lo que no es necesario especificar las preferencias de selección. Por tanto, el evento +SEIZE\_00 adquiere el MT de +TAKE\_00 (1), al que añadimos el PS con una expresión detallada del significado propio de +SEIZE\_00 (2): “una persona (x1) toma algo (x2) de forma legal”.

(1) MT: (x1:+HUMAN\_00)Theme (x2: +ARTEFACT\_00)Referent

(2) PS: +(e1: +TAKE\_00 (x1)Theme (x2)Referent (f1: (e2

+(e2: +BE\_01 (x1) Theme (x2: +LEGAL\_00)Attribute)Manner?

En el proceso de subsunción se hace necesario introducir conceptos paraguas que facilitan el mecanismo de herencia a nivel medio (conceptos básicos) en la ontología, como es el caso de la entidad ORGANIZED\_CRIME\_00. Así, las actividades típicas del crimen organizado (*corruption, counterfeiting, cybercrime, money laundering,...*) son conceptos terminales hipónimos del concepto básico +ORGANIZED\_CRIME\_00.

Mediante un proceso de negociación decidimos la estructura de la ontología (Periñán y Arcas 2010a: 2669) para así evitar que la subjetividad de los ingenieros del conocimiento influya en la taxonomía.

Una vez que los ingenieros han decidido qué conceptos se van a incluir en la taxonomía como conceptos básicos o terminales, se transformará el significado de los mismos en COREL. El desafío posterior para el ingeniero de conocimiento radicará en la definición de los conceptos terminales con un catálogo cerrado de conceptos básicos.

De acuerdo al análisis de términos realizados en la sección anterior y al catálogo cerrado de metaconceptos, la taxonomía propuesta para el dominio del crimen organizado y el terrorismo clasificada en eventos y entidades sería la siguiente:

A continuación presentamos la ruta conceptual de cada concepto incluido en los apéndices 1 y 2 (ordenados alfabéticamente) basándonos en la taxonomía existente en la ontología nuclear, donde los “conceptos espejo” aparecen con un subrayado sencillo y los conceptos especializados con un subrayado doble.

#### **#EVENT**

#EVENT>#MATERIAL>+DO\_00>+FREEZE\_01

#EVENT>#MATERIAL>+DO\_00>+LAUNDER\_MONEY\_00

#EVENT>#MATERIAL>+DO\_00>+OBTAIN\_00>+TAKE\_00>+SEIZE\_00

**#ENTITY**

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>

+AGGRESSION\_00

#ENTITY>#PHYSICAL>#OBJECT>#SELF\_CONNECTED\_OBJECT>>

+ARTIFICIAL\_OBJECT\_00>+SOLID\_00>+BLACK\_MARKET\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>

+ORGANIZED\_CRIME\_00>\$CAROUSEL\_FRAUD\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>+ORGA

NIZED\_CRIME\_00>\$COUNTERFEITING\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>

+ORGANIZED\_CRIME\_00>\$CUCKOO\_SMURFING\_00

#ENTITY>#PHYSICAL>#OBJECT>#SELF\_CONNECTED\_OBJECT>+NATURAL\_

OBJECT\_00>+CORPUSCULAR\_00>+SOLID\_00>+BANKNOTE\_00>

\$DIRTY\_MONEY\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>

+FORGERY\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>+ORGA

NIZED\_CRIME\_00>+MONEY\_LAUNDERING\_00>\$SLAYERING\_00

#ENTITY>#PHYSICAL>#OBJECT>#SELF\_CONNECTED\_OBJECT>+NATURAL\_

OBJECT\_00>+SUBSTANCE\_00>+SOLID\_00>+ORGANISM\_00>+HUMAN\_00>

\$LONE\_SHARK\_00

#ENTITY>#PHYSICAL>#OBJECT>#SELF\_CONNECTED\_OBJECT>+NATURAL\_

OBJECT\_00>+SUBSTANCE\_00>+SOLID\_00>+ORGANISM\_00>+HUMAN\_00>

\$LONE\_WOLF\_00

#ENTITY>#PHYSICAL>#OBJECT>#SELF\_CONNECTED\_OBJECT>+NATURAL\_

OBJECT\_00>+SUBSTANCE\_00>+SOLID\_00>+ORGANISM\_00>+HUMAN\_00>

+MASTERMIND\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00>+MONEY\_LAUNDERING\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>+ORGA  
NIZED\_CRIME\_00>+CYBERCRIME\_00>+SPHARMING\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00> +CYBERCRIME\_00>+PHISHING\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00>+RACKETEERING\_00

#ENTITY>#REGION\_00>+PLACE\_00>+ARTIFICIAL\_AREA\_00>+COUNTRY\_00  
>+SAFE\_HAVEN\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00>+SELF\_LAUNDERING\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00>+SMUGGLING\_00

#ENTITY>#PHYSICAL>#OBJECT>#COLLECTION>+GROUP\_00>  
+SYNDICATE\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00> +TAX\_EVASION\_00

#ENTITY>#REGION\_00>+PLACE\_00>+ARTIFICIAL\_AREA\_00>+COUNTRY\_00  
>+TAX\_HAVEN\_00

#ENTITY>#PHYSICAL>#OBJECT>#SELF\_CONNECTED\_OBJECT>+NATURAL\_  
OBJECT\_00>+SUBSTANCE\_00>+SOLID\_00>+ORGANISM\_00>+HUMAN\_00>  
+TERRORIST\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+TORTURE\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+ORGANIZED\_CRIME\_00>+VISHING\_00

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE\_00>+CRIME\_00>  
+TORTURE\_00>\$WATERBOARDING\_00

## **4.5. Conclusiones**

En este capítulo primero hemos partido de las nociones de “corpus” y “término” adoptadas en esta tesis, claves para el desarrollo de una ontología satélite del crimen organizado y el terrorismo a partir del GCTC. Después hemos analizado los distintos pasos de la extracción terminológica, donde hemos visto que *FunGramKB Term Extractor* permite la extracción semiautomática de términos basada en la importancia estadística de cada término en un corpus determinado. A continuación, hemos analizado en detalle la fase de conceptualización de los términos, donde hemos seleccionado una muestra significativa de “conceptos espejo” y conceptos especializados. Para finalizar hemos elaborado una propuesta de jerarquización cognitiva de dichos conceptos.

En la fase de conceptualización, en la consulta de las unidades n-gramáticas en las fuentes lexicográficas, tanto generales como especializadas, nos hemos encontrado con una gran cantidad de “conceptos espejo”, donde en muchos de ellos se ha hecho tremendamente difícil distinguir el conocimiento experto del conocimiento compartido porque mucha terminología se ha hecho del uso común (p. ej. *money laundering* y *tax haven*). De hecho, nos hemos encontrado con casos en los que el conocimiento compartido y el conocimiento experto aparecen mezclados en los *Learner’s Dictionaries* (p. ej. *organized crime* y *forgery*). Si bien estos resultados eran predecibles puesto que el derecho es una disciplina social que refleja mucho conocimiento del sentido común.

Además, hemos visto, por un lado, un claro trasvase de terminología propia del derecho penal en la lengua común (p. ej. *black market* y *dirty money*) y, por otro, casos en los que el conocimiento del sentido común aparece mezclado con el conocimiento experto en la misma palabra en los *Learner’s Dictionaries* (p. ej. *organized crime* y *forgery*), mientras que los diccionarios especializados presentan una terminología específica para el conocimiento experto (*syndicate* y *counterfeiting*). De hecho, si preguntáramos a un hablante medio qué entiende por *blanqueo de dinero*, o por *paraíso fiscal* podría darnos una definición con conocimiento del sentido común, pero no sabría explicarnos la parte

legal del término, es decir, no sabría decirnos cuáles son las repercusiones jurídicas del mismo.



## Capítulo 5

# CONCLUSIONES Y SUGERENCIAS PARA POSTERIORES INVESTIGACIONES

### 5.1. Conclusiones

En este último capítulo ofrecemos un pequeño resumen de lo expuesto en cada capítulo y de las conclusiones a las que hemos llegado en cada uno de ellos, además de varias sugerencias para posteriores investigaciones.

Como anticipábamos en el capítulo 1, el desarrollo de ontologías de representación del conocimiento legal es un área interdisciplinar donde convergen las disciplinas del derecho, gestión del conocimiento e IA. El trabajo presentado en esta tesis partía de la hipótesis de que el modelo multinivel de la ontología nuclear de *FunGramKB* se puede exportar a cualquier ontología satélite con independencia de cuál sea su dominio temático. Como muestra de ello hemos presentado el diseño y desarrollo de una ontología satélite del crimen organizado y el terrorismo vinculada a la ontología nuclear de la base de conocimiento de *FunGramKB*, para demostrar que tanto el modelo teórico de *FunGramKB* como sus herramientas son reutilizables en la elaboración de cualquier ontología satélite de un dominio concreto, como se muestra concretamente en el capítulo 4.

Para ello hemos comenzado con el análisis de la problemática suscitada por el término “ontología” para luego ofrecer una clasificación de las ontologías consideradas más significativas por la literatura, las metodologías, lenguajes y herramientas utilizadas para su diseño, como base fundacional en el estado de la cuestión de las ontologías legales (capítulo 2). En dicho capítulo hemos visto que diversos autores coinciden en señalar la problemática que supone llamar ontología legal a elaboraciones muy diferentes entre sí sin distinguir si se trata de representaciones de conocimiento, bases de conocimiento, taxonomías léxicas o de ontologías propiamente dichas, aunque a veces puedan combinar diversos elementos. También hemos destacado que, al igual que la ontología nuclear de *FunGramKB*, nuestra ontología satélite del crimen organizado y

el terrorismo está basada en la semántica profunda, por lo que tiene un enfoque conceptualista que permite definir los términos de forma conceptual, por lo que son independientes de cada lengua. Así se evitan los problemas presentados por el léxico de cada lengua integrada en la ontología satélite (inglés, español e italiano) ya que los conceptos son unidades abstractas independientes de cada lengua. Por otra parte, como el dominio del derecho es una ciencia social, hay una gran cantidad de términos del derecho en el lenguaje general (por ejemplo, “juez” y “abogado”) que están incluidos en la ontología nuclear de *FunGramKB* con conocimiento del sentido común. En otras palabras, y como hemos señalado anteriormente en relación a los “conceptos espejo”, en la ontología nuclear aparece reflejado el conocimiento del sentido común y en la satélite el conocimiento especializado, por lo que, de ese modo se reutiliza en la ontología satélite la información contenida en los PS de los “conceptos espejo” de la ontología nuclear a la vez que se minimiza la redundancia informativa. De hecho, la ontología satélite del crimen organizado y el terrorismo presentada en esta tesis ofrece una nueva perspectiva a las ontologías legales ya que, por un lado tiene como propósito estructurar la información para una fácil recuperación en contextos profesiones y, por otro, la resolución de tareas en situaciones reales.

A continuación, en el capítulo 3, hemos analizado en detalle la base de conocimiento léxico-conceptual multipropósito *FunGramKB*, diseñada para la realización de tareas de procesamiento del lenguaje natural, que, entre otras, puede ser reutilizada para la recuperación y extracción de información, traducción automática y en diversas lenguas. También hemos ofrecido una descripción detallada del lenguaje de notación COREL que, con una sintaxis y semántica propias permite construir los MT y los PS que dan cohesión al módulo conceptual de *FunGramKB*, donde la representación del conocimiento contiene una mayor expresividad. También hemos señalado que hemos elegido esta base de conocimiento y no otra porque su modelo teórico y sus herramientas pueden ser reutilizadas para la elaboración de cualquier ontología satélite, con independencia de cuál sea su dominio temático. De hecho, el modelo multinivel de la ontología nuclear (metaconceptos, conceptos básicos y terminales), que permite la representación conceptual reflejo de la estructura cognitiva humana, es exportado a la ontología satélite del crimen organizado y el terrorismo, donde adquiere una especial

incidencia ya que se facilita sustancialmente la representación del conocimiento experto, tan difícil de representar en algunos casos.

Seguidamente, en el capítulo 4 hemos presentado el proceso de elaboración de un corpus especializado mediante la recopilación de textos, que nos han servido para la extracción terminológica. Además, hemos hecho un análisis del proceso de extracción terminológica, para lo que hemos presentado *FunGramKB Term Extractor* como herramienta válida en la extracción terminológica que asiste al terminólogo en la explotación de corpus especializados, previa a la fase de conceptualización de los términos. Asimismo hemos mostrado que mediante la extracción de términos basada en corpus no sólo se crean ontologías satélite más completas y mejor definidas, sino también se enriquece el conocimiento especializado y en el modelado conceptual del mundo. Como prueba de ello hemos ofrecido una muestra de conceptualización de los términos (sección 4.3.1.) y de la jerarquización cognitiva (sección 4.4.) de la ontología satélite del crimen organizado y el terrorismo. Dicha muestra conceptual se complementa con los dos apéndices ofrecidos al final de la tesis.

En el proceso de elaboración de definiciones hemos tenido que resolver varios problemas. Por un lado nos hemos encontrado con un gran número de “conceptos espejo” donde en muchos casos su conocimiento experto era difícil de definir dada la gran cantidad de sentido común presente en las definiciones de los *Learner’s Dictionaries* (p. ej. *dirty money* y *tax haven*); y, por otro, cómo la labor del terminólogo se complica principalmente en la consulta de los bigramas y trigramas en las fuentes lexicográficas tanto generales como especializadas dado que muchas de estas unidades n-gramáticas no se encuentran en ninguna de esas fuentes. Para solucionar estos problemas la consulta al GCTC ha sido de vital importancia.

En el proceso de modelado ontológico hemos presentado los “conceptos espejo” y el modelo en espiral, que nos sirven tanto para la reutilización de la información conceptual ya existente en la ontología nuclear como para reducir la redundancia informativa. De igual modo, los ejemplos mostrados del dominio del crimen organizado y el terrorismo nos han servido para probar que el enfoque conceptualista de *FunGramKB* nos permite reutilizar la ontología nuclear en el desarrollo de ontologías satélite ya que éstas la enriquecen a través del conocimiento experto formalizado en COREL.

Tras la fase de conceptualización de los términos, y tomando como base la metodología COHERENT, hemos establecido los siguientes criterios ontológicos: productividad del concepto, evitar la circularidad de las definiciones agrupando los sinónimos y cuasi-sinónimos en un mismo concepto, definir el concepto negativo como negación del positivo y destacar el conocimiento experto sobre el conocimiento compartido mediante los PS. Mediante la aplicación de estos criterios ontológicos se demuestra, por tanto, que la reutilización de un mismo modelo teórico y herramientas que sirvieron para la construcción de la ontología nuclear es válida para la construcción y modelado de ontologías satélite.

El lenguaje está en continuo movimiento y como tal sufre transformaciones constantes. Vivimos en tiempos de cambiante transformación en los que los términos especializados se mezclan con el lenguaje común continuamente. Con el desarrollo de las tecnologías de la comunicación (teléfonos móviles, internet,...) el trasvase de terminología a la lengua común es imparable, que en muchos casos se da en forma de préstamos de la lengua inglesa (*WhatsApp, trending topic...*). Con respecto a la terminología del crimen organizado y el terrorismo, lamentablemente estamos familiarizados con ella porque las noticias sobre la materia son frecuentes en los distintos medios de comunicación. De hecho, si preguntáramos a un hablante medio qué entiende por *dinero negro* o por *kamikaze* podría explicarnos de qué se trata en virtud al conocimiento del sentido común del término, aunque no sabría decirnos cuáles son las repercusiones jurídicas del mismo.

En definitiva, esta tesis ha pretendido servir de base metodológica para la elaboración de futuras ontologías satélite que se desarrollen dentro del marco de *FunGramKB*, puesto que por el momento no está definida una metodología para el desarrollo de ontologías satélite vinculadas a la ontología nuclear de *FunGramKB*. En otras palabras, esta tesis ha pretendido servir de plataforma de trabajo para futuros estudios más completos sobre la materia.

Por último señalar las razones que me han llevado a realizar esta tesis en español. El motivo principal que me ha incentivado y animado a realizar la tesis en español no ha sido la comodidad de hacerla en mi lengua materna (que de hecho no ha sido tal dado que prácticamente toda la bibliografía sobre la materia está en inglés), sino realizar un

trabajo de investigación que sirva para animar a que se hagan futuros estudios en español. Puesto que es labor del investigador abrir nuevos campos de investigación a la comunidad científica de la forma más atractiva posible, aspiro a que mi pequeña aportación sea suficientemente inspiradora y motivadora para aquellos que quieran ampliar conocimientos en esta área.

La lingüística, como rama del conocimiento, necesita de una innovación constante que le permita la apertura de nuevos campos de investigación que atraigan a nuevos investigadores motivados por las nuevas aproximaciones al estudio de la lengua. Por ello, me gustaría que esta tesis sirviera de plataforma de trabajo para futuras investigaciones en español sobre ontologías.

Como bien dice el Talmud, al que hacíamos referencia al principio de esta tesis: “el hombre que tiene conocimientos lo tiene todo, si no tiene conocimientos no tiene nada”. Por tanto, es función del investigador procurar que a la humanidad nunca le falten conocimientos nuevos.

## **5.2. Sugerencias para posteriores investigaciones**

En la fase de extracción de términos aparecieron un gran número de términos del lenguaje jurídico general, lo que dificultó la toma de decisiones para elaborar la jerarquía conceptual. El subdominio del crimen organizado y el terrorismo, al ser parte del derecho penal, se ve contagiado de un gran número de términos del derecho procesal penal, por lo que durante el proceso de construcción de la ontología se hace determinante la toma de decisiones de los términos a incluir en la misma. En esta fase juega, por tanto, un importantísimo papel la consulta a los expertos para decidir los términos esencialmente relevantes en la ontología satélite.

Dada la gran cantidad de términos del derecho penal obtenidos a partir del GCTC, y que escapan a las dimensiones de esta tesis, como sugerencia para posteriores investigaciones estarían una ontología nuclear del derecho penal y un glosario terminológico y diccionarios especializados para labores traductológicas o didácticas.



## Bibliografía

- Abecker, A., A. Bernardi, K. Hinkelmann, O. Kühn y M. Sintek. 1998. Toward a Technology for Organizational Memories. *IEEE Intelligent Systems* 13 (3): 40-48.
- Aguado, G., E. Montiel y J. Ramos. 2007. Multilingüalidat en una aplicaci3n basada en el conocimiento. *Procesamiento del Lenguaje Natural*, 38: 77-97.
- Alcaraz, E. y B. Hughes. 2007. *Diccionario de t3rminos jur3dicos*. 10<sup>a</sup> edici3n. Barcelona: Ariel.
- Allen, J. 1983. Maintaining knowledge about temporal intervals, en P. Adarraga y J. Zaccagnini, (Eds.) *Psicología e Inteligencia Artificial*, págs. 141-186. Madrid: Trotta.
- Allen, J. 1984. Towards a general theory of action and time. *Artificial Intelligence* 23: 123-154.
- Antoniou, G. y F. van Harmelen. 2004. *A Semantic Web Primer*. Cooperative Information Systems. The MIT Press.
- Arp3rez, J., O. Corcho, M. Fern3ndez y A. G3mez. 2003. WebODE in a nutshell. *AI Magazine*. 24 (3): 37-48.
- Arp3rez, J., A. G3mez, A. Lozano y S. Pinto. 1998. (ONTO)<sup>2</sup> Agent: An ontology-based www broker to select ontologies, en A. Gomez-Perez y V. R. Benjamins (Eds.), *Proceedings of the Workshop on Applications of Ontologies and Problem-Solving Methods*, realizado junto con ECAI-98, págs. 16-24. Brighton, Gran Bretaña, agosto de 1998. ECAI.
- Asaro, C., M. Biasiotti, P. Guidotti, M. Papini, M. Sagri y D. Tiscornia. 2003. A Domain Ontology: Italian Crime Ontology, en *ICAAIL 2003 Workshop on Legal Ontologies and Web Based Legal Information Management*. Edimburgo, Escocia, Gran Bretaña.

- Atkins, B., J. Clear y N. Ostler. 1992. Corpus Design Criteria. *Literary and Linguistic Computing*. Vol. 7, No. 1: 1-16.
- Aussenac, N., B. Biébow y S. Szulman. 2000. Corpus analysis for conceptual modeling, en *Workshop on Ontologies and Texts, Knowledge Engineering and Knowledge Management: Methods, Models and Tool, (EKAW'2000)*. Juan-les-pins, Francia.
- Barlas, C. 1995. The imprimatur project, en *KnowRight '95: Proceedings of the conference on Intellectual property rights and new technologies*, págs. 264–272, Munich, Alemania. R. Oldenbourg Verlag GmbH.
- Bateman, J., R. Henschel y F. Rinaldi. 1995. *The Generalised Upper Model 2.0. Technical Report*. IPSI/GMD, Darmstadt. Disponible en línea. Último acceso 01/05/2012.  
[<http://www.fb10.uni-bremen.de/anglistik/langpro/webSPACE/jb/gum/gum-2.pdf>].
- Bateman, J., R. Kasper, J. Moore y R. Whitney. 1990. *A General Organization of Knowledge for Natural Language Processing: The Penman Upper Model. Technical Report*. USC/Information Sciences Institute, Marina del Rey, California. Disponible en línea. Último acceso 01/05/2012.  
[<http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/um89/UM89-penman.pdf>].
- Bechhofer, S., I. Horrocks, C. Goble y R. Stevens. 2001. OilEd: a reasonable ontology editor for the Semantic Web, en F. Baader, G. Brewka y T. Eiter (Eds.) *Joint German/Austrian conference on Artificial Intelligence (KI '01)*. Viena, Austria. *Lecture Notes in Artificial Intelligence LNAI 2174*. Springer-Verlag, Berlín, Alemania, págs. 396-408.
- Beckett, D. y B. McBride. 2004. Resource Description Framework (RDF) Model and XML Syntax Specification. W3C Recommendation. 10 de febrero de 2004. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/TR/REC-rdf-syntax/>].

- Benjamins, V., P. Casanovas, J. Breuker y A. Gangemi (Eds.). 2005a. *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. Lecture Notes in Artificial Intelligence 3369*. Springer, Verlag, Berlin, Heidelberg.
- Benjamins, V., P. Casanovas, J. Breuker y A. Gangemi. 2005b. Law and the Semantic Web, an Introduction, en V. Benjamins, P. Casanovas, J. Breuker y A. Gangemi (Eds.), *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. Lecture Notes in Artificial Intelligence 3369*: 1-17. Springer, Verlag, Berlin, Heidelberg.
- Benjamins, V., P. Casanovas, J. Contreras, J. M. Lopez Cobo y L. Lemus. 2005. Iuriservice: An Intelligent Frequently Asked Questions System to Assist Newly Appointed Judges, en V. Benjamins, P. Casanovas, J. Breuker y A. Gangemi (Eds.), *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. Lecture Notes in Artificial Intelligence 3369*: 201-217. Springer, Verlag, Berlin, Heidelberg.
- Benjamins, V., J. Contreras, P. Casanovas, M. Ayuso, M. Becue, L. Lemus y C. Urios. 2003. Ontologies of Professional Legal Knowledge as the Basis for Intelligent IT Support for Judges, en *ICAIL 2003 Workshop on Legal Ontologies & Web Based Legal Information Management*. 28 de junio de 2003. Edimburgo, Escocia.
- Benjamins, V., D. Fensel, S. Decker y A. Gómez. 1999. KA<sup>2</sup>: Building ontologies for the Internet: a midterm report. *International Journal of Human Computer Studies* 51: 687-712.
- Benton, J. y S. Kambhampati. 1997. Toward Distributed Use of Large-Scale Ontologies, en A. Farquhar, M. Gruninger, A. Cómez, M. Uschold y P. van der Vet (Eds.), *Proceedings of the AAAI '97 Spring Symposium on Ontological Engineering*, págs. 138-148. Stanford University, California.
- Benton, J. y S. Kambhampati. 2002. OntoEdit: Collaborative Ontology Engineering for the Semantic Web, en I. Horrocks y J. Hendler (Eds.), *Proceedings of the First International Semantic Web Conference (ESWC '02). Lecture Notes in*

- Computer Science LNCS 2342*, págs. 221-235. Springer-Verlag, Berlín, Alemania, Cerdeña, Italia.
- Bergamaschi, S., S. Castano, D. Beneventano y M. Vincini. 2001. Semantic Integration of Heterogeneous Information Sources. Special Issue on Intelligent Information Integration. *Data & Knowledge Engineering* 36 (3): 215-249.
- Bernaras A., I. Laresgoiti y J. Corera 1996. Building and reusing ontologies for electrical network applications, en W. Wahlster (Ed.), *European Conference on Artificial Intelligence (ECAI '96)*. Budapest, Hungría. John Wiley and sons, Chichester, Gran Bretaña, págs. 298-302.
- Berners-Lee, T. 1999. *Weaving the web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor*. Harper-Collins Publishers, New York.
- Berners-Lee, T. 2000. *Semantic Web on XML. XML 2000*. Washington D.C., 6 de diciembre de 2000. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>].
- Berners-Lee, T., J. Hendler y O. Lassila. 2001. The Semantic Web, en *Scientific American*, mayo 2001.
- Biber, D. 1993. Representativeness in Corpus Design, *Literary and Linguistic Computing* 8 (4): 243-57.
- Blázquez, M., M. Fernández, J. García y A. Gómez. 1998. Building Ontologies at the Knowledge Level using the Ontology Design Environment, en B. Gaines y M. Musen (Eds.), *11<sup>th</sup> International Workshop on Knowledge Acquisition, Modeling Management (KAW '98)*. Banff, Canadá. SHARE 4: 1-15.
- Boer, Alexander. 2009. *Legal Theory, Sources of Law and the Semantic Web*. Frontiers in Artificial Intelligence and Applications. Vol. 195. Amsterdam: IOS Press.
- Boer, A., M. di Bello, K. van den Berg, T. Gordon, A. Förhéczy, y R. Vas. 2007. *Specification of the legal knowledge interchange format Summary. Deliverable 1.1., ESTRELLA project (IST-2004-027655)*. Disponible en línea. Último acceso 01/05/2012. [<http://dare.uva.nl/document/167498>].

- Boer, A., R. Hoekstra y R. Winkels. 2001. The CLIME Ontology. *Proceedings of the Second International Workshop on Legal Ontologies*, University of Amsterdam, págs. 37-47.
- Borgo, S., N. Guarino, y C. Masolo. 1996. Stratified ontologies: the case of physical objects, en *Proceedings of the Workshop on Ontological Engineering*, realizado junto con ECAI-96, págs. 5–15. Budapest, agosto de 1996. ECAI.
- Borst, W. N. 1997. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. Tesis Doctoral. Centre for Telematica and Information Technology. University of Twente. Enschede, Holanda.
- Bossini, F., M. Gleeson y S. Arana. 2008. *Diccionario bilingüe de términos legales Inglés-Español/Español-Inglés*. 5ª edición. Granada: Comares.
- Bray, T., J. Paoli, C.M. Sperberg-McQueen y E. Maler. 2000. *Extensible Markup Language (XML) 1.0. (Fifth Edition)W3C Recommendation*. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/TR/xml/>].
- Breuker, J. 2003. The construction and use of ontologies of criminal law in the e-Court European project, en V. Golaczynski (Ed.), *Proceedings of Means of electronic communication in court administration*, págs. 15–40. ILV, Wrocław.
- Breuker, J. 2004. Constructing a legal core ontology: LRI-Core, en *Proceedings of the Workshop on Ontologies and their Applications*, Sao Luis, Maranhao, Brasil.
- Breuker, J., A. Elhag, E. Petkov y R. Winkels. 2002. Ontologies for legal information serving and knowledge management, en T. Bench-Capon, A. Daskalopulu y R. Winkels (Eds.), *Legal Knowledge and Information Systems*. Jurix 2002: The Fifteenth Annual Conference, págs. 73–82, Amsterdam. IOS Press.
- Breuker, J., A. Elhag, E. Petkov y R. Winkels. 2003. IT support for the judiciary: Use of ontologies in the e-court project, en *Proceedings of the ICCS Conference on Conceptual Graphs*, págs. 17-26.
- Breuker, J., P. Casanovas, M. Klein y E. Francesconi (Eds.). 2009a. *Law, Ontologies and the Semantic Web*. Frontiers in Artificial Intelligence and Applications. Vol. 188. Amsterdam: IOS Press.

- Breuker, J., P. Casanovas, M. Klein y E. Francesconi 2009b. The Flood, the Channels and the Dykes: Managing Legal Information in a Globalized and Digital World en J. Breuker, P. Casanovas, M. Klein y E. Francesconi (Eds.), *Law, Ontologies and the Semantic Web*. Frontiers in Artificial Intelligence and Applications. Vol. 188: 3-18. Amsterdam: IOS Press.
- Breuker, J. y R. Hoekstra. 2004a. Core concepts of law: Taking common-sense seriously, en *Proceedings of Formal Ontologies in Information Systems. FOIS-2004*, págs. 210–221. IOS-Press.
- Breuker, J. y R. Hoekstra. 2004b. Epistemology and ontology in core ontologies: FOLaw and LRI-Core, two core ontologies for law, en A. Gangemi y S. Borgo (Eds). *Proceedings of Workshop on Core Ontologies in Ontology Engineering in the 14th International Conference (EKAW'04.)* Whittlebury Hall. CEUR.
- Breuker, J., R. Hoekstra, A. Boer, K. van den Berg, R. Rubino, G. Sartor, M. Palmirani, A. Wyner y T. Bench-Capon. 2007. *Owl ontology of basic legal concepts (LKIF-Core). Deliverable 1.4 D.1.4, ESTRELLA project (IST-2004-027655)*. Disponible en línea. Último acceso 01/05/2012. [<http://www.estrellaproject.org/doc/D1.4-OWL-Ontology-of-Basic-Legal-Concepts.pdf>].
- Breuker, J., A. Valente y R. Winkels. 1997. Legal Ontologies: A Functional View, en *Proceedings of the First International Workshop on Legal Ontologies*, Melbourne, Australia, págs. 23-36.
- Breuker, J., A. Valente y R. Winkels. 2005. Use and Reuse of Legal Ontologies in Knowledge Engineering and Information Management, en V. Benjamins, P. Casanovas, J. Breuker y A. Gangemi (Eds.), *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. in Artificial Intelligence 3369*: 36-64. Springer, Verlag, Berlin, Heidelberg.
- Brickley, D. y R. Guha. 2004. *RDF Vocabulary Description Language. 1.0.: RDF Schema. W3C Working Draft*. 10 de febrero de 2004. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/TR/PR-rdf-schema>].

- Bruijn, J. De. 2006. *The Web Service Modeling Language WSMML. Deliverable D 16.1v0.21. Informe técnico*. ESSI WSMO working group. Disponible en línea. Último acceso 01/05/2012. [<http://www.wsmo.org/TR/d16/d16.1/v0.21>].
- Cabré, M. 1993. *La terminología. Teoría, metodología, aplicaciones*. Empúries: Barcelona.
- Cabré, M. 1999. *La terminología: representación y comunicación*. Universidad Pompeu Fabra.
- Carrión, M. (En prensa). Extracción y análisis de unidades léxico-conceptuales del dominio jurídico: un acercamiento metodológico desde FunGramKB, en *Revista Electrónica de Lingüística Aplicada 11*.
- Casellas, N. 2008. *Modelling Legal Knowledge through Ontologies. OPJK: the Ontology of Professional Judicial Knowledge*. Tesis Doctoral. Universidad Autónoma de Barcelona.
- Casellas, N. 2011. *Legal Ontology Engineering: Methodologies, modeling trends, and the Ontology of the Professional Judicial Knowledge*. Law, Governance and Technology Series 3, 2011. Volumen 3. London / New York: Springer-Verlag.
- Casanovas, P., X. Binefa, C. Gracia, E. Teodoro, N. Galera, M. Blázquez, M. Poblet, J. Carrabina, M. Monton, C. Montero, J. Serrano y J. López. 2009. The e-Sentencias Prototype: A Procedural Ontology for Legal Multimedia Applications in the Spanish Civil Courts, en J. Breuker, P. Casanovas, M. Klein y E. Francesconi (Eds.), *Law, Ontologies and the Semantic Web*, págs. 199-219.
- Casanovas, P., N. Casellas y J. Vallbé. 2009. An Ontology-Based Decision Support System for Judges, en J. Breuker, P. Casanovas, M. Klein y E. Francesconi (Eds.), *Law, Ontologies and the Semantic Web*. Frontiers in Artificial Intelligence and Applications. Vol. 188. Amsterdam: IOS Press, págs. 165-175.
- Chandrasekaran, B., J. Josephson, y V. Benjamins. 1999. What are ontologies, and why do we need them? en *Intelligent Systems and their Applications*, IEEE, 14 (1): 20–26, January 1999.

- Chaudhri V., A. Farquhar, R. Fikes, P. Karp y J. Rice. 1998. *Open Knowledge Base Connectivity 2.0.3. Technical Report*. Disponible en línea. Último acceso 01/05/2012. [<http://www.ai.sri.com/~okbc/okbc-2-0-3.pdf>].
- Conen, W. y R. Klasing. 2001. *Logical interpretations of RDFS – Compatibility Guide. Working paper*. Noviembre de 2001. Disponible en línea. Último acceso 01/05/2012. [[http://nestroy.wi-inf.uni-essen.de/rdf/new\\_interpretation](http://nestroy.wi-inf.uni-essen.de/rdf/new_interpretation)].
- Corcho, O., M. Fernández, A. Gómez y A. López. 2005. Building Legal Ontologies with METHONTOLOGY and WebODE, en V. Benjamins, P. Casanovas, J. Breuker y A. Gangemi (Eds.), *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. Lecture Notes in Artificial Intelligence 3369*: 142-157. Springer, Verlag, Berlin, Heidelberg.
- Corcho, O. y A. Gómez. 2000. A Roadmap to Ontology Specification Languages en *12th International Conference on Knowledge Engineering and Knowledge Management (EKAW'00)*. Octubre 2-6, 2000. Juan-les-Pins, Francia. *Lecture Notes in Artificial Intelligence 1937*: págs. 80-96. Springer, Verlag, Berlin, Heidelberg.
- Corcho, O. y A. Gómez. 2001. Solving integration problems of e-commerce standards and initiatives through ontological mappings, en A. Gómez, M. Grüninger, H. Stuckenschmidt y M. Uschold (Eds.), *IJCAI '01 Workshop on Ontologies and Information Sharing*. Sheattle, Washington, págs. 131-140. CEUR Workshop Proceedings. 47: 131-140. Amsterdam. Holanda.
- Davis, R., H. Shrobe y P. Szolovits. 1993. What is a knowledge representation? en *AI Magazine*, Spring, 17 – 33.
- Dean, M. y G. Schreiber. 2004. *OWL Web Ontology Language Reference. W3C Recommendation*. 10 de febrero de 2004. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/TR/2004/REC-owl-ref-20040210/>].
- Dean, M. y G. Schreiber. 2009. *OWL 2 Web Ontology Language Document Overview. W3C Recommendation*. 27 de octubre de 2009. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/TR/owl2-overview/>].

- Decker, S., M. Erdmann, D. Fensel y R. Studer. 1999. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information, en R. Meersman, Z. Tari y S. Stevens (Eds.), *Semantic Issues in Multimedia Systems (DS-8)*. Rotorua, Nueva Zelanda. Kluwer Academic Publisher, Boston, Massachusetts, págs. 351-369.
- Delgado, J., I. Gallego, R. García, y R. Gil. 2002. An ontology for intellectual property rights: IPRonto, en R. Malyankar (Ed.), *Collected Posters of the First International Semantic Web Conference (ISWC 2002)*. Cerdeña, Italia, pág. 111.
- Delgado, J., I. Gallego, S. Llorente, y R. García. 2003. IPRonto: An ontology for digital rights management, en D. Bourcier (Ed.), *Legal Knowledge and Information Systems. Jurix 2003: The Sixteenth Annual Conference*, 111-120, Amsterdam. IOS Press.
- Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). Disponible en línea. Último acceso 01/05/2012. [<http://www.loa.istc.cnr.it/DOLCE.html>].
- Despres, S. y S. Szulman. 2007. Merging of legal micro-ontologies from European Directives. *Artificial Intelligence and Law* 15: 187–200.
- Dieng-Kuntz, R., O. Corby, A. Giboin y M. Ribière. 1998. Methods and Tools for Corporate Knowledge Management, en B. R. Gaines y M. Musen (Eds.), *Eleventh Workshop on Knowledge Acquisition, Modeling Management (KAW '98)*. Banf, Canadá, KM 3: 1-20.
- Dik, S. 1997. *The theory of Functional Grammar. Part 1: The structure of the clause*. Berlín/Nueva York: Mouton de Gruyter.
- Dixon, R. 1991. *A new approach to English grammar, on semantic principles*. Oxford: Clarendon Press.
- Doering, J. y M. Herberger. 2003. Ontologies in the legal domain. *Artificial Intelligence and Law*.
- Domingue, J. 1998. Tadzebao and WebOnto: Discussing, Browsing, and Editing Ontologies on the Web, en B. Gaines y M. Musen (Eds.), *11<sup>th</sup> International*

- Workshop on Knowledge Acquisition, Modelling and Management (KAW '98)*. Banff, Canadá, KM 4: 1-20.
- Faber, P., C. López y M. Tercedor. 2001. La utilización de técnicas de corpus en la representación del conocimiento médico. *Terminology* 7 (2): 167-197. Ámsterdam/Filadelfia: John Benjamins.
- Faber, P., R. Mairal y P. Magaña 2011. Linking a Domain-Specific Ontology to a General Ontology. *Proceedings of the 24th International Flairs (Florida Artificial Intelligence Research Society) Conference*. AAAI Press (Association for the Advancement of Artificial Intelligence).
- Farquhar, A., R. Fikes y J. Rice. 1997. The Ontolingua Server: A Tool for Collaborative Ontology Construction. *International Journal of Human Computer Studies* 46 (6): 707-727.
- Felices, A. y A. Marín. (En prensa): La semántica profunda como fundamento para desarrollar una subontología jurídica en el contexto de las ontologías de ámbito legal. En: Actas del XXIX Congreso Internacional de AESLA.
- Felices, A., P. Ureña y A. Alameda. 2011. FunGramKB y la adquisición terminológica. *Anglogermánica Online* 2011: 66-86.
- Fensel, D. 2001. *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer-Verlag. Berlín, Alemania.
- Fensel, D., I. Horrocks, F. van Harmelen, S. Decker, M. Erdmann y M Klein. 2000. OIL in a nutshell, en R. Dieng y O Corby (Eds.), *12<sup>th</sup> International Conference in Knowledge Engineering and Knowledge Management (EKAW '00)*. Juan-Les-Pins, Francia. *Lecture Notes in Artificial Intelligence LNAI 1937*. Springer-Verlag. Berlín, Alemania, págs. 1-16.
- Fernández, M., A. Gómez y N. Juristo. 1997. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Spring Symposium on Ontological Engineering of AAAI. Stanford University, California, págs. 33-40.

- Fernández, M., A. Gómez, A. Pazos y J. Pazos. 1999. Building a Chemical Ontology Using Methontology and the Ontology Design Environment. *IEEE Intelligent Systems & their applications* 4 (1): 37-46.
- Flowerdew, L. 2004. The Argument for Using English Specialized Corpora to Understand Academic and Professional Settings, en U. Connor y T. Upton (Eds.) *Discourse in the Professions: Perspectives from Corpus Linguistics*. Amsterdam: John Benjamins, págs. 11-33.
- Fox, M. S. 1992. The TOVE Project: A Common-sense Model of the Enterprise, en F. Belli y F. Radermacher (Eds.), *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Lecture Notes in Artificial Intelligence LNAI 604*. Springer-Verlag, Berlín, Alemania, págs. 25-34.
- Francesconi, E., P. Spinosa y D. Tiscornia. 2007. A linguistic-ontological support for multilingual legislative drafting: the DALOS Project.
- Francis, W. 1982. Problems Assembling and Computerizing Large Corpora, en Johansson, S. (Ed.), *Computer Corpora in English Language Research*, págs. 124-136.
- FunGramKB Suite. [<http://www.fungramkb.com/>].
- Gangemi, A., N. Guarino, C. Masolo, A. Oltramari y L. Schneider. 2002. Sweetening ontologies with DOLCE, en Asunción Gómez-Pérez y R. Benjamins (Eds.), *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web: 13th International Conference, EKAW 2002, Sigüenza, 1-4*.
- Gangemi, A., N. Guarino y A. Oltramari. 2001. Conceptual analysis of lexical taxonomies: the case of Wordnet top-level, en B. Smith y C. Welty (Eds.), *Internal Conference on Formal Ontology in Information Systems (FOIS '01)*. Ogunquit, Maine. ACM Press, Nueva York, págs. 3-15.
- Gangemi, A., A. Prisco, M. T. Sagri, G. Steve, y D. Tiscornia 2003. Some ontological tools to support legal regulatory compliance, with a case study, en R. Meersman, y Z. Tari (Eds.), *On The Move to Meaningful Internet Systems 2003: OTM 2003 Workshops*. Sicilia, Italia. 3-7 de noviembre. Volumen 2889 de *Lecture Notes in Computer Science*, págs. 607–620. Springer-Verlag.

- Gangemi, A., D. Pisanelli y G. Steve. 1998. Some Requirements and Experiences in Engineering Terminological Ontologies over the WWW en B. R. Gaines, M. A. Musen (Eds.), *11<sup>th</sup> International Workshop on Knowledge Acquisition, Modeling and Management (KAW '98)*. Banff, Canadá, SHARE 10: 1-20.
- Gangemi, A., M. Sagri y D. Tiscornia. 2003. Metadata for content description in legal information, en *14th International Workshop on Database and Expert Systems Applications (DEXA'03)*. 1-5 de septiembre, Praga, República Checa, pág. 745. Washington, DC, EEUU. *IEEE Computer Society*.
- Gangemi, A., M. Sagri y D. Tiscornia. 2005. A Constructive Framework for Legal Ontologies, en V. Benjamins, P. Casanovas, J. Breuker y A. Gangemi (Eds.), *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. Lecture Notes in Artificial Intelligence 3369*: 97-124. Springer, Verlag, Berlin, Heidelberg.
- Garner, B. (Ed.). 2009. Black's Law Dictionary. 9<sup>a</sup> edición. USA.
- Genesereth, M., R. Fikes. 1992. *Knowledge Interchange Format. Version 3.0. Reference Manual. Technical Report Logic-92-1*. Computer Science Department. Stanford University, California. Disponible en línea. Último acceso 01/05/2012. [<http://logic.stanford.edu/kif/Hypertext/kif-manual.html>].
- Goddard, C. y A. Wierzbicka (Eds.) 2002. *Meaning and universal grammar*. Ámsterdam: John Benjamins.
- Gómez, A. 1994a. Some Ideas and Examples to Evaluate Ontologies. *Knowledge Systems Laboratory*, Stanford University, California.
- Gómez, A. 1994b. From Knowledge Based Systems to Knowledge Sharing Technology: Evaluation and Assessment. *Knowledge Systems Laboratory*. Stanford University, California.
- Gómez, A. 1996. Towards a method to conceptualize domain ontologies, en P. van der Vet (Ed.) *ECAI' 96 Workshop on Ontological Engineering*. Budapest, Hungría, págs. 41-52.

- Gómez, A. 1999a. Knowledge Sharing and Reuse: Ontologies and Applications. *Tutorial on Ontological Engineering: IJCAI'99*.
- Gómez, A. 1999b. Ontological engineering: A state of the art. *Expert Update: Knowledge Based Systems and Applied Artificial Intelligence*. British Computer Society, 2 (3): 33–43.
- Gómez, A. 2004. Ontology Evaluation. *International Handbook on Information Systems*. Springer, págs. 251-274.
- Gómez, A. 2006. Evaluation of Ontologies. *International Journal of Intelligent Systems*. 16(3), págs. 391-409.
- Gómez, A. y V. Benjamins. 1999. Overview of knowledge, sharing and reuse components: Ontologies and problem-solving methods en *Proceedings of the Workshop on Ontologies and Problem-Solving Methods* in conjunction with the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99) Stocolmo, Suecia, págs. 1.1–1.15.
- Gómez, A. y O. Corcho. 2002. Ontology Languages for the Semantic Web, en *IEEE Intelligent Systems*, 17 (1). págs. 54-60.
- Gómez, A., M. Fernández y O. Corcho. 2004. *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. London / New York: Springer-Verlag.
- Gómez, A., M. Fernández y O. Corcho. 2008. Ontologías, en J. Palma y R. Marín (Eds.), *Inteligencia artificial: Métodos, técnicas y aplicaciones*, págs. 171-205. Madrid: Mc Graw-Hill / Interamericana de España S.A.
- Gómez, A., N. Juristo, C. Montes y J. Pazos. 1997. *Ingeniería del Conocimiento: Diseño y Construcción de Sistemas Expertos*. Ceura, Madrid, España.
- Gordijn, J., H. de Bruin, J. Akkermans. 2001. Scenario methods for viewpoint integration in e-business requirements engineering, en Sprague R. H. Jr (Ed.) 34<sup>th</sup> Hawaii International Conference On System Sciences (HICSS-34)-volume 7 p. 7032. Hawai. IEEE CS Press, Los Alamitos, California.

Gruber, T. 1992a. *Ontolingua: A mechanism to support portable ontologies*. Technical report, Knowledge Systems Laboratory, Stanford University. Disponible en línea. Último acceso 01/05/2012.

[<http://greenstone.isofts.kiev.ua/gSDL/collect/ontology/index/assoc/HASH7f00.dir/doc.pdf>].

Gruber, T. 1992b. What is an ontology? Disponible en línea. Último acceso 01/05/2012. [<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>].

Gruber, T. 1993. A Translation Approach to Portable Ontology Specifications, en *Knowledge Acquisition*, 5 (2): 199-220.

Gruber, T. 1995. Toward Principles for the Design of Ontologies Used for Knowledge Sharing en *International Journal Human-Computer Studies*. 43 (5-6): 907-928.

Gruber, T. 2006. Where the social web meets the semantic web, en I. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold y L. Aroyo (Eds.), *The Semantic Web - ISWC 2006*, Proceedings of the 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9 (Keynote Abstract), *Lecture Notes in Computer Science*, volumen 4273, pág. 994. Springer.

Gruber, T. 2007. Ontologies, Web 2.0 and Beyond. Presentación en *Ontolog Ontology Summit*, 24 de abril de 2007. Disponible en línea. Último acceso 01/05/2012. [<http://tomgruber.org/writing/ontolog-social-web-keynote.pdf>].

Gruber, T. 2008. Ontology. Entry in the *Encyclopedia of Database Systems*, Ling Liu y M. Tamer Özsu (Eds.), Springer-Verlag. Disponible en línea. Último acceso 01/05/2012. [<http://tomgruber.org/writing/ontology-definition-2007.htm>].

Gruber, T. y G. Olsen. 1994. An Ontology for Engineering Mathematics, en J. Doyle, P. Torasso, y E. Sandewall (Eds.), *Fourth International Conference on Principles of Knowledge Representation and Reasoning*. Bonn, Alemania, Morgan Kaufmann Publishers, San Francisco, California, págs. 258-269.

- Grüninger, M. y M. Fox. 1995. Methodology for the design and evaluation of ontologies, en D. Skuce (ed) *IJCAI 95 Workshop on Basic Ontological Issues in Knowledge Sharing*, págs. 6.1-6.10.
- Guarino, N. 1995. Formal Ontology, Conceptual Analysis and Knowledge Representation, en *International Journal of Human-Computer Studies*, 43 (5-6): 625–640.
- Guarino, N. 1998. Formal Ontology and Information Systems. *Proceedings of FOIS'98*, Trento, Italia, 6-8 junio 1998, págs. 3-15. Amsterdam, IOS Press.
- Guarino, N. y L. Boldrin. 1993. Ontological requirements for knowledge sharing. Comunicación presentada en el *IJCAI workshop for knowledge sharing and information interchange*, Chambéry, Francia.
- Guarino, N. y P. Giaretta. 1995. Ontologies and Knowledge Bases: Towards a Terminological Clarification, en N. Mars (Ed.). *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS '95)*. University of Twente, Enschede, The Netherlands. IOS Press, Amsterdam, The Netherlands, págs. 25-31.
- Guarino, N., D. Oberle y S. Staab. 2009. What is an Ontology? en S. Staab y R. Studer (Eds.), *Handbook on Ontologies*, International Handbooks on Information Systems, págs. 1-17. Springer-Verlag Berlin Heidelberg.
- Guarino, N. y C. Welty. 2000a. A formal ontology of properties, en R. Dieng, O. Corby (Eds.), *Proceedings of the 12<sup>th</sup> International Conference in Knowledge Engineering and Knowledge Management (EKAW '00)*. Juan-Les-Pins, Francia. *Lecture Notes in Artificial Intelligence LNAI 1937*. Springer-Verlag. Berlín, Alemania, págs. 97-112.
- Guarino, N. y C. Welty. 2000b. Ontological analysis of taxonomic relationships, en *Proceedings of the 19<sup>th</sup> International Conference on Conceptual Modeling*, Berlín-New York: Springer, págs. 210-224.
- Guarino, N. y C. Welty. 2002. Evaluating ontological decisions with OntoClean, *Communications of the ACM* 45(2): 61-65.

- Halliday, M. 1985. *An introduction to Functional Grammar*. Londres: Edward Arnold.
- Hoekstra, R., J. Breuker, M. D. Bello y A. Boer. 2007. The LKIF Core ontology of basic legal concepts, en P. Casanovas, M. A. Biasiotti, E. Francesconi, y M. Sagri (Eds.), *Proceedings of the Workshop on Legal Ontologies and Artificial Intelligence Techniques (LOAIT 2007)* en the *International Conference on AI and Law (ICAIL '07)*. Stanford, EEUU. 4 de junio, págs. 43–63.
- Hoekstra, R., J. Breuker, M. D. Bello, y A. Boer. 2009. LKIF Core: Principled ontology development for the legal domain, en J. Breuker, P. Casanovas, M. Klein y E. Francesconi (Eds.), *Law, Ontologies and the Semantic Web. Channelling the Legal Information Flood*, volume 188 of *Frontiers in Artificial Intelligence and Applications*. IOS PRESS, págs. 21-52.
- Horrocks, I., D. Fensel, F. van Harmelen, S. Decker, M. Erdmann y M. Klein. 2000. OIL in a nutshell, en R. Dieng, O. Corby (Eds.), *12<sup>th</sup> International Conference in Knowledge Engineering and Knowledge Management (EKAW '00)*. Juan-Les-Pins, Francia. *Lecture Notes in Artificial Intelligence LNAI 193*). Springer-Verlag. Berlín, Alemania, págs. 1-16.
- Horrocks, I. y F. van Harmelen. 2001. *Reference Description of the DAML+OIL (March 2001) Ontology Markup Language. Technical Report*. Disponible en línea. Último acceso 01/05/2012. [<http://www.daml.org/2001/03/reference.html>].
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- IEEE. 1990. *IEEE Standard Glossary of Software Engineering Terminology*. IEEE Computer Society. New York. IEEE Std 610.12-1990. Disponible en línea. Último acceso 01/05/2012. [<http://www.idi.ntnu.no/grupper/su/publ/ese/ieee-se-glossary-610.12-1990.pdf>].
- IEEE. 1997. *IEEE Standard for Developing Software Life Cycle Processes*. IEEE Computer Society. New York. IEEE Std 1074-1997. Disponible en línea. Último acceso 01/05/2012. [[http://arantxa.ii.uam.es/~sacuna/is1/normas/IEEE\\_Std\\_1074\\_1997.pdf](http://arantxa.ii.uam.es/~sacuna/is1/normas/IEEE_Std_1074_1997.pdf)].

- Jiménez, R. y M. Cabello. 2010. An account of selection restrictions within a conceptual framework: Its relevance for Role and Reference Grammar. Comunicación presentada en *The 2009 International Conference in Role and Reference Grammar*, págs. 131-146. University of California, Berkeley, U.S.A.
- Jiménez, R. y A. Luzondo. 2011. Building ontological meaning in a lexico-conceptual knowledge base. *Onomázein* 23: 11-40.
- Jiménez, R., A. Luzondo y M. Cabello. 2011. FunGramKB y la organización ontológica. *Anglogermánica Online* 2011: 16-36.
- KACTUS. 1996. *The Kactus Booklet version 1.0*. Esprit Project 8145 KACTUS. Disponible en línea. Último acceso 01/05/2012. [<http://hcs.science.uva.nl/projects/NewKACTUS/Reports.html>].
- Karp, P., V. Chaudhri y J. Thomere. 1999. *XOL: An XML-Based Ontology Exchange Language. Version 0.3. Technical Report*. Disponible en línea. Último acceso 01/05/2012. [<http://www.ai.sri.com/~pkarp/xol/xol.html>].
- Kelsen, H. 1991. *General Theory of Norms*. Clarendon Press: Oxford.
- Kietz, J.U., A. Maedche y R. Volz. 2000 A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet, en N. Aussenac-Gilles, B. Biébow y S. Szulman (Eds.) *EKAW'00 Workshop on Ontologies and Texts*. Juan-Les-Pins, Francia. *CEUR Workshop Proceedings* 51: 4.1-4.14. Amsterdam, Holanda.
- Kifer, M., G. Lausen y J. Wu. 1995. Logical Foundations of Object-Oriented and Frame-Based Languages. *Journal of the ACM* 42(4): 741-843.
- Klein, M., W. van Steenbergen, E. Bruijntjes, F. van Harmelen, y A. R. Lodder. 2006. Supporting layman in finding relevant court decisions in the best-project, en NWO TOKEN Symposium 2006, Leiden, 31 de marzo.
- Klinker, G., C. Bholá, G Dallemagne, D Marques, J McDermott. 1991. Usable and reusable programming constructs. *Knowledge Acquisition* 3: 117-136.
- Knight, K. y S. Luck. 1994. Building a Large-Scale Knowledge Base for Machine Translation, en B. Hayes, R. Korf (Eds.), *12<sup>th</sup> National Conference on Artificial*

- Intelligence (AAAI '94)*. Seattle Washington. AAAI Press, Menlo Park, California, 1:773-778.
- Knublauch, H., R. Ferguson, N. Noy y M. Musen. 2004. The Protegé OWL Plugin: An Open Development Environment for Semantic Web Applications, en S. McIlraith y D. Plexousakis (Eds.), *Proceedings of the 3<sup>rd</sup> International Semantic Web Conference (ISWC '04) Lecture Notes in Computer Science LNC 3298*, págs. 229-243. Springer-Verlag, Berlín, Alemania, Hiroshima, Japón.
- Koenig, J., G. Mauner y B. Bienvenue. 2003. Arguments for adjuncts. *Cognition* 89: 67-103.
- Koester, A. 2010. Building small specialised corpora, en A. O’Keeffe y M. McCarthy (Eds.). *The Routledge Handbook of Corpus Linguistics*: 66-79. US / Canada: Routledge.
- Lame, G. 2000. Knowledge Acquisition from Texts Towards an Ontology of French Law, en N. Aussenac-Gilles y S. Szulman (Eds.), *Proceedings of the Workshop on Ontologies and Text in the 12th European Workshop on Knowledge Acquisition, Modeling and Management (EKAW '00)*. Juan-Les-Pins, France, page (Application).
- Lame, G. 2001. Constructing an IR-oriented legal ontology, en R. Winkels (Ed.), *Proceedings of the Second International Workshop on Legal Ontologies (LEGONT) in JURIX 2001*. Amsterdam (Holanda) págs. 31–36.
- Lame, G. 2002. Construction d’ontologie a partie de textes. Une ontologie du droit dedieè à la recherche d’information sur le Web. Tesis Doctoral, L’Ecole des Mines de Paris.
- Lame, G. 2005. Using NLP techniques to identify legal ontology components: concepts and relations, en Benjamins, V. R., Casanovas, P., Breuker, J., and Gangemi, A. (Eds.) *Law and the Semantic Web. Legal Ontologies, Methodologies, Legal Information retrieval, and Applications*, volume 3369 of *Lecture Notes in Computer Science*, págs. 169–184. Springer-Verlag Berlin Heidelberg.

- Lame, G. y Desprès, S. 2005. Updating ontologies in the legal domain, en *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL 2005)*, 6-11 de junio, Universidad de Bolonia (Italia), volumen 155-162, New York. Association for Computing Machinery.
- Lassila, O. y D. McGuinness. 2001. *The Role of Frame-Based Representation on the Semantic Web. Technical Report KSL-01-02. Knowledge Systems Laboratory. Stanford University. Stanford, California. Disponible en línea. Último acceso 01/05/2012.*  
[<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.104.3946>].
- Laurence, S. y E. Margolis. 1999. Concepts and cognitive science en E. Margolis y S. Laurence (Eds.), *Concepts: Core readings*, págs. 3-81. Cambridge (Mass.): Bradford Books/MIT Press.
- Law, J. y E. Martin. (Eds.). 2009. *Oxford Dictionary of Law*. 9ª edición. Oxford: Oxford University Press.
- Leary, R., W. Vandenberghe, y J. Zeleznikow. 2003. Towards a financial fraud ontology; a legal modelling approach, en *Proceedings of the ICAIL 2003 Workshop on Legal Ontologies and Web Based Legal Information Management*.
- Lehmann, J., J. Breuker y B. Brouwer. 2004. CAUSATIOnt in AI and Law. *Artificial Intelligence and Law*, 12(4): 279–315.
- Lenat, D. B. y R. V. Guha. 1990. *Building large Knowledge-based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, Boston, Massachusetts.
- Lindberg, D. A. B., B. L. Humphreys y A. T. Mc Cray. 1993. The Unified Medical Language System. *Methods of Information in Medicine*, 32: 281 – 291.
- Luke, S. y J. Heflin. 2000. *SHOE 1.0. Proposed Specification. Technical Report. Parallel Understanding Systems Group. Department of Computer Science. University of Maryland. Disponible en línea. Último acceso 01/05/2012.*  
[<http://www.cs.umd.edu/projects/plus/SHOE/spec.html>].
- MacGregor, R. 1991. Inside the LOOM classifier. *SIGART bulletin*. 2 (3): 70-76.

- Maedche, A., B. Motik, L. Stojanovic, R. Studer y R. Volz. 2003. Ontologies for Enterprise Knowledge Management. *IEEE Intelligent Systems*, 18 (2), págs. 26-33.
- Mahesh, K. 1996. Ontology development for machine translation: ideology and methodology. Technical report MCCS-96-292. Computing Research Laboratory, New Mexico State University, Las Cruces.
- Mahesh, K. y S. Nirenburgh. 1995. Semantic classification for practical natural language processing, en R.P. Schwartz B. H. Kwasnik, C. Beghtol, P. J. Smith y E. Jacobs (Eds.), *6<sup>th</sup> ASIS SIG/CR Classification Research Workshop: An Interdisciplinary Meeting*. Chicago, Illinois, págs. 79-94.
- Mairal, R. y C. Perriñán. 2009. The anatomy of the lexicon component within the framework of a conceptual knowledge base. *Revista Española de Lingüística Aplicada* 22: 217-244.
- Mairal, R. y C. Perriñán. 2010. Role and Reference Grammar and Ontological Engineering, en J. Cifuentes, A. Gómez, A. Lillo, J. Mateo y F. Yus (Eds.), *Los caminos de la lengua. Estudios en homenaje a Enrique Alcaraz Varó*, págs. 649-665. Alicante: Universidad de Alicante.
- Mairal, R., C. Perriñán y E. Samaniego 2011. Using ontologies for terminological knowledge representation: a preliminary discussion, en N. Talaván, E. Martín y F. Palazón (Eds.), *Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10* . págs. 267-280. UNED: Madrid.
- Masolo, C., S. Borgo, A. Gangemi, N. Guarino, A. Oltramari y L. Schneider. 2002. *The WonderWeb Library of Foundational Ontologies and the DOLCE Ontology. Wonderweb Deliverable D17*. Disponible en línea. Último acceso 01/05/2012.  
[\[http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf\]](http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf).
- McCarthy, J. 1980. Circumscription -- A Form of Non-Monotonic Reasoning. *Artificial Intelligence*, 5 (13): 27-39.

- McCarty, L. T. 1989. A language for legal discourse, I. Basic Features. *Proceedings of the Second International Conference on Artificial Intelligence and Law*. Vancouver, Canadá, págs. 180-189.
- McGuinness, D. L. 2003. Ontologies come of age, en D. Fensel, J. Hendler, H. Lieberman, y W. Wahlster (Eds.), *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, págs. 171–194. The MIT Press.
- McGuinness, D. y F. van Harmelen. 2004. *OWL Web Ontology Language. W3C Recommendation*. 10 February 2004. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/TR/owl-features/>].
- Meyer, Ch. 2002. *English Corpus Linguistics*. Cambridge University Press.
- Miller, G. 1995. Wordnet: a lexical database for English. *Communications of the ACM* 38 (11): 39-41.
- Miller, G., R. Beckwith, C. Fellbaum, D. Gross y K. Miller. 1990. *International Journal of Lexicography* 3(4): 235-244.
- Mizoguchi, R., J. Vanwelkenhuysen y M. Ikeda. 1995. Task ontology for reuse of problem solving knowledge, en N. Mars (Ed.), *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing (KBKS'95)*, págs. 46–59. IOS Press, Amsterdam, Holanda.
- Moens, M. 2001. Innovative techniques for legal text retrieval. *Artificial Intelligence and Law* 9: 29–57.
- Mommers, L. 1998. On the ontological status and representation of legal concepts, en J. Hage, T. Bench-Capon, A. Koers, C.de Vey Mestdagh. y C. Grütters (Eds.), *Proceedings of the Eleventh Conference of Legal Knowledge-based Systems (JURIX'98)*, págs. 45–58. Gerard Noodt Instituut (GNI), Nijmegen.
- Mommers, L. 2001. A knowledge-based ontology of the legal domain, en R. Winkels (Ed.), R., *Proceedings of the Second International Workshop on Legal Ontologies (LEGONT) en JURIX 2001*, págs. 1–7. Amsterdam (Netherlands).
- Mommers, L. 2002. *Applied Legal Epistemology. Building a Knowledge-Based Ontology of the Legal Domain*. PhD thesis, Leiden University.

- Moreno, A. 2000. OntoTerm: un sistema abierto de representación conceptual. *Actas del XVI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*. Octubre de 2000, Vigo, España.
- Moreno, A. y C. Pérez. 2000. Reusing the Mikrokosmos ontology for concept-based multilingual terminology databases. *Proceedings of the 2<sup>nd</sup> International Conference on Language Resources and Evaluation (LREC 2000)*. Junio de 2000, Atenas, Grecia, págs. 1061-1067.
- Motta, E. 1999. *Reusable components for Knowledge Modelling: Principles and Case Studies in Parametric Design*. IOS Press, Amsterdam, Holanda.
- Neches, R., R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, y W. Swartout. 1991. Enabling technology for knowledge sharing. *AI Magazine*, 12 (3): 36-56.
- Niles, I. y A. Pease 2001a. Origins of the Standard Upper Merged Ontology: a proposal for the IEEE Standard Upper Ontology, en *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology*. Seattle.
- Niles, I. y A. Pease 2001b. Towards a Standard Upper Ontology, en *Proceedings of the Second International Conference on Formal Ontology in Information Systems (FOIS-2001)*. Ogunquit, Maine.
- Noy, N., R. Ferguson y M. Musen. 2000. The Knowledge Model of Protégé-2000: Combining interoperability and flexibility, en R. Dieng y O. Corby (Eds.), *Proceedings of the 12<sup>th</sup> International Conference in Knowledge Engineering and Knowledge Management (EKAW'00) Lecture Notes in Artificial Intelligence LNAI 1937*, págs. 17-32. Springer-Verlag. Berlín, Alemania, Juan-Les-Pins, Francia.
- Noy, N. y M. Musen. 2000. PROMPT: Algorithm and tool for automated ontology merging and alignment, en P. Rosenbloom, H. Kautz, B. Porter, R. Dechter, R. Sutton y V. Mittal (Eds.), *Proceedings of the 17<sup>th</sup> National Conference on Artificial Intelligence (AAAI '00)*, págs. 450-455. Austin, Tejas.

- Ontolingua. 1997. *Ontolingua System Reference Manual*. Disponible en línea. Último acceso 01/05/2012. [<http://www-ksl-svc.stanford.edu:5915/doc/frame-editor/index.html>].
- Pease, R. y T. Carrico. 1997. *JTF-ATD Core Plan Representation: A Progress Report*, en A. Farquhar, M. Gruninger, A. Gómez, M. Uschold y P. van der Vet (Eds.), *AAAI' 97 Spring Symposium on Ontological Engineering*. Stanford University, California, págs. 95-99.
- Pease, R. e I. Niles. 2002. IEEE Standard Upper Ontology: A Progress Report. *Knowledge Engineering Review*. Special Issue on Ontologies and Agents, 17 (1): 65-70.
- Periñán, C. y F. Arcas. 2004. Meaning postulates in a lexico-conceptual knowledge base, 15th International Workshop on Databases and Expert Systems Applications, IEEE, Los Alamitos (California): 38-42.
- Periñán, C. y F. Arcas. 2005. Microconceptual-Knowledge Spreading in FunGramKB. *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*, ACTA Press, Anaheim-Calgary-Zurich, págs. 239-244.
- Periñán, C. y F. Arcas. 2007. Deep semantics in an NLP knowledge base, *12th Conference of the Spanish Association for Artificial Intelligence*, Universidad de Salamanca: 279-288.
- Periñán, C. y F. Arcas. 2008. A cognitive approach to qualities for NLP. *Procesamiento del Lenguaje Natural* 41: 137-144.
- Periñán, C. y R. Mairal. 2009. Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural* 43: 265-273.
- Periñán, C. y F. Arcas. 2010a. The architecture of FunGramKB. *7th International Conference on Language Resources and Evaluation, 17-23 mayo 2010, Valeta (Malta). Proceedings of the Seventh International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA), págs. 2667-2674. (ISBN: 2-9517408-6-7).

- Periñán, C. y F. Arcas. 2010b. Ontological Commitments in FunGramKB. *Procesamiento del Lenguaje Natural* 44: 27-34.
- Periñán, C. y F. Arcas. 2011. Introducción a FunGramKB. *Anglogermánica Online* 2011: 1-15.
- Periñán, C. y R. Mairal. 2010a. Enhancing UniArab with FunGramKB. *Procesamiento del Lenguaje Natural* 44: 19-26.
- Periñán, C. y R. Mairal. 2010b. La gramática de COREL: un lenguaje de representación conceptual. *Onomázein* 21: 11-45.
- Periñán, C. y R. Mairal. 2010c. Tutorial de FunGramKB. Disponible en línea. Último acceso 01/05/2012. [<http://www.fungramkb.com/resources/papers/tutorial.pdf>].
- Periñán, C. y R. Mairal. 2011. The COHERENT methodology in FunGramKB. *Onomázein* 24: 13-33.
- Peters, W., M. Sagri y D. Tiscornia. 2007. The structuring of legal knowledge in LOIS. *Artificial Intelligence and Law*, 15 (2): 117–135.
- Procter, P. (Ed.). 1978. *Longman Dictionary of Contemporary English*. Harlow (Essex): Longman.
- Protégé. 2000. The Protégé Project. Disponible en línea. Último acceso 01/05/2012. [<http://protege.stanford.edu>].
- Ragett D., A. Le Hors e I. Jacobs. 1999. *HTML 4.01. Specification. W3C Recommendation*. [<http://www.w3.org/TR/html401>].
- Rector, A, S. Bechhofer, C. Goble, I. Horrocks, W. Nowlan y W. Solomon. 1997. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine* 9: 139-171.
- Rector, A., W. Nowlan, S. Kay, C. Goble y T. Howkins. 1993. A framework for modelling the electronic medical record. *Methods of Information in Medicine*, 32: 109 – 119.

- Rector, A. W. D. Solomon, W. A. Nowlan y T. W. Rush. 1995. A Terminology Server for Medical Language and Medical Information Systems. *Methods of Information in Medicine*, 34: 147-157.
- Reppen, R. 2010. Building a corpus: What are the key considerations? en A. O’Keeffe y M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics*, págs. 66-79. US/Canada: Routledge.
- Rubino, R., A. Rotolo y G. Sartor. 2007. An owl ontology of norms and normative judgements, en C. Biagioli, y G. Sartor (Eds.), *Proceedings of the V Legislative XML Workshop*. Florencia, Italia, 14-16 junio, 2006, págs. 173–188. European Press Academic Publishing.
- Rusell, S. y P. Norvig. 1995. *Artificial Intelligence: A Modern Approach*. Englewood Cliffs, New Jersey.
- Sager, J. 1993. *Curso práctico sobre el procesamiento de la terminología*. Fundación Germán Sánchez Ruipérez. Madrid: Pirámide.
- Sagri, M., D. Tiscornia y F. Bertagna. 2004. Jur-wordnet, en P. Sojka, K. Pala, P. Smrz, C. Fellbaum, y P. Vossen (Eds.), *GWC 2004 Proceedings*, págs. 305–310. Masaryk University, Brno.
- Saias, J.y P. Quaresma, P. 2003. Using NLP techniques to create legal ontologies in a logic programming based web information retrieval system, en *Proceedings of the ICAIL 2003 Workshop on Legal Ontologies and Web Based Legal Information Management*.
- Sánchez, I. 2008. *La criminalidad organizada. Aspectos penales, procesales, administrativos y policiales*. Madrid: Dykinson.
- Schreiber, G., H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbolt, W. van de Velde y B. Wielinga. 1999. *Knowledge engineering and management. The CommonKADS Methodology*. MIT Press. Cambridge, Massachusetts.
- Schreiber, A., B. Wielinga y W. Jansweijer. 1995. The KACTUS View on the “O” World, en D. Skuce (Ed.), *IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing*, págs. 15.1-15.10.

- Shaheed, J., A. Yip y J. Cunningham. 2005. A Top-Level Language-Biased Legal Ontology, en Jos Lehmann, et al. (Eds), *LOAIT - Legal Ontologies and Artificial Intelligence Techniques*. Bolonia, *IAAIL Workshop Series No 4*, págs. 13-24.
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Smith, B. 2003. Ontology, en Floridi, L., *Blackwell Guide to the Philosophy of Computing and Information*, págs. 155–166. Oxford: Blackwell.
- Smith, B. y C. Welty. 2001. Ontology: Towards a new synthesis. *Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS2001)*. ACM Press.
- Sowa, J. F. 1999. *Knowledge representation: Logical, Philosophical, and Computational Foundations*. Brooks Cole Publishing Co., Pacific Grove, California.
- Staab, S., M. Erdmann, A. Maedche y S. Decker. 2000. An Extensible Approach for Modeling Ontologies in RDF(S), en P. Costantopoulos (Ed.), *ECDL 2000 Workshop on the Semantic Web*. Lisboa, Portugal, págs. 11-22.
- Staab, S., H. P. Schnurr, R. Studer y Y. Sure. 2001. Knowledge Processes and Ontologies. *IEEE Intelligent Systems* 16 (1): 26-34.
- Stamper, R. 1991. The role of semantics in legal expert reasoning and legal systems. *Ratio Juris*, Vol. 2 No. 4, págs. 219-44.
- Stamper, R 1996. Signs, information, norms and systems, en B. Holmqvist y P. Andersen (Eds.), *Signs of Work*. De Gruyter: Berlín.
- Standard Upper Ontology Working Group (SUO) IEEE P1600.1. Disponible en línea. Último acceso 01/05/2012. [<http://suo.ieee.org>].
- Steels, L. 1993. Corporate Knowledge Management, en Cuenca (Ed.) *Knowledge oriented software design: extended papers from the IFIP TC 12 Workshop on Artificial Intelligence from the Information Processing Perspective, AIFIPP*. Madrid, España. Publicado en 1993 en *IFIP Transactions: Computer Science and Technology*. Amsterdam, Holanda, págs. 91-116.

- Studer, R., V. R. Benjamins y D. Fensel. 1998. Knowledge engineering: Principles and methods. *Data Knowledge Engineering*, 25 (1-2): 161–197.
- Suggested Upper Merged Ontology (SUMO). Disponible en línea. Último acceso 01/05/2012. [<http://suo.ieee.org/SUO/SUMO/index.html>].
- Sure, Y., M. Erdmann, J. Angele, S. Staab, R. Studer y D. Wenke. 2002. OntoEdit: Collaborative Ontology Engineering for the Semantic Web, en I. Horrocks y J. A. Hendler (Eds.), *First International Semantic Web Conference (ISWC '02)*. Cerdeña, Italia. *Lecture Notes in Computer Science LNCS 2342*. Springer-Verlag, Berlín, Alemania, págs. 221-235.
- Sure, Y., S. Staab y J. Angele. 2002. OntoEdit: Guiding Ontology Development by Methodology and Inferencing, en R. Meersman y Z. Tari (Eds.), *Confederated International Conferences CoopIS, DOA y ODBASE 2002*. Universidad de California, Irvine. *Lecture Notes in Computer Science LNCS 2519*. Springer-Verlag, Berlín, Alemania, págs. 1205-1222.
- Swartout, P., R. Patil, K. Knight y T. Russ. 1997. Toward distributed Use of Large-Scale Ontologies, en A. Farquhar, M. Gruinger, A. Gómez, M. Uschold y P. van der Vet (Eds.) *AAAI '97 Spring Symposium on Ontological Engineering*. Stanford University, California, págs. 138-148.
- Temmerman, R. 2000. *Towards new ways of terminology description: the sociocognitive-approach*. Amsterdam: John Benjamins.
- The Legal Knowledge Interchange Format (LKIF). Deliverable 4.1*. Disponible en línea. Último acceso 01/05/2012. [<http://carneades.github.com/resources/Home/LKIF-Specification.pdf>].
- Tijerino, Y. A. y R. Mizoguchi. 1993. MULTIS II: enabling end-users to design problem-solving engines via two-level task ontologies, en N. Aussenac, G. Boy B. Gaines, M. Linster, J. G. Ganascia, Y. Kodratoff (Eds.), *7<sup>th</sup> European Workshop on Knowledge Acquisition for Knowledge-Based Systems (EKAW '93)*. Toulouse y Caylus, Francia. *Lecture Notes in Computer Science LNCS 723*. Springer-Verlag, Berlín, Alemania, págs. 340-359.

- Tiscornia, D. 2007. The LOIS project: Lexical ontologies for legal information sharing, en C. Biagioli, E. Francesconi y G. Sartor (Eds.), *Proceedings of the V Legislative XML Workshop*, págs. 189–204. European Press Academic Publishing.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Benjamins.
- Tu, S. W., H. Eriksson, J. Gennari, Y. Shahar y M. Musen. 1995. Ontology-based configuration of problem-solving methods and generation of knowledge acquisition tools: the application of Protégé-II to protocol-based decision support. *Artificial Intelligence in Medicine* 7: 257 – 289.
- Ureña, P., A. Alameda y A. Felices. 2011. Towards a specialised corpus of organized crime and terrorism. María Luisa Carrió y otros (Eds.), *La investigación y la enseñanza aplicadas a las lenguas de especialidad y a la tecnología*. Universitat Politècnica de Valencia, Valencia, págs. 301-306.
- Uschold, M. 1996. Building ontologies: Towards a unified methodology, en *Expert Systems '96*, Cambridge, UK.
- Uschold, M. 2003. Where are the semantics in the semantic web? *AI Magazine*, 24 (3): 25–36.
- Uschold, M. y M. Grüninger. 1996. Ontologies: Principles, methods, and applications. *Knowledge Engineering Review*, 11 (2): 93–155.
- Uschold, M. y R. Jasper. 1999. A Framework for Understanding and Classifying Ontology Applications, en V. Benjamins (Ed.), *IJCAI '99 Workshop on Ontology and Problem Solving Methods: Lessons Learned and Future Trends*. Stocolmo, Suecia. CEUR Workshop Proceedings 18: 11.1-11.12. Amsterdam, Holanda.
- Uschold, M. y M. King. 1995. Towards a Methodology for Building Ontologies, en D. Suce (Ed.), *IJCAI '95 Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal, Canadá, págs. 6.1-6.10.
- Uschold, M., M. King, S. Moralee y Y. Zorgios. 1998. The Enterprise Ontology. *The Knowledge Engineering Review* 13 (1): 31-89.

- Valente, A. 1995. *A modelling approach to Legal Knowledge Engineering*. IOS Press: Amsterdam y Tokio.
- Valente, A. 2005. Types and Roles of Legal Ontologies, en V. Benjamins, P. Casanovas, J. Breuker y A. Gangemi (Eds.), *Law and the Semantic Web: Legal Ontologies, Methodologies, Legal Information Retrieval and Applications. Lecture Notes in Artificial Intelligence 3369*: 65-76. Springer, Verlag, Berlin, Heidelberg.
- Valente, A. y J. Breuker. 1994a. A Functional Ontology of Law, en G. Bargellini y S. Binazzi, S. (Eds.), *Towards a Global Expert System in Law*. Cedam Publishers, Padua, Italia.
- Valente, A. y J. Breuker. 1994b. Ontologies: the missing link between Legal Theory and AI & Law, en A. Soeteman (Ed.), *JURIX 94*, LKoninklijke Vermande, Lelystad, 1994, 138-150.
- Valente, A. y J. Breuker. 1996. Towards principled core ontologies, en *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW'96)* Banff, Alberta, Canada, Banff, Alberta, Canada.
- Van der Spek, R., E. Kruizinga y G. van Heijst, G. 1996. Organizing Corporate Memories, en B. Gaines y M. Musen (Eds.), *Tenth Knowledge Acquisition for Knowledge-Based Systems Workshop (KAW '96)*. Banff, Canadá, págs. 42.1-42.17.
- Van Harmelen, F., P. Patel-Schneider e I. Horrocks. 2001. *Reference description of the DAML+OIL (March 2001) ontology mark-up language*. Disponible en línea. Último acceso 01/05/2012. [<http://www.daml.org/2001/03/reference.html>].
- Van Heijst, G., A. Schreiber y W. Wielinga. 1997. *Using explicit ontologies in KBS development*. International Journal of Human-Computer Studies. 45: 183-292.
- Van Kralingen, R. 1995. *Frame-based Conceptual Models of Statute Law*, Computer/Law Series, 16, Kluwer Law International: La Haya.

- Van Kralingen, R. 1997. A conceptual frame-based ontology for the law, en P. Visser, y R. Winkels (Eds.), *Proceedings of the First International Workshop on Legal Ontologies (LEGONT'97)*, Melbourne, Australia, págs. 15–22.
- Van Kralingen, R., E. Oskamp y E. Reuring. 1993. Norm frames in the representation of laws, en J. Svensson, J. Wassink, y B. van Buggenhout. (Eds.), *Legal knowledge based systems JURIX 93: Intelligent Tools for Drafting Legislation, Computer-Supported Comparison of Law*, págs. 11–22, Lelystad. Koninklijke Vermande.
- Van Valin, R. 2005. *Exploring the Syntax-Semantic Interface*. Cambridge: Cambridge University Press.
- Van Valin, R. y R. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.
- Viegas, E., K. Mahesh, S. Nirenburg y S. Beale. 1999. Semantics in Action, en Saint-Dizier (Ed.), *Predicative Forms in Natural Language and Lexical Knowledge Bases*. Dordrecht: Kluwer, págs. 171-204.
- Visser, P. 1995. Framer: Source code of a legal knowledge system that performs assessment and planning. *Reports on Technical Research in Law Vol. 2, No. 1*, University of Leiden, Leiden, Holanda.
- Visser, P. y T. Bench-Capon. 1996. The formal specification of a legal ontology, en van R. van Kralingen, H. van den Herik, J. Prins, M. Sergot y J. Zeleznikow (Eds.), *Proceedings of JURIX'96, Legal Knowledge Based Systems (Foundations of Legal Knowledge Systems)*, págs. 15–24. Tilburg University Press.
- Visser, P. y T. Bench-Capon. 1998. A Comparison of Four Ontologies for the Design of Legal Knowledge Systems, *Artificial Intelligence and Law* 6: 27-57.
- Visser, P. y T. Bench-Capon. 1999. Ontologies in the Design of Legal Knowledge Systems: Towards a Library of Legal Domain Ontologies. In *Proceedings of the Conference on Applied Ontology*. Buffalo University, NY, págs. 76–85.

- Vossen, P. (Ed.). 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, Dordrecht, Holanda. Disponible en línea. Último acceso 01/05/2012. [<http://acl.ldc.upenn.edu/J/J99/J99-4008.pdf>].
- Vossen, P. (Ed.). 1999. *EuroWordNet General Document. Version 3*. Disponible en línea. Último acceso 01/05/2012. [<http://vossen.info/docs/2002/EWNGeneral.pdf>].
- Waterman, D. A. 1986. *A Guide to Expert Systems*. Addison-Wesley, Boston, Massachussetts.
- Welty, C. y N. Guarino. 2001. Supporting Ontological Analysis of Taxonomic Relationships. *Data and Knowledge Engineering* 39 (1): 51-74.
- Wielinga, B. y A. Schreiber. 1993. Reusable and shareable knowledge bases: A European perspective, en Fuchi, K. (Ed.), *Proceedings of the International Conference on Building and Sharing of Very Large-Scale Knowledge Bases (KB&KS'93)*, 1-4 diciembre. Tokio, págs. 103–115, Tokio, Japón.
- Wielinga, B., A. Schreiber y J. Breuker. 1992. KADS: a modeling approach to knowledge engineering. *Knowledge Acquisition* 4:1-162.
- Winkels, R., A. Boer y R. Hoekstra. 2002. Clime: Lessons learned in legal information serving, en *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI 2002)*, Lyon, France, págs. 230–234, Amsterdam, Holanda. IOS Press.
- Winkels, R., D. Bosscher, A. Boer., y R. Hoekstra. 2000. Extended conceptual retrieval, en J. Breuker, R. Leenes y R. Winkels (Eds.), *Legal Knowledge and Information Systems (JURIX-2000)*, págs. 85–98, Amsterdam. IOS Press.
- W3C Semantic Web. Disponible en línea. Último acceso 01/05/2012. [<http://www.w3.org/2001/sw/>].
- Zeleznikow, J y A. Stranieri. 2001. An ontology for the construction of legal decision support systems, en *Proceedings of the 2nd International Workshop of Legal Ontologies in Conjunction with the 14th Annual International Conference on Legal Knowledge and Information Systems (JURIX'01)*.

## **Diccionarios y recursos en línea**

CAMBRIDGE: *Cambridge Learner's Dictionary*.

[<http://dictionary.cambridge.org/dictionary/learner-english/>].

COLLINS: *Collins English Dictionary*. [<http://www.collinsdictionary.com/>].

GOOGLE: [<http://www.google.com>].

LONGMAN: *Longman Dictionary of Contemporary English*.

[<http://www.ldoceonline.com/>].

ONELOOK: *Onelook dictionary search*. [<http://www.onelook.com/>].

THE FREE DICTIONARY: *The free dictionary by Farlex*.

[<http://www.thefreedictionary.com/>].

WIKIPEDIA: *Wikipedia, the free encyclopedia*.

[[http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)].

WIKTIONARY: *Wiktionary, the free dictionary*.

[[http://en.wiktionary.org/wiki/Wiktionary:Main\\_Page](http://en.wiktionary.org/wiki/Wiktionary:Main_Page)].

WORDREFERENCE: *Wordreference English-Spanish Legal Dictionary*.

[<http://www.wordreference.com/enesl/>].

## Apéndices

### Apéndice 1. PS de 20 “conceptos espejo” de la ontología satélite del crimen organizado y el terrorismo (por orden alfabético).

#### 1. +AGGRESSION\_00

##### Conceptualización en la ontología nuclear:

+AGGRESSION\_00: spoken or physical behaviour which is threatening or involves harm to someone or something.

+(e1: +BE\_00 (x1: +AGGRESSION\_00)Theme (x2: +CRIME\_00)Referent)

+(e2: +ATTACK\_00 (x3)Theme (x4)Referent (f1: +VOICE\_00 ^ +HAND\_00)Instrument)

(f1: (e3: +DO\_00 (x5: +HARM\_00)Theme (x4)Referent))Purpose)

Unidades léxicas asociadas: *aggression, agresión.*

##### Conceptualización en la ontología satélite:

+AGGRESSION\_00: the breach of international law by a country.

+(e1: +BE\_00 (x1: +AGGRESSION\_00)Theme (x2: +CRIME\_00)Referent)

+(e2: +ATTACK\_00 (x3: +COUNTRY\_00)Theme (x4: +COUNTRY\_00)Referent (f1: +ARMY\_00)Instrument)

Unidades léxicas asociadas: *aggression, ataque, agresión.*

#### 2. +BLACK\_MARKET\_00

##### Conceptualización en la ontología nuclear:

+BLACK\_MARKET\_00: the system by which people illegally buy and sell foreign money, goods that are difficult to obtain etc.

+(e1: +BE\_00 (x1: +BLACK\_MARKET\_00)Theme (x2)Referent)

+(e2: +SELL\_00 (x3)Theme (x4: +MONEY\_00 ^ +GOODS\_00)Referent)

(f1: (e3: n+BE\_01 (x3)Theme (x5: +LEGAL\_00)Attribute)Manner))

Unidades léxicas asociadas: *black market, mercado negro.*

Conceptualización en la ontología satélite:

+BLACK\_MARKET\_00: An illegal market for goods that are controlled or prohibited by the government.

+(e1: +BE\_00 (x1: +BLACK\_MARKET\_00)Theme (x2)Referent)

+(e2: +SELL\_00 (x3)Theme (x4: +MONEY\_00 ^ +GOODS\_00)Referent)

(f1: (e3: n+BE\_01 (x3)Theme (x5: +LEGAL\_00)Attribute)Manner))

+(e4: +CONTROL\_00 (x6: +GOVERNMENT\_00)Theme (x4)Referent)

+(e5: +FORBID\_00 (x6)Theme (x4)Referent (x7)Goal)

Unidades léxicas asociadas: *black economy, black market, shadow economy, underground economy, underground market, mercado negro, economía sumergida.*

### **3. +CRIME\_00**

Conceptualización en la ontología nuclear:

+CRIME\_00: an act punishable by law; usually considered an evil act.

+(e1: +BE\_00 (x1: +CRIME\_00)Theme (x2: +OCCURRENCE\_00)Referent)

+(e2: n +BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)

Unidades léxicas asociadas: *crime, law-breaking, delito, ilegalidad, infracción, quebrantamiento, quebranto, transgresión, vulneración.*

Conceptualización en la ontología satélite:

+CRIME\_00: a violation of the law punishable by the state in criminal proceedings.

+(e1: +BE\_00 (x1: +CRIME\_00)Theme (x2: +OCCURRENCE\_00)Referent)

+(e2: n +BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)

+(e3: +BE\_00 (x1)Theme (x4: +FELONY\_00 ^ +MISDEMEANOR\_00)Referent)

\*(e4: +PUNISH\_00 (x5: +COUNTRY\_00)Theme (x6: +CRIMINAL\_00)Referent)

\*(e5: +BE\_02 (x6)Theme (x7: +CRIMINAL\_PROCEEDING\_00)Location)

(f1: +IN\_00)Position)

Unidades léxicas asociadas: *crime, criminal wrong, offence, delito, ofensa.*

#### 4. +DIRTY\_MONEY

Conceptualización en la ontología nuclear:

\$DIRTY\_MONEY\_00: money obtained illegally.

+(e1: +BE\_00 (x1: \$DIRTY\_MONEY\_00)Theme (x2: m+BANKNOTE\_00)Referent)

(f1: (e2: n+BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)Purpose)

Unidades léxicas asociadas: *dirty money, dinero negro.*

Conceptualización en la ontología satélite:

+DIRTY\_MONEY\_00: money illegally gained, transferred or used.

+(e1: +BE\_00 (x1: \$DIRTY\_MONEY\_00)Theme (x2: m+BANKNOTE\_00)Referent)

(f1: (e2: +OBTAIN\_00 (x3: +CRIMINAL\_00)Theme (x1)Referent) (e2: n+BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)Purpose)

(f2: (e3: +TRANSFER\_00 (x3)Agent (x1)Theme (x4)Origin (x5: n+LEGAL\_00)Goal)

(f3: (e4: +USE\_00 (x3)Theme (x1)Referent (e2: n+BE\_01 (x1)Theme (x3: +LEGAL\_00)Attribute)Manner)

Unidades léxicas asociadas: *dirty money, dinero negro.*

## **5. +FORGERY\_00**

### Conceptualización en la ontología nuclear:

+FORGERY\_00: the crime of making an illegal copy of documents or money.

+(e1: +BE\_00 (x1: +FORGERY\_00)Theme (x2: +CRIME\_00)Referent)

(f1: (e2: +CREATE\_00 (x3: +CRIMINAL\_00)Theme (x4: +DOCUMENT\_00 ^  
+MONEY\_00)Referent (f2)Instrument (e3: n+BE\_01 (x5)Theme (x4:  
+TRUE\_00)Attribute))Result)

Unidades léxicas asociadas: *forgery, falsificación.*

### Conceptualización en la ontología satélite:

+FORGERY\_00: the crime of making a false document or altering a real one to be used as if genuine.

+(e1: +BE\_00 (x1: +FORGERY\_00)Theme (x2: +CRIME\_00)Referent)

(f1: (e2: +CREATE\_00 (x3: +CRIMINAL\_00)Theme (x4: +DOCUMENT\_00)Referent  
(f2)Instrument (e3: n+BE\_01 (x5)Theme (x4: +TRUE\_00)Attribute))Result)

+(e4: +CHANGE\_00 (x3)Theme (x4)Referent (f3: (e5: +BECOME\_00 (x4)Theme (x5:  
+REAL\_00)Attribute))Result)

Unidades léxicas asociadas: *false making, forgery, falsificación documental.*

## **6. \$FREEZE\_01**

### Conceptualización en la ontología nuclear:

+FREEZE\_01: To cause to become fixed and unable to increase <to freeze interest rates> <to freeze prices>.

+(e1: +STOP\_00 (x1: +GOVERNMENT\_00)Theme (x2: s+PRICE\_00 ^  
s+INTEREST\_RATE\_00)Referent)

Unidades léxicas asociadas: *freeze, congelar.*

Conceptualización en la ontología satélite:

+FREEZE\_01: to make immobile by government mandate or banking action <to freeze assets>.

+(e1: +STOP\_00 (x1: +GOVERNMENT\_00)Theme (x2: s+PRICE\_00 ^ s+INTEREST\_RATE\_00)Referent)

(f1: (e2: n+PERMIT\_00 (x1)Theme (x3:(e3: +SELL\_00 (x4)Agent (x5: +ASSETS\_00)Theme (x6)Origin (x7)Goal))Referent(x8)Goal))Manner)

(f2: (e4: n+TRANSFER\_00 (x9: +BANK\_00)Agent (x10: +MONEY\_00)Theme (x11)Origin (x12)Goal)Manner)

Unidades léxicas asociadas: *freeze, congelar*.

**7. +LAUNDER\_MONEY\_00**Conceptualización en la ontología nuclear:

+LAUNDER\_MONEY\_00: to put illegal money into bank accounts or businesses.

+(e1: +PUT\_00 (x1)Agent (x2: +DIRTY\_MONEY\_00)Theme (x5: +BANK\_00 ^+BUSINESS\_00)Location (x3)Origin (x4)Goal (f1: (e2: +BECOME\_00 (x4)Theme (x6: +LEGAL\_00)Purpose)

Unidades léxicas asociadas: *launder money, blanquear dinero*.

Conceptualización en la ontología satélite:

+LAUNDER\_MONEY\_00: to transfer illegally obtained money through legitimate people or accounts so that its original source cannot be traced

+(e1: +CHANGE\_00 (x1)Theme (x2: +DIRTY\_MONEY\_00)Referent (f1: (e2: +BE\_01 (x2)Theme (x3: +LEGAL\_00)Attribute))Purpose)

+(e2: +PUT\_00 (x1)Agent (x2)Theme (x5: +BANK\_00 ^+BUSINESS\_00)Location (x3)Origin (x4)Goal (f1: (e3: +BECOME\_00 (x4)Theme (x6: +LEGAL\_00)Purpose)

Unidades léxicas asociadas: *launder money, blanquear dinero*.

## **8. +MASTERMIND\_00**

### Conceptualización en la ontología nuclear:

+MASTERMIND\_00: a person who plans every detail of a complicated event or activity and makes sure that it happens.

+(e1: +BE\_00 (x1: +MASTERMIND\_00)Theme (x2: +HUMAN\_00)Referent)

+(e2: +PLAN\_00 (x1)Theme (x3: +OCCURRENCE\_00)Referent)

+(e3: +BE\_01 (x3)Theme (x4: +DIFFICULT\_00)Attribute)

Unidades léxicas asociadas: *mastermind, cerebro*.

### Conceptualización en la ontología satélite:

+MASTERMIND\_00: someone who plans and organizes a criminal operation.

+(e1: +BE\_00 (x1: +MASTERMIND\_00)Theme (x2: +HUMAN\_00)Referent)

+(e2: +PLAN\_00 (x1)Theme (x3: +OCCURRENCE\_00)Referent)

+(e3: +BE\_00 (x3)Theme (x4: +CRIME\_00)Referent)

Unidades léxicas asociadas: *mastermind, cerebro de una banda criminal*.

## **9. +MONEY\_LAUNDERING\_00**

### Conceptualización en la ontología nuclear:

+MONEY\_LAUNDERING\_00: the act of putting illegal money into bank accounts or businesses.

+(e1: +BE\_00 (x1: +MONEY\_LAUNDERING\_00)Theme (x2)Referent)

+(e2: +PUT\_00 (x3)Agent (x4: +DIRTY\_MONEY\_00)Theme (x7: s+BANK\_00 ^ s+BUSINESS\_00)Location (x5) Origin (x6) Goal)

Unidades léxicas asociadas: *money laundering, blanqueo de dinero*.

Conceptualización en la ontología satélite:

+MONEY\_LAUNDERING\_00: the act of transferring illegally obtained money through legitimate people or accounts so that its original source cannot be traced.

+(e1: +BE\_00 (x1: +MONEY\_LAUNDERING\_00)Theme (x2)Referent)

+(e2: +PUT\_00 (x3)Agent (x4: +DIRTY\_MONEY\_00)Theme (x7: s+BANK\_00 ^ s+BUSINESS\_00)Location (x5) Origin (x6) Goal)

(f1: (e3: +SEEM\_00 (x4)Theme (x8: +LEGAL\_00)Attribute)Purpose) & (f2: (e4: n+DISCOVER\_00 (x9: +GOVERNMENT\_00)Theme (x10: +ORIGIN\_00)Referent)Purpose)

Unidades léxicas asociadas: *money laundering, blanqueo de dinero.*

**10. +ORGANIZED\_CRIME\_00**Conceptualización en la ontología nuclear:

+ORGANIZED\_CRIME\_00: criminal organization or the crimes committed by such organization.

+(e1: +BE\_00 (x1: +ORGANIZED\_CRIME\_00)Theme (x2: +CRIME\_00)Referent)

+(e2: +BE\_00 (x1)Theme (x3: +ORGANIZATION\_00)Referent)

+(e3: n +BE\_01 (x3) Theme (x4: +LEGAL\_00)Attribute)

Unidades léxicas asociadas: *criminal organization, gang, mafia, mob, organized crime, organized criminal group, ring, banda criminal, crimen organizado.*

Conceptualización en la ontología satélite:

+ORGANIZED\_CRIME\_00: criminal activities coordinated and controlled through a central syndicate.

+(e1: +BE\_00 (x1: +ORGANIZED\_CRIME\_00)Theme (x2: +CRIME\_00)Referent)

+(e2: +CONTROL\_00 (x3: +SYNDICATE\_00)Theme (x2)Referent)

Unidades léxicas asociadas: *organized crime, crimen organizado.*

## 11. \$PHISHING\_00

### Conceptualización en la ontología nuclear:

\$PHISHING\_00: the criminal activity of sending emails to trick someone into giving away information such as their credit card number and their password.

+(e1: +BE\_00 (x1: \$PHISHING\_00)Theme (x2: +CYBERCRIME\_00)

+(e2: +SEND\_00 (x3: +CRIMINAL\_00)Agent (x4: +E\_MAIL\_00)Theme (x5)Origin (x6)Goal (f1: +THE\_INTERNET\_00)Instrument)

+(e3: n+BE\_01 (x4)Theme (x7: +TRUE\_00)Attribute)

(f1: (e4: +DECEIVE\_00 (x3)Theme (x6)Referent) (e5: +GIVE\_00 (x6)Agent (x8: +CREDIT\_CARD\_NUMBER\_00 ^ +PASSWORD\_00)Theme (x6)Origin (x5)Goal (f2: +HAND\_00)Instrument)Purpose)

Unidades léxicas asociadas: *phishing*.

### Conceptualización en la ontología satélite:

\$PHISHING\_00: The sending of a fraudulent electronic communication to steal an identity or sell it to another party for illegal purposes.

+(e1: +BE\_00 (x1: \$PHISHING\_00)Theme (x2: +CYBERCRIME\_00)

+(e2: +SEND\_00 (x3: +CRIMINAL\_00)Agent (x4: +E\_MAIL\_00)Theme (x5)Referent (x6)Goal (f1: +THE\_INTERNET\_00)Instrument)

+(e3: n+BE\_01 (x4)Theme (x7: +TRUE\_00)Attribute)

(f2: (e4: +STEAL\_00 (x3)Theme (x8: +IDENTITY\_00)Referent)Purpose) (f3: +PLACE\_00)Origin ^ (f4: (e5: +SELL\_00 (x3)Agent (x8)Theme (x9)Origin (x10)Goal)Purpose)

Unidades léxicas asociadas: *phishing*.

**12. \$RACKETEERING\_00**Conceptualización en la ontología nuclear:

\$RACKETEERING\_00: making money from a dishonest or illegal activity.

+(e1: +BE\_00 (x1: \$RACKETEERING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +DECEIVE\_00 (x3: +CRIMINAL\_00) Theme (x4) Referent) (f1: (e3: +OBTAIN\_00 (x3) Theme (x5: +MONEY\_00)Referent)Purpose)

Unidades léxicas asociadas: *extortion, racketeering, extorsión.*

Conceptualización en la ontología satélite:

\$RACKETEERING\_00: an organized crime activity that involves extortion of money from businesses by intimidation or violence.

+(e1: +BE\_00 (x1: \$RACKETEERING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +DECEIVE\_00 (x3: +SYNDICATE\_00) Theme (x4: +BUSINESS\_00) Referent (f1: (e3 +OBTAIN\_00 (x3) Theme (x5: +MONEY\_00)Referent)Purpose)

+(e4: +USE\_00 (x3)Theme (x6: +VIOLENCE\_00)Referent (f1: x6)Instrument)

Unidades léxicas asociadas: *extortion, racketeering, extorsión, latrocinio.*

**13. \$SAFE\_HAVEN\_00**Conceptualización en la ontología nuclear:

SAFE\_HAVEN\_00: a refuge for people in danger.

+(e1: +BE\_00 (x1: \$SAFE\_HAVEN\_00)Theme (x2: +PLACE\_00)Referent)

+(e2: +BE\_02 (x1)Theme (x3: +REFUGE\_00)Location (f1: +HUMAN\_00)Beneficiary (f2: +DANGER\_00)Scene)

Unidades léxicas asociadas: *safe haven, refugio.*

Conceptualización en la ontología satélite:

\$SAFE\_HAVEN\_00: a refuge with military protection for a persecuted ethnic minority; a refuge for criminals or terrorists.

+(e1: +BE\_00 (x1: \$SAFE\_HAVEN\_00)Theme (x2: +PLACE\_00)Referent)

\*(e2: +BE\_02 (x1)Theme (x3: +REFUGE\_00)Location (f1: +ETHNIC\_MINORITY\_00)Beneficiary (f2: +ARMY\_00)Instrument)

\*(e3: +BE\_02 (x1)Theme (x3: +REFUGE\_00)Location (f3: +CRIMINAL\_00 ^ +TERRORIST\_00)Beneficiary)

Unidades léxicas asociadas: *safe haven, refugio*.

**14. +SEIZE\_00**

Conceptualización en la ontología nuclear:

+SEIZE\_00: take hold of; also metaphorically: "Fear seized the prisoners"

+(e1: +TAKE\_00 (x1)Theme (x2)Referent (f1: +HAND\_00)Instrument (f2: +FAST\_00 & +STRONG\_00)Manner)

Unidades léxicas asociadas: *clutch at, grab, seize, snatch, agarrar, apropiarse de, arrebatarse, asir, incautarse de*.

Conceptualización en la ontología satélite:

+SEIZE\_00: to take legal possession of a person or property.

+(e1: +TAKE\_00 (x1)Theme (x2: +HUMAN\_00 ^ +PROPERTY\_00)Referent (f1: +HAND\_00)Instrument (f2: +LEGAL\_00)Manner)

Unidades léxicas asociadas: *confiscate, seize, incautar, confiscar, aprehender, decomisar*.

**15. \$SMUGGLING\_00**Conceptualización en la ontología nuclear:

\$SMUGGLING\_00: the crime of taking something illegally from one country to another in secret.

+(e1: +BE\_00 (x1: \$SMUGGLING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +TAKE\_00 (x3: +CRIMINAL ^ +SYNDICATE\_00)Agent (x4: +MONEY\_00 ^ s+TOBACCO ^ s+DRUG\_00 ^ s+WEAPON\_00)Theme (x7)Location (x5: +COUNTRY\_00)Origin (x6: +COUNTRY\_00)Goal) (f1: +SECRET\_00)Manner)

Unidades léxicas asociadas: *contaband, smuggling, contrabando*.

Conceptualización en la ontología satélite:

\$SMUGGLING\_00: the crime of importing or exporting illegal articles on which duties have not been paid. Smuggled goods are liable to confiscation.

+(e1: +BE\_00 (x1: \$SMUGGLING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +TAKE\_00 (x3: +CRIMINAL ^ +SYNDICATE\_00)Agent (x4: s+HUMAN\_00 ^ s+MONEY\_00 ^ s+TOBACCO ^ s+DRUG\_00 ^ s+WEAPON\_00)Theme (x7)Location (x5: +COUNTRY\_00)Origin (x6: +COUNTRY\_00)Goal)

(f1: (e3: n+PAY\_00 (x3)Agent (x8: s+DUTY\_00)Theme (x9)Origin (x10: +GOVERNMENT\_00)Goal)Scene)

\*(e4: +SEIZE\_00 (x11: +OFFICER\_00)Theme (x4)Referent)

Unidades léxicas asociadas: *contaband, smuggling, contrabando*.

## **16. \$SYNDICATE\_00**

### Conceptualización en la ontología nuclear:

\$SYNDICATE\_00: a group of people or companies who join together in order to achieve a particular aim.

\*((e1: +BE\_00 (x1: \$SYNDICATE\_00)Theme (x2: +GROUP\_00)Referent)(e2: +COMPRISE\_00 (x2)Theme (x3: s +COMPANY\_00)Referent))

Unidades léxicas asociadas: *corporation, syndicate, corporación, sindicato.*

### Conceptualización en la ontología satélite:

+SYNDICATE\_00: a group that organizes criminal activities.

+(((e1: +BE\_00 (x1: +SYNDICATE\_00)Theme (x2: +GROUP\_00)Referent)(e2: +COMPRISE\_00 (x2)Theme (x3: s +CRIMINAL\_00)Referent) (e3: +ORGANIZE\_00 (x1)Theme (x4: +CRIME)Referent)))

Unidades léxicas asociadas: *criminal organization, organized criminal group, banda criminal, sindicato (del crimen).*

## **17. \$TAX\_EVASION\_00**

### Conceptualización en la ontología nuclear:

\$TAX\_EVASION\_00: the crime of paying too little tax.

Unidades léxicas asociadas: *tax evasion, tax fraud, defraudación impositiva, fraude fiscal.*

+(e1: +BE\_00 (x1: \$TAX\_EVASION\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +PAY\_00 (x3)Agent (x4: p+TAX\_00)Theme (x5)Origin (x6: +GOVERNMENT\_00)Goal)

Conceptualización en la ontología satélite:

\$TAX\_EVASION\_00: the crime of avoiding the lawful assessment of taxes; for example, by concealing income.

+(e1: +BE\_00 (x1: \$TAX\_EVASION\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +PAY\_00 (x3)Agent (x4: p+TAX\_00)Theme (x5)Origin (x6: +GOVERNMENT\_00)Goal (f1: (e3: +HIDE\_00 (x3)Theme (x7: +INCOME\_00)Referent))Manner)

Unidades léxicas asociadas: *tax evasion, tax fraud, defraudación impositiva, fraude fiscal.*

**18. +TAX\_HAVEN\_00**Conceptualización en la ontología nuclear:

+TAX\_HAVEN\_00: a country with low or no taxation.

+(e1: +BE\_00 (x1: +TAX\_HAVEN\_00)Theme (x2: +COUNTRY\_00)Referent)

+(e2: +PAY\_00 (x3: +HUMAN\_00)Agent (x4: p+TAX\_00)Theme (x5)Origin (x6: +GOVERNMENT\_00)Goal (f1:x1)Location)

Unidades léxicas asociadas: *fiscal paradise, tax haven, paraíso fiscal.*

Conceptualización en la ontología satélite:

+TAX\_HAVEN\_00: A jurisdiction, esp. a country, that imposes little or no tax on the profits from transactions carried on there or on persons resident there.

+(e1: +BE\_00 (x1: +TAX\_HAVEN\_00)Theme (x2: +COUNTRY\_00)Referent)

+(e2: +MOVE\_00 (x3)Agent (x4)Theme (x5)Location (x6)Origin (x7)Goal (f1: (e3: +LIVE\_01 (x4)Theme (x7)Location))Purpose (f2: (e3: +PAY\_00 (x4)Agent (x8: p+TAX\_00)Theme (x9)Origin (x10: +GOVERNMENT\_00)Goal) (f3: x7)Location))Purpose)

Unidades léxicas asociadas: *financial haven, fiscal paradise, tax haven, paraíso fiscal.*

## **19. +TERRORIST\_00**

### Conceptualización en la ontología nuclear:

+TERRORIST\_00: someone who uses terror as a political weapon.

+(e1: +BE\_00 (x1: +TERRORIST\_00)Theme (x2: +HUMAN\_00) Referent)

+(e2: +USE\_00 (x1)Theme (x3: +TERROR\_00)Referent (f1: +WEAPON\_00)Instrument)

+(e3: +BE\_01 (x3)Theme (x4: +POLITICAL\_00)Attribute)

Unidades léxicas asociadas: *terrorist, terrorista.*

### Conceptualización en la ontología satélite:

+TERRORIST\_00: someone who commits a terrorist act; participates as an accomplice in terrorist act; or organizes others to commit a terrorist act.

+(e1: +BE\_00 (x1: +TERRORIST\_00)Theme (x2: +HUMAN\_00) Referent)

+(e2: +USE\_00 (x1)Theme (x3: +TERROR\_00)Referent (f1: +WEAPON\_00)Instrument)

+(e3: +BE\_01 (x3)Theme (x4: +POLITICAL\_00)Attribute)

+(e4: +ORGANIZE\_00 ^ +PARTICIPATE\_00 (x1)Theme (x5: +TERRORIST\_CELL ^ +TERRORIST\_ACT\_00)Referent)

Unidades léxicas asociadas: *terrorist, terrorista.*

## **20. +TORTURE\_00**

### Conceptualización en la ontología nuclear:

+TORTURE\_00: the cause of severe pain to get information.

+(e1: +BE\_00 (x1: +TORTURE\_01)Theme (x2: +CRIME\_00)Referent)

+(e2: +DO\_00 (x3)Theme (x4: m+PAIN\_00)Referent)

(f1: (e3: +OBTAIN\_00 (x3)Theme (x5: +INFORMATION\_00)Referent))Purpose)

Unidades léxicas asociadas: *torture, tortura*.

Conceptualización en la ontología satélite:

+TORTURE\_00: the cause of severe pain to the body or mind to punish, get a confession or information or to obtain sadistic pleasure (to be cruel).

+(e1: +BE\_00 (x1: +TORTURE\_01)Theme (x2: +CRIME\_00)Referent)

+(e2: +DO\_00 (x3: +HUMAN\_00)Theme (x4: m+PAIN\_00)Referent)

+(e3: +BE\_02 (x4)Theme (x5: +BODY\_AREA\_00 ^ +MIND\_00)Location (f1: +IN\_00)Position)

(f1: (e4: +PUNISH\_00 (x3)Theme (x6: +VICTIM\_00)Referent)Purpose) ^

(f2: (e5: +OBTAIN\_00 (x3)Theme (x7: +CONFESSION ^ +INFORMATION\_00)Referent))Purpose) ^

(f3: (e6: +BE\_01 (x3)Theme (x8: +CRUEL\_00)Attribute)Purpose))

Unidades léxicas asociadas: *torture, tortura*.

## **Apéndice 2. PS de 10 conceptos especializados de la ontología satélite del crimen organizado y el terrorismo (por orden alfabético).**

### **1. \$CAROUSEL\_FRAUD\_00**

\$CAROUSEL\_FRAUD\_00: the crime of importing goods from a tax free country. The goods are sold with taxes included but without paying them to the government.

+(e1: +BE\_00 (x1: \$CAROUSEL\_FRAUD\_00)Theme (x2: +FRAUD\_00)Referent)

+(e2: +SELL\_00 (x3: +CRIMINAL\_00)Agent (x4: s+ALCOHOL^s+TOBACCO\_00)Theme (x5)Origin (x6)Goal (f1: (e3 n + PAY\_00 (x3)Agent (7: +TAX\_00)Theme (x5) Origin (x8: +GOVERNMENT\_00)Goal))Condition) (f2: (e3: +SELL\_00 (x3)Agent (x4)Theme (x9)Origin (x10)Goal) (f3: (e4: +PAY\_00 (x10)Agent (x11: +TAX\_00)Theme (x9)Origin (x3)Goal))Condition)

Unidades léxicas asociadas: *carousel fraud, fraude carrusel.*

### **2. \$COUNTERFEITING\_00**

\$COUNTERFEITING\_00: unlawful making of false money.

+(e1: +BE\_00 (x1: \$COUNTERFEITING\_00)Theme (x2: ORGANISED\_CRIME\_00)Referent)

(f1: (e2: +CREATE\_00 (x3: +CRIMINAL\_00)Theme (x4: s+MONEY\_00)Referent (f2)Instrument (e3: n+BE\_01 (x4)Theme (x5: +TRUE\_00)Attribute))Result)

+(e4: +CHANGE\_00 (x3)Theme (x4)Referent (f3: (e5: +BECOME\_00 (x4)Theme (x6: +REAL\_00)Attribute))Result)

Unidades léxicas asociadas: *counterfeiting, falsificación de moneda.*

### 3. \$CUCKOO\_SMURFING\_00

\$CUCKOO\_SMURFING\_00: the crime of putting illegal money into the third parties bank accounts.

+(e1: +BE\_00 (x1: \$CUCKOO\_SMURFING\_00)Theme (x2: +MONEY\_LAUNDERING\_00)Referent)

+(e2: +PUT\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x7: +BANK\_ACCOUNT\_00)Location (x5)Origin (x6)Goal

(f1: (e3+BE\_02 (x7)Theme (x8)Location) (f2: +IN\_00)Condition))Result)

Unidades léxicas asociadas: *cuckoo smurfing*.

### 4. \$LAYERING\_00

\$LAYERING\_00: the second stage of money laundering by which the dirty money is transferred to offshore accounts where it is moved in and out.

+(e1: +BE\_00 (x1: \$LAYERING\_00)Theme (x2: +MONEY\_LAUNDERING)Referent)

+(e2: +TRANSFER\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x5)Origin (x6)Goal (x7: +TAX\_HAVEN\_00)Location (f1: +IN\_00 & +OUT\_00)Position)

Unidades léxicas asociadas: *layering*.

### 5. \$LONE\_SHARK\_00

\$LONE\_SHARK\_00: someone who lends money at an illegal interest rate and uses force to obtain repayment.

+(e1: +BE\_00 (x1: \$LONE\_SHARK\_00)Theme (x2: +CRIMINAL\_00)Referent)

+(e2: +LEND\_00 (x1)Agent (x3: +MONEY\_00)Theme (x4)Origin (x5)Goal (f1: (e3 +BE\_01 (x6: +INTEREST\_RATE\_00)Theme (x7: n+LEGAL\_00)Attribute))Condition)

+(e4: +USE\_00 (x1)Theme (x8: +FORCE\_00)Referent (f2: (e5: + fut +RETURN\_00 (x5)Agent (x3)Theme (x8)Location (x9)Origin (x1)Goal))Condition)

Unidades léxicas asociadas: *lone shark*.

## **6. \$LONE\_WOLF\_00**

\$LONE\_WOLF\_00: someone who commits violent acts in support of some group, movement, or ideology, but does so alone, outside of any command structure.

+(e1: +BE\_00 (x1: \$LONE\_WOLF\_00)Theme (x2: +TERRORIST\_00)Referent)

+(e2: +USE\_00 (x1)Theme (x3: +VIOLENCE\_00)Referent (f1: (e3: +BE\_01 (x1)Theme (x4: +ALONE\_00)Attribute)

+(e4: +DEFEND\_00 (x1)Theme (x5: +GROUP\_00 ^ +MOVEMENT\_00 ^ +IDEOLOGY\_00)Referent)

Unidades léxicas asociadas: *lone wolf, lone wolf fighter*.

## **7. \$PHARMING\_00**

\$PHARMING\_00: a hacker's attack to redirect the website's traffic to a false one.

+(e1: +BE\_00 (x1: \$PHARMING\_00)Theme (x2: +CYBERCRIME\_00)Referent)

+(e2: +ATTACK\_00 (x3: +HACKER\_00)Theme (x4: +WEBSITE\_00)Referent (f1) Instrument

(f2: (e3: +COMMAND\_00 (x3)Theme (x5: +TRAFFIC\_00)Referent (x6: +WEBSITE\_00)Goal)Purpose)

Unidades léxicas asociadas: *pharming*.

## **8. \$SELF\_LAUNDERING\_00**

\$SELF\_LAUNDERING\_00: the crime of putting illegal money into the criminal's own bank account.

+(e1: +BE\_00 (x1: \$ SELF\_LAUNDERING\_00)Theme (x2: +MONEY\_LAUNDERING)Referent)

+(e2: +TRANSFER\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x5)Origin (x6: +TAX\_HAVEN\_00)Goal

+(e2: +PUT\_00 (x3: +CRIMINAL\_00)Agent (x4: +DIRTY\_MONEY\_00)Theme (x7: +BANK\_ACCOUNT\_00)Location (x5)Origin (x6)Goal

(f1: (e3: +BE\_02 (x7)Theme (x6)Location) (f2: +IN\_00)Condition))Result)

Unidades léxicas asociadas: *self laundering, autoblanqueo*.

## 9. \$VISHING\_00

\$VISHING\_00: the criminal practice of using social engineering over the telephone system to gain access to private personal and financial information. The term is a combination of "voice" and phishing.

+(e1: +BE\_00 (x1: \$VISHING\_00)Theme (x2: +ORGANIZED\_CRIME\_00)Referent)

+(e2: +USE\_00 (x3)Theme (x4: +ENGINEERING\_00)Referent (f1: (e3: +BE\_02 (x1)Theme (x5: +TELEPHONE\_00)Location)

((e4: +OBTAIN\_00 (x3)Theme (x6: +INFORMATION\_00)Referent)(e5: +BE\_01 (x6)Theme (x7: +FINANCIAL\_00)Attribute)(f2: +VOICE\_00)Instrument))

Unidades léxicas asociadas: *vishing*.

## 10. \$WATERBOARDING\_00

\$WATERBOARDING\_00: a form of torture with water.

+(e1: +BE\_00 (x1: \$WATERBOARDING\_00)Theme (x2: +TORTURE\_00)Referent)

+(e2: +USE\_00 (x3: +CRIMINAL ^ +SYNDICATE\_00 ^ +TERRORIST\_00)Theme (x4 +WATER\_00) Referent)

+(e3: +FEEL\_00 (x3)Agent (x5: +VICTIM\_00)Theme (x6: +AFRAID\_00)Attribute)

(f1: (e4: ing+DROWN\_00 (x3)Theme (x5: +VICTIM\_00)Referent)Result)

Unidades léxicas asociadas: *waterboarding*.

