
La jerarquización cognitiva de las entidades en la ontología satélite del crimen organizado y el terrorismo en FunGramKB

MARÍA DE GRACIA CARRIÓN DELGADO
UNED
mdgcarrion@madrid.uned.es

ÁNGEL FELICES LAGO
Universidad de Granada
afelices@ugr.es

RESUMEN

*FunGramKB*¹ es una base de conocimiento léxico-conceptual multipropósito y multilingüe, diseñada para su aplicación en tareas de procesamiento del lenguaje natural (Periñán y Arcas 2004; Mairal y Periñán, 2009). Su estructura modular refleja tres niveles independientes de conocimiento (léxico, gramatical y ontológico), pero relacionados entre sí, lo cual permite aplicar la base de conocimiento a labores de traducción automática o de razonamiento artificial en diversas lenguas (Periñán y Mairal, 2010), en función de un módulo conceptual compartido por todas las lenguas. Así, la Ontología está compuesta de dos estructuras: un módulo de propósito general (la ontología nuclear) y varios módulos terminológicos de dominios específicos. La presente contribución muestra cómo se desarrolla la jerarquización cognitiva de las entidades en la Ontología satélite del crimen organizado y el terrorismo² vinculada a la Ontología nuclear de *FunGramKB*. A modo de ilustración ofrecemos ejemplos de este dominio donde presentamos casos de «conceptos espejo» y de conceptos especializados.

Palabras clave: *FunGramKB*, terminología, procesamiento del lenguaje natural, ontología nuclear, ontología satélite, concepto espejo.

1. INTRODUCCIÓN

A modo de introducción presentamos *FunGramKB*, que es la base de conocimiento en la que se enmarca nuestro trabajo.

FunGramKB es una base de conocimiento léxico-conceptual multipropósito y multilingüe que ha sido diseñada para la realización de tareas de procesamiento del lenguaje

¹ *FunGramKB* ha sido diseñada para trabajar con siete lenguas: alemán, búlgaro, catalán, español, francés, inglés e italiano.

² 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.

natural tales como la recuperación y extracción de información, traducción automática, etc. Su carácter multilingüe hace que pueda ser reutilizada en diversas lenguas.

Por otro lado, en *FunGramKB* están representados los tres niveles principales de conocimiento, esto es, léxico, gramatical y conceptual a través de módulos independientes pero interrelacionados entre sí (Periñán y Arcas, 2011: 2-3):

Nivel léxico:

- En el *Lexicón* se encuentra almacenada la información morfosintáctica, pragmática y de colocaciones de las unidades léxicas que componen la base de conocimiento.
- El *Morficón* asiste al analizador y generalizador en el tratamiento de la morfología flexiva ya que aporta la información morfosintáctica, pragmática y de colocaciones.

Nivel gramatical:

- El *Gramaticón* almacena los esquemas construccionales 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 satélite (geología, 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 hacer una tortilla de patatas.
- El *Onomasticón* almacena conocimiento enciclopédico sobre instancias de entidades y eventos como Amancio Ortega 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 1, en *FunGramKB*, los módulos léxico y gramatical son dependientes de cada lengua, mientras que el módulo conceptual es compartido por todas las lenguas que conforman la base de conocimiento.

LA JERARQUIZACIÓN COGNITIVA DE LAS ENTIDADES EN LA ONTOLOGÍA

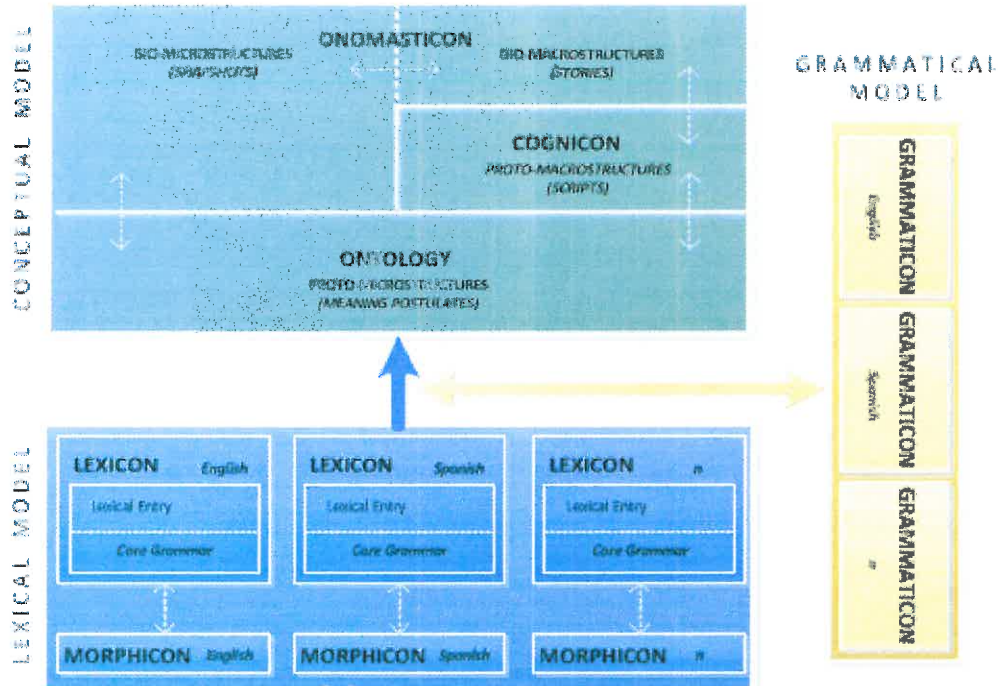


Figura 1. La suite FunGramKB (<http://www.fungramkb.com>).

Así, FunGramKB tiene carácter conceptualista ya que la Ontología nuclear 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 a través de relaciones superficiales entre unidades léxicas (como por ejemplo en FrameNet o MultiWordnet), *FunGramKB* se basa en la denominada «semántica profunda», que muestra un mayor potencial expresivo (Periñán y Arcas, 2007; Felices y Marín, 2012).

Como *FunGramKB* tiene un propósito general y no está pensada en principio para dominios específicos, puede ser ampliada y enriquecida con conocimiento experto. Así pues, la ontología nuclear se puede dotar de conocimiento experto a través de enlaces con ontologías satélite (Faber, Mairal y Magaña 2011: 564), como se muestra en la Figura 2 a continuación.

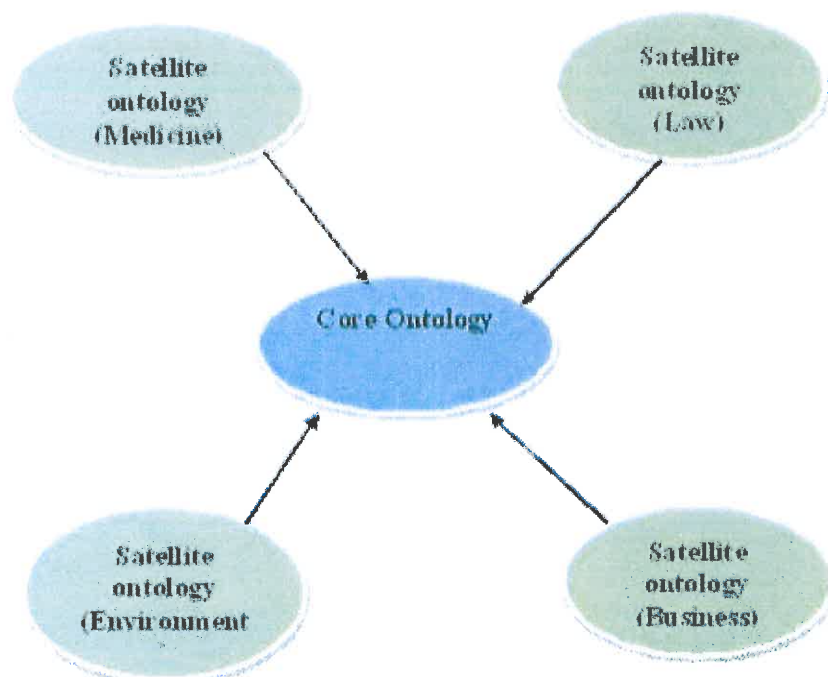


Figura 2. La ontología nuclear y las ontologías satélite (Mairal, Periñán y Samaniego, 2011: 5).

Por otro lado, el modelo multinivel de la ontología nuclear de *FunGramKB*, que analizaremos en detalle en la sección 2.2, está dividido en metaconceptos, conceptos básicos y terminales que reflejan la representación conceptual de la estructura cognitiva humana.

Asimismo, *FunGramKB suite* permite reutilizar la metodología y las herramientas en cualquier ontología satélite, independientemente de cuál sea su dominio temático. Con lo cual, los ingenieros del conocimiento elaboran una ontología común a los diversos lexicones de las lenguas implementadas en la base de conocimiento.

En este estudio mostramos cómo se desarrolla la jerarquización cognitiva de las entidades en la ontología satélite del crimen organizado y el terrorismo que estará vinculada a la ontología nuclear de *FunGramKB*. Y como se mostrará en la sección 2.2, mediante el modelado de los denominados «concepto espejo» se reducirá la redundancia informativa y se reutilizará la información conceptual de la Ontología nuclear.

En consecuencia, el desarrollo del contenido de este trabajo se estructura del siguiente modo: En primer lugar, presentamos *FunGramKB Term Extractor*, una nueva herramienta para la extracción automática de términos candidatos. Seguidamente, exponemos los aspectos principales del análisis terminológico, cuyo objetivo es la definición de unidades especializadas. Para ello explicamos la metodología *COHERENT* donde se ofrecen varios ejemplos de conceptualización, dos de «conceptos espejo» y dos de términos especializados; finalmente se efectuará una representación de sus correspondientes propuestas de jerarquización.

2. FUNGRAMKB TERM EXTRACTOR

FunGramKB Term extractor es una herramienta semiautomática que facilita la labor del terminólogo en la selección de términos ganadores. Así, en una primera fase, permite obtener de forma automática una lista de términos candidatos y descartar los términos falsos cuyo índice de frecuencia *tf-idf* es menor de tres. En una fase posterior, el terminólogo realiza un primer filtrado manual de unidades léxicas (unigramas) y sintagmáticas no terminológicas (bigramas y trigramas). A continuación, el terminólogo ha de realizar un segundo y último filtrado para obtener una lista de términos ganadores.

En la fase manual del filtrado es necesario el uso y consulta minuciosa de diccionarios especializados para determinar cuáles de los términos candidatos serán finalmente ganadores.

El extractor de *FunGramKB* permite, a partir de un corpus de textos, obtener automáticamente una lista de términos candidato especializados y representativos de un dominio concreto. A partir de ahí comienza la labor del terminólogo, quien elabora de forma manual el filtrado de los términos y la definición de conceptos. Dichos conceptos conformarán la ontología satélite que estará vinculada a la ontología nuclear de la base de conocimiento de *FunGramKB*.

En la figura 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. Aquí nos centraremos en *view*, *search* y *corpus* que son los correspondientes a la extracción terminológica.

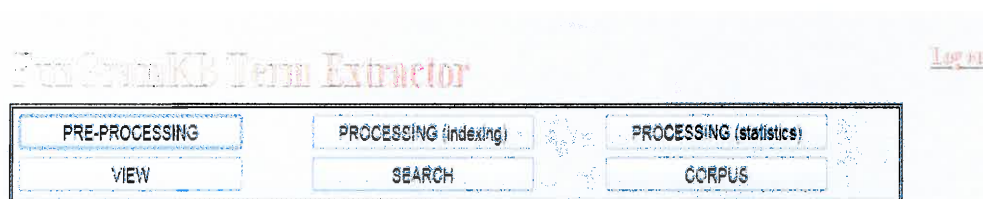


Figura 3. Pantalla principal del extractor de *FunGramKB*.

De las tres herramientas mencionadas anteriormente, *corpus* es especialmente útil ya que nos permite consultar los documentos que conforman nuestro corpus especializado. En la figura 4 podemos ver ejemplos del corpus del crimen organizado y el terrorismo *GCTC*³. Destacar que sólo para el inglés se han subido al extractor 621 textos.

³ *GCTC (Global Crime Term Corpus)*, corpus de textos del crimen organizado y el terrorismo utilizado para este estudio.

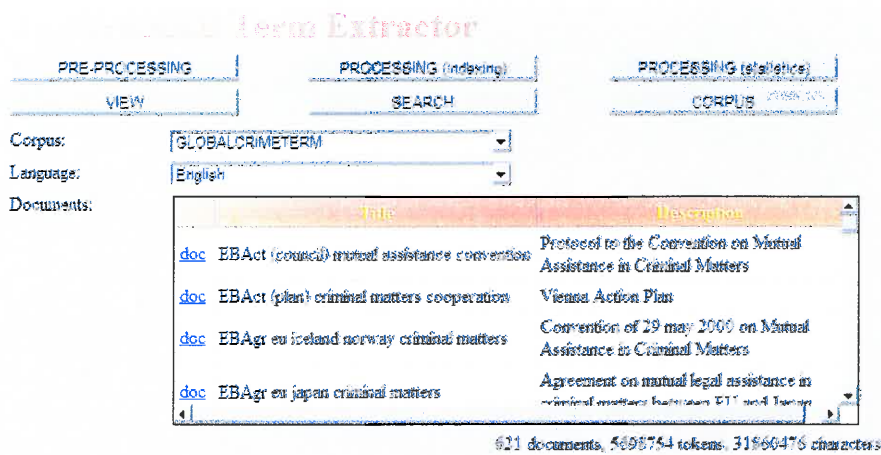


Figura 4. Pantalla de visualización del corpus.

Por su parte, el elemento view nos permite ver la lista de términos candidatos. Así, en las figuras 5, 6 y 7 podemos ver ejemplos de trigramas, bigramas y unigramas⁴ candidatos del dominio del crimen organizado y terrorismo respectivamente.

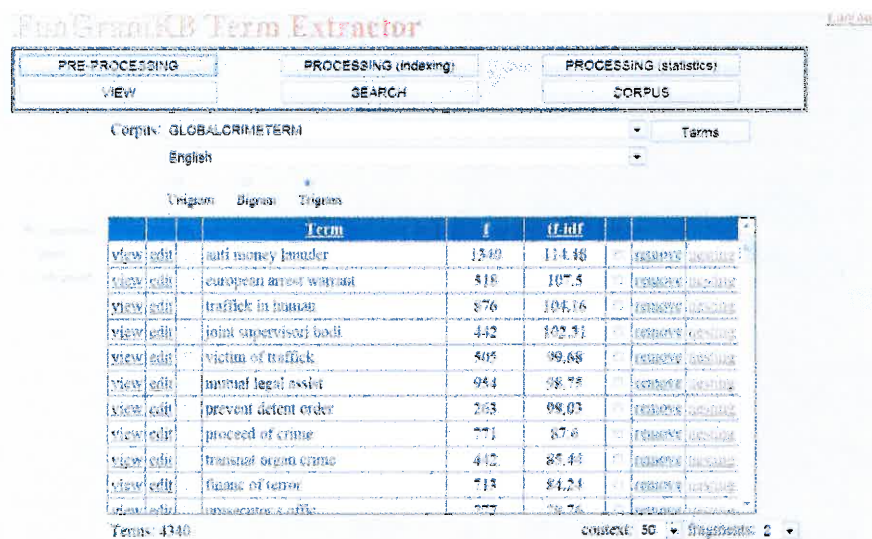


Figura 5. Tabla de trigramas candidatos.

⁴ 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, en *FunGramKB Term Extractor* encontramos tres tipos básicos de «n-gramas»: los unigramas (palabras individuales), los bigramas o secuencias de dos palabras y los trigramas (formaciones unificadas de tres palabras).

LA JERARQUIZACIÓN COGNITIVA DE LAS ENTIDADES EN LA ONTOLOGÍA

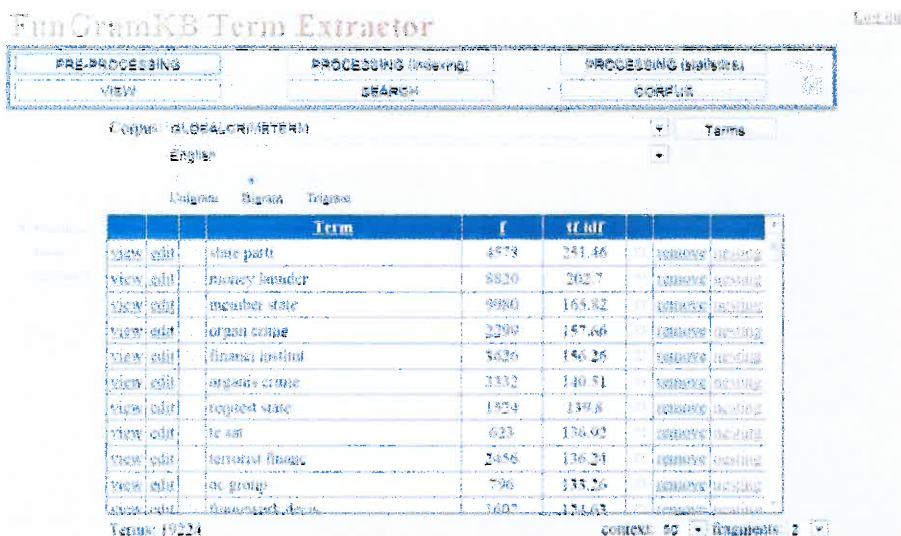


Figura 6. Tabla de bigramas candidatos.

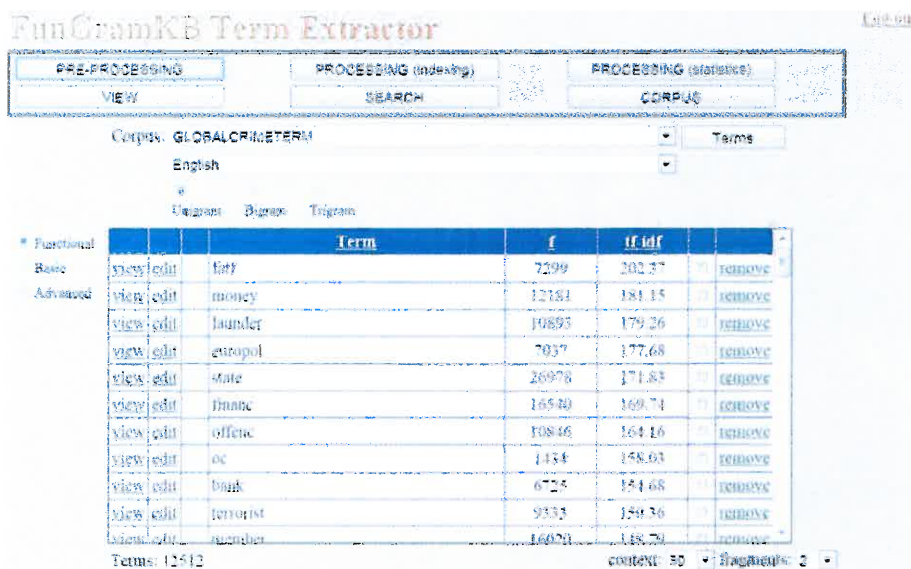


Figura 7. Tabla de unigramas candidatos.

2.1. El análisis terminológico

Durante el proceso de análisis, el terminólogo considerará término candidato aquel que potencialmente sea relevante para un dominio léxico concreto con la ayuda del índice de frecuencia del extractor. En el análisis de los términos candidatos el experto ha

de decidir cuáles términos son ganadores y cuáles falsos candidatos. Así, con la ayuda del extractor, considerará también término ganador aquel que sea propio del ámbito de conocimiento específico. Por el contrario determinará como falso candidato aquel que, tras su análisis, no se considere especializado al pertenecer al lenguaje general (Felices, Ureña y Alameda, 2011). En este caso serán de gran utilidad los *Learners' Dictionaries*, ya que sus definiciones ayudarán al terminólogo a identificar las unidades terminológicas.

2.2.. Conceptualización de términos ganadores

Una vez elaborada la selección de términos ganadores, el terminólogo ha de elaborar la jerarquización y conceptualización de los mismos. Se trata de identificar las palabras definitorias del dominio temático (i.e., los conceptos básicos) que servirán de base para la definición de otros conceptos más específicos ; esto es, los conceptos terminales.

La Ontología nuclear utiliza el lenguaje de representación *COREL (COncceptual Representation Language)*⁵ para modelar el significado de sus conceptos. En la jerarquía conceptual de *FunGramKB* los conceptos pertenecen a tres niveles. El nivel superior está formado por 42 metaconceptos que representan dimensiones cognitivas en mayúsculas precedidas por el signo «#». Este resultado es fruto del análisis de algunas de las ontologías lingüísticas más relevantes como SUMO (Niles y Pease, 2001a, 2001b) o DOLCE (Gangemi et al., 2002; Masolo et al., 2002). La Ontología de *FunGramKB* contiene a su vez tres subontologías cuyos metaconceptos son *#ENTITY*, *#EVENT* y *#QUALITY* (figura 8), que permiten la organización interna de nombres, verbos y adjetivos respectivamente.

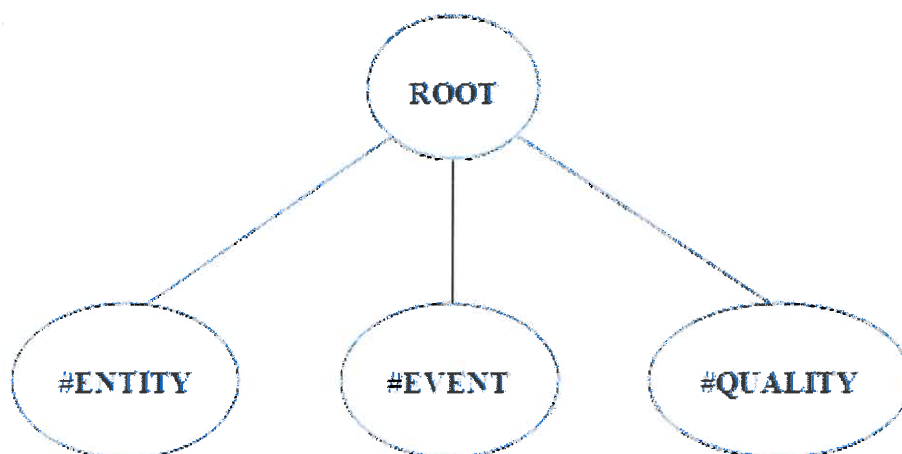


Figura 8: Tipos de conceptos.

⁵ Véase Perrián y Mairal (2010) para una descripción detallada del lenguaje COREL.

Ejemplos de metaconceptos son #ABSTRACT, #PHYSICAL, #OBJECT, #PROCESS, etc) que podemos ver en la figura 9.

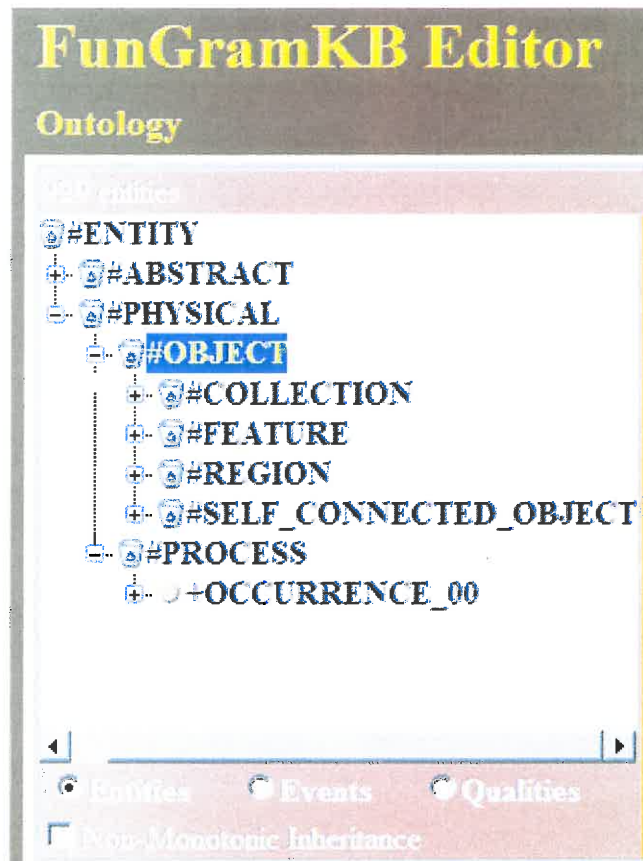


Figura 9. Representación parcial de los metaconceptos de las entidades.

Seguidamente encontramos los conceptos básicos, que están en un nivel intermedio, se representan con el signo + y van seguidos de un guión bajo y un índice numérico (por ejemplo, +JUDGE_00, +LAWYER_00, +PRISONER_00, etc, de los cuales podemos ver algunos ejemplos en la figura 9); y por último están los conceptos terminales, precedidos por el signo \$ y también seguidos de un guión bajo y de un índice numérico (por ejemplo \$MANKIND_00, \$PENALTY_00, \$SYNDICATE_00, etc).

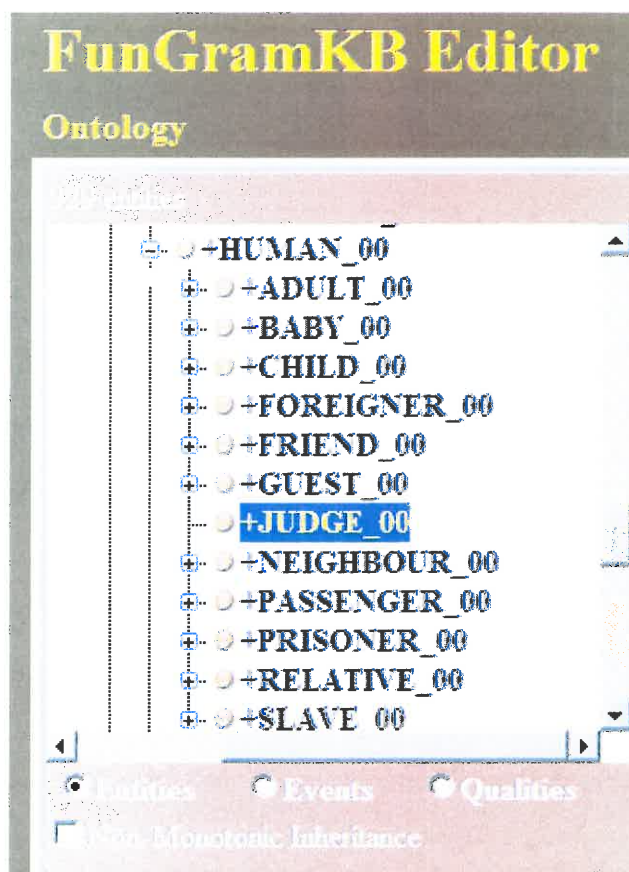


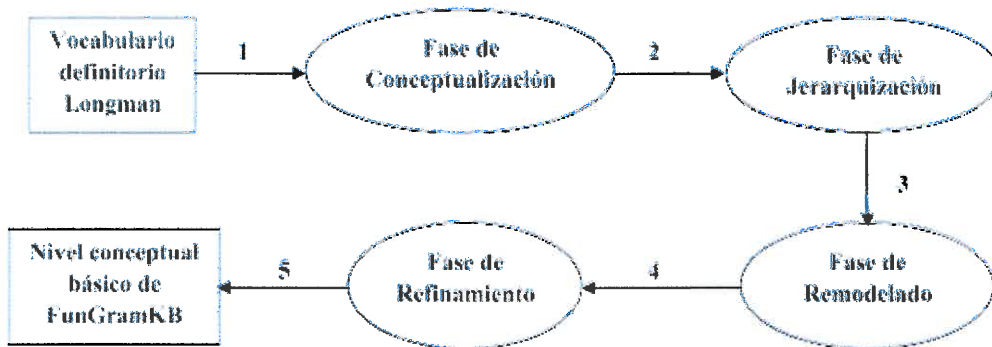
Figura 10. Representación parcial de las entidades.

3. LA METODOLOGÍA COHERENT

La metodología *COHERENT* (*CO*nceptualization + *Hi*Erarchization + *RE*modelling + *refinemeNT*) diseñada por Periñán y Mairal (2011), se utiliza para la construcción del nivel conceptual básico de la ontología nuclear de la base de conocimiento *FunGramKB* y también sirve de base para el desarrollo de ontologías satélite vinculadas a ella (Carrión, en prensa).

A continuación presentamos las fases de su desarrollo. En un primer paso se identificaron los conceptos básicos del *Longman Defining Vocabulary (LDV)* del *Longman Dictionary of Contemporary English* (Procter, 1978), el cual ya ha sido probado como referencia para establecer el vocabulario básico de un lenguaje artificial. Sin embargo, hubo que realizar una revisión profunda para poder desarrollar el mapa conceptual. Concretamente, tanto la población como la 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 11.

LA JERARQUIZACIÓN COGNITIVA DE LAS ENTIDADES EN LA ONTOLOGÍA



1. Lista de 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.

Figura 11. La metodología COHERENT (adaptado de Perrián y Mairal, 2011:20).

Durante la fase de conceptualización, se convirtió la lista de palabras inglesas del *LDV* en un inventario de unidades conceptuales interlingüísticas. Para ello hubo que rechazar algunas unidades léxicas (es decir, algunas palabras del *LDV* no se asignaron a conceptos básicos sino a conceptos terminales) y se tuvieron que realizar grupos cognitivos (esto es, se agruparon algunas palabras del *LDV* en un mismo concepto básico).

Como hemos mencionado anteriormente, la Ontología de *FunGramKB* se divide a su vez en tres subontologías (*#ENTITY*, *#EVENT* y *#QUALITY*) mediante la cual los conceptos se clasifican jerárquicamente. De este modo, los conceptos de la fase anterior se distribuyeron en cada una de estas tres subontologías y después se organizaron jerárquicamente de acuerdo a su relación *IS-A*. Durante este proceso fue necesario introducir algunos «conceptos paraguas» y así poder explotar de manera más eficiente el mecanismo de herencia de la ontología a nivel medio (conceptos básicos).

En la fase de remodelado algunos conceptos básicos se degradaron a «subconceptos», dado que, aunque compartan el mismo Postulado de Significado (PS) presentaban diferencias en las preferencias de selección de alguno de los argumentos. Por consiguiente, esto afectó a la jerarquía ontológica. En la figura 12 podemos ver las distintas capas presentes en la ontología).



Figura 12. La jerarquía conceptual en FunGramKB (Jiménez, Luzondo y Cabello, 2011: 19).

En la fase de refinamiento se eliminaron los conceptos básicos que resultaron ser muy poco productivos. Es decir, si un concepto básico determinado no formaba parte de los PS de un gran número de conceptos, o bien se le fusionaba a su superordinado, o bien se le degradaba a concepto terminal, dependiendo de la presencia o ausencia de los conceptos subordinados. Hay que destacar que esta fase se aplicó cuando el nivel terminal de la ontología nuclear de *FunGramKB* ya estaba considerablemente poblado.

Como resultado de estas cuatro fases se obtuvo un inventario de aproximadamente 1.300 conceptos básicos que han servido de base para poblar la ontología de conceptos terminales, proceso que sigue en desarrollo a día de hoy. En el momento de realización de este artículo, la ontología nuclear de *FunGramKB* contaba ya de 404 eventos, 929 entidades y 312 cualidades.

Con respecto a la ontología satélite del crimen organizado y el terrorismo, nos encontramos aún desarrollando tanto la fase de conceptualización como la de jerarquización, por lo que ambas están aún sujetas a modificaciones. Por consiguiente, los resultados que presentamos aquí son los correspondientes a un estudio preliminar.

3.1. La jerarquización cognitiva

Basamos la jerarquía conceptual en la relación de subsunción (*IS-A*) que permite la herencia múltiple no monótonica mediante la cual los conceptos están relacionados con las unidades léxicas, pero no dependen de ellas. Así, la definición en *COREL* está formada por el *GENUS+DIFFERENTIAE*, donde el *GENUS* delimita la jerarquía de conceptos ya

que es compartido por sus conceptos hipónimos, como podemos ver en los ejemplos a continuación:

```
#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE_00>+CRIME_00>
+ORGANIZED_CRIME_00>+RACKETEERING_00
#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE_00>+CRIME_00>
+ORGANIZED_CRIME_00>+SMUGGLING_00
```

A continuación presentamos y analizamos en detalle dos ejemplos de lo que entendemos como «conceptos espejo»:

El significado de *tax haven* será general o especializado según el tipo de diccionario que consultemos. El ejemplo (3) muestra una definición general de la palabra *tax haven* obtenida del diccionario *Logman Dictionary of Contemporary English*, mientras que el ejemplo (4) muestra una definición especializada extraída del diccionario jurídico *Black's Law Dictionary* (9.ª edición, 2009):

(3) *Tax haven*: a place where people go to live to avoid paying high taxes in their own country.

(4) *Tax haven*: A jurisdiction, esp. a country, that imposes little or no tax on the profits from transactions carried on there or on persons resident there.

Las diferencias entre una entrada y otra son evidentes, aun tratándose de dos definiciones de una misma unidad léxico-conceptual, la segunda es cualitativamente más precisa. Podríamos decir, por tanto, que nos encontramos ante un caso de «concepto espejo» al tratarse de un término que contiene un gran componente de conocimiento del sentido común, pero que en la ontología satélite del crimen organizado y terrorismo adquiriría un matiz más especializado ya que se trata de un concepto clave en este ámbito. Por tanto, en el proceso de población de la ontología satélite sobre el crimen organizado y terrorismo tendríamos que crear el concepto +TAX_HAVEN_00 que sería espejo de su equivalente en la Ontología nuclear, pero con una formalización en *COREL* más específica que reflejaría su contenido especializado, que mostramos a continuación.

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 (inglés y español): *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 (inglés y español): *financical haven, fiscal paradise, tax haven, paraíso fiscal*.

Así, la propuesta de jerarquización de este concepto es la siguiente:

#ENTITY>#REGION>#PLACE>+ARTIFICIAL_AREA_00>+COUNTRY_00>
TAX_HAVEN_00

Por otro lado, podemos encontrarnos ante el hecho de que muchos bigramas y trigramas no aparecen en los diccionarios especializados, como ocurre con el «concepto espejo» *dirty money*. El *Cambridge Learner's Dictionary* y el *GTC* lo definen así (ejemplos 1 y 2 respectivamente):

(1) *Dirty money*: Money that someone gets in an unfair, illegal, or dishonest way.

(2) *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. (*GTC*).

Como este concepto no aparece en las fuentes lexicográficas especializadas, es necesaria la consulta directa al *GTC* en contraposición a los *Learner's Dictionaries*, para poder escindir así los rasgos propios del sentido común de los equivalentes al conocimiento experto que nos permitan elaborar su definición. Aunque podríamos decir que estamos ante un concepto altamente frecuente en el lenguaje común, la consulta de textos especializados nos que indica se trata de un concepto clave en el ámbito del crimen organizado.

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 (inglés y español): *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 (inglés y español): *dirty money*, *dinero negro*.

La propuesta de jerarquización de este concepto queda formalizada del siguiente modo:

#ENTITY>#PHYSICAL>#OBJECT>#SELF_CONNECTED_OBJECT>
+NATURAL_OBJECT_00>+CORPUSCULAR_00>+SOLID_00>+BANKNOTE_00>
+DIRTY_MONEY_00

Seguidamente presentamos y analizamos un par de ejemplos de conceptos especializados. El primero (ejemplo 5) es *carousel fraud*, término de reciente creación dado que no se encuentra en los diccionarios especializados.

(5) *Carousel fraud*: The crime of importing goods from a tax free country. The goods are sold with taxes included but without paying them to the government.

Conceptualización en la ontología satélite:

\$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 (inglés y español): *carousel fraud, fraude carrusel*.

La propuesta de jerarquización de este término queda descrita del modo siguiente:

#ENTITY>#PHYSICAL>#PROCESS>
#OCCURRENCE>+CRIME_00>
+FRAUD_00>CC

El segundo ejemplo que hemos seleccionado es *waterboarding*, que se define de la siguiente manera:

(6) *Waterboarding*: a form of torture with water.

Conceptualización en la ontología satélite:

\$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 (inglés y español): *waterboarding, tortura por ahogamiento*

La propuesta de jerarquización de este término se presenta a continuación:

#ENTITY>#PHYSICAL>#PROCESS>
#OCCURRENCE>+CRIME_00>
+TORTURE_00>\$WATERBOARDING_00

4. CONCLUSIONES

Una de las conclusiones más plausibles del conjunto de ejemplos ofrecidos en los apartados precedentes es que, a nuestro juicio, a partir de un catálogo cerrado de meta-conceptos podemos definir los «conceptos espejo» y los conceptos especializados de la ontología satélite seleccionada. De hecho, los «conceptos espejo» aparecen reflejados tanto en la ontología nuclear como en la satélite y, donde la primera refleja el conocimiento de lo que se entiende comúnmente como sentido común, y la segunda procura proyectar el conocimiento experto o especializado, aun cuando el referente conceptual sea siempre el mismo en los dos casos.

Por otro lado, entre las dificultades halladas durante la fase de conceptualización, merece destacarse el gran número de «conceptos espejo» donde el conocimiento experto

era muy difícil de definir dada la alta coincidencia de rasgos propios del sentido común presentes en las definiciones de los *Learner's Dictionaries* (en casos como *dirty money* y *tax haven*) si se contrastan con las extraídas de fuentes terminológicas. También hemos podido observar que la labor del terminólogo se complica sustancialmente con la consulta de bigramas y trigramas en los recursos lexicográficos especializados, dado que muchas de estas unidades n-gramáticas no se encuentran en ninguna de esas fuentes. Para solucionar estos problemas la consulta directa al *GCTC* ha sido de vital importancia, dado que facilita la inferencia del significado a partir de cotejar un número representativo de ejemplos.

En resumen, los ejemplos mostrados del dominio del crimen organizado y el terrorismo nos han servido para probar que el enfoque conceptualista de la base de conocimiento *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 el lenguaje de interfaz *COREL*.

REFERENCIAS BIBLIOGRÁFICAS

- CARRIÓN, M. (2012). Extracción y análisis de unidades léxico-conceptuales del dominio jurídico: un acercamiento metodológico desde FunGramKB, en *RaeL* 11, 25-39.
- 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).
- FELICES A. y A. MARÍN. 2012: La semántica profunda como fundamento para desarrollar una subontología jurídica en el contexto de las ontologías de ámbito legal. En: *Empiricism and Analytical Tools for 21st Century Applied Linguistics. Selected papers from the XXIX International Conference of the Spanish Association of Applied Linguistics (AESLA)*. Izaskun Elorza, Ovidi Carbonell i Cortés, Reyes Albarrán, Blanca García Riaza, Miriam Pérez-Veneros (eds.) Salamanca: Universidad, 441-450.
- FELICES, A., P. UREÑA y A. ALAMEDA. 2011. FunGramKB y la adquisición terminológica. *Anglogermánica Online* 2011: 66-86.
- GANGEMI, ALDO et al. 2002. Sweetening ontologies with DOLCE. En: Asunción Gómez-Pérez y Richard Benjamins (eds.) *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web: 13th International Conference, EKAW 2002*, Sigüenza, 1-4.
- GARNER, BRIAN A., ed. 2009. *Black's Law Dictionary*. 9ª edición. USA.
- JIMÉNEZ, R., A. LUZONDO y M. BEATRIZ CABELLO. 2011. FunGramKB y la organización ontológica. *Anglogermánica Online* 2011: 16-36.
- MAIRAL, R. y C. PERIÑÁN. 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., C. PERIÑÁN y E. SAMANIEGO. 2011. Using ontologies for terminological knowledge representation: a preliminary discussion. *Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10* (eds. N. Talaván, E. Martín Monje y F. Palazón): 267-280. UNED: Madrid.
- MASOLO, C., S. BORGO, A. GANGEMI, N. GUARINO, A. OLTRAMARI y L. SCHEIDER. 2002. The WonderWeb Library of Foundational Ontologies and the DOLCE Ontology. *WonderWeb Deliverable D18*. Disponible en línea. Último acceso 30/12/2012.

- [<http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf>].
- NILES, Ian 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.
- 2001b. Towards a Standard Upper Ontology. En: *Proceedings of the Second International Conference on Formal Ontology in Information Systems*. Ogunquit.
- 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.
- 2007. Deep semantics in an NLP knowledge base, *12th Conference of the Spanish Association for Artificial Intelligence*, Universidad de Salamanca: 279-288.
- 2010. The architecture of FunGramKB, en *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. 17-23 mayo 2010, Valeta (Malta). European Language Resources Association (ELRA). 2667-2674.
- 2011. Introducción a FunGramKB. *Anglogermánica Online* 2011: 1-15.
- PERIÑÁN, C. y R. MAIRAL. 2010. La gramática de COREL: un lenguaje de representación conceptual. *Onomázein* 21: 11-45.
- PROCTER, P. (ed.). 1978. *Longman Dictionary of Contemporary English*. Harlow (Essex): Longman.
- 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. 301-306.
- VAN VALIN, ROBERT D. JR. y R. LAPOLLA. 1997. *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- VAN VALIN, ROBERT D. JR. 2005. *Exploring the Syntax-Semantic Interface*. Cambridge: Cambridge University Press.

DICCIONARIOS

- CAMBRIDGE: *Cambridge Learner's Dictionary*. [<http://dictionary.cambridge.org/dictionary/learner-english/>].
- LONGMAN: *Longman Dictionary of Contemporary English*. [<http://www.ldoceonline.com/>].
- OXFORD: *Oxford Dictionary of Law*. 2009. 9.^a edición. Oxford: Oxford University Press.