

EXTRACCIÓN Y ANÁLISIS DE UNIDADES LÉXICO-CONCEPTUALES DEL DOMINIO JURÍDICO: UN ACERCAMIENTO METODOLÓGICO DESDE FUNGRAMKB¹

MARÍA DE GRACIA CARRIÓN DELGADO
UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

Resumen: *FunGramKB* es una base de conocimiento que ha sido diseñada para su aplicación en tareas de procesamiento del lenguaje natural (Periñán-Pascual y Arcas-Túnez 2004, 2005; Mairal-Usón y Periñán-Pascual 2009; Periñán-Pascual y Mairal-Usón 2010). Su estructura modular, compuesta de tres componentes –conceptual, léxico y gramatical–, permite aplicar la base de conocimiento en labores de traducción automática o razonamiento artificial (Periñán-Pascual y Mairal-Usón 2009b, 2010b). Una de las vías de ampliación del componente conceptual de *FunGramKB*, representado en la ontología nuclear, consiste en la creación de ontologías satélite. La presente contribución ofrece una metodología para la gestión y recuperación de unidades terminológicas a partir de la explotación de corpórea especializados (Felices-Lago, Ureña Gómez-Moreno y Alameda-Hernández 2011). El denominado modelado de ‘conceptos espejo’ se propone como estrategia para la reutilización de información conceptual de la ontología nuclear. Para ilustrar el análisis terminológico-conceptual se aportan ejemplos del dominio del crimen organizado y el terrorismo.

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

Abstract: *FunGramKB* is a knowledge base that has been designed for its use in tasks of natural language processing (Periñán-Pascual y Arcas-Túnez 2004, 2005; Mairal-Usón y Periñán-Pascual 2009; Periñán-Pascual y Mairal-Usón 2010). Its modular structure, composed of three components –conceptual, lexical and grammatical– allows the application of the knowledge base for work in related machine translation or artificial reasoning (Periñán-Pascual y Mairal-Usón 2009b, 2010b). One way to extend the conceptual component of *FunGramKB*, represented in the core ontology, is the creation of satellite ontologies. In the present contribution we provide a methodology for the management and recovery of terminological units from the operation of specialized corpora (Felices-Lago, Ureña Gómez-Moreno y Alameda-Hernández 2011). The so-called modeling of ‘mirror concepts’ is proposed as a strategy for reusing the conceptual information of the core ontology. To illustrate the terminological-conceptual analysis, this contribution provides examples from the domain of organized crime and terrorism.

Key words: *FunGramKB*, knowledge base, natural language processing, core ontology, satellite ontology, mirror concept.

1. Introducción

*FunGramKB*² (*Functional Grammar Knowledge Base*) 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, traducción automática, etc) y en diversas lenguas³. Por otro lado, *FunGramKB* está formada por tres

¹Esta contribución 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.

² <http://www.fungramkb.com>

³ *FunGramKB* ha sido diseñada para trabajar con siete lenguas: alemán, búlgaro, catalán, español, francés, inglés e italiano.

niveles principales de conocimiento, esto es, léxico, gramatical y conceptual con módulos independientes pero interrelacionados a la vez (Periñán-Pascual y Arcas-Túnez 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 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 subontologías satélites.
- 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 un gazpacho.
- El Onomasticón almacena conocimiento enciclopédico, sobre instancias de entidades y eventos como Steve Jobs o el 11-S. 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* cada lengua tiene su propio módulo léxico y gramatical, mientras que el módulo conceptual es compartido por todas las lenguas. Así, *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-Pascual y Arcas-Túnez 2007; Felices-Lago y Marín-Rubiales, en prensa).

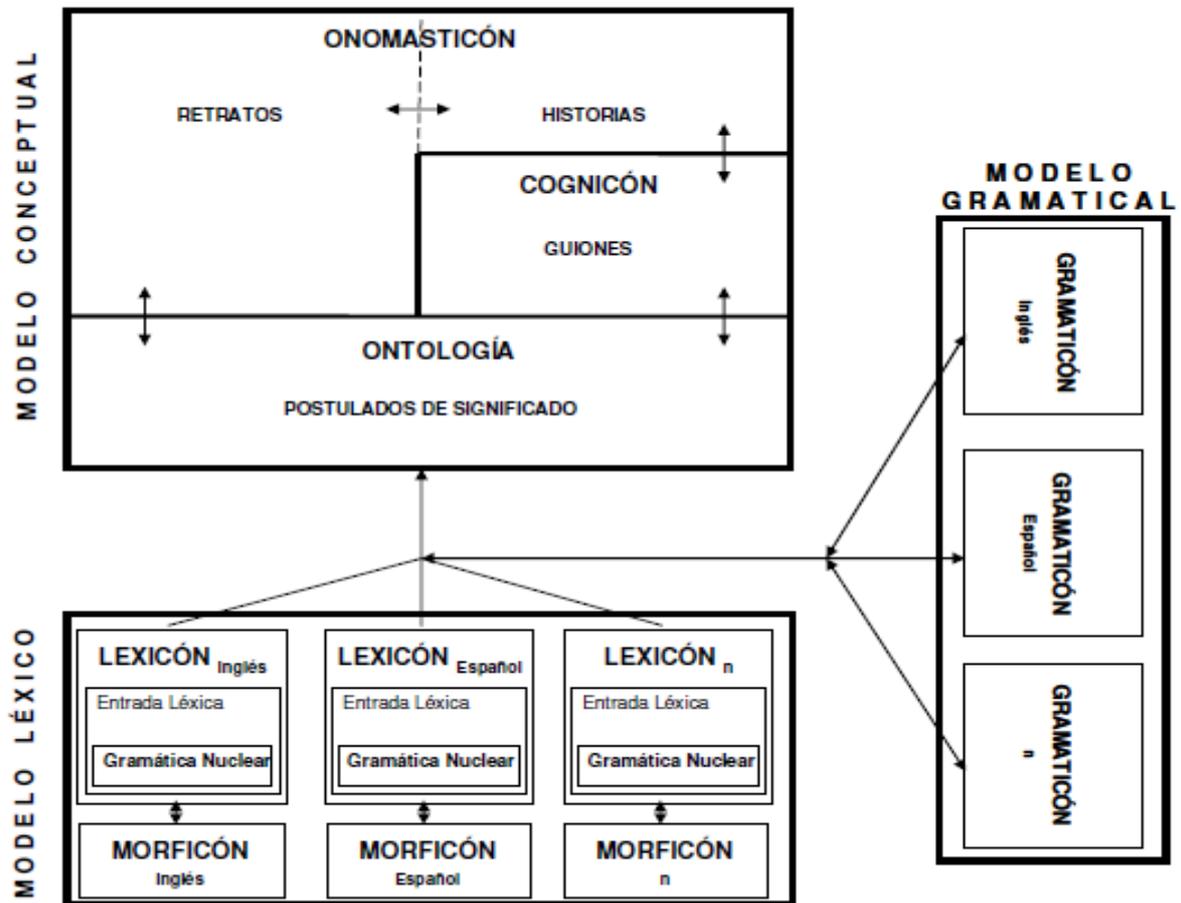


Figura 1: La arquitectura modular de FunGramKB (Periñán-Pascual y Arcas-Túnez 2011: 3).

Como *FunGramKB* tiene un propósito general y no 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 2 (cfr. Faber, Mairal y Magaña 2011: 564).

En esta contribución 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 2.2., mediante el modelado de los denominados ‘concepto espejo’ se reduce la redundancia informativa y se reutiliza la información conceptual de la Ontología Nuclear.

La estructura de este artículo es la siguiente. 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 y para finalizar ofrecemos dos ejemplos de conceptualización, uno de un ‘concepto espejo’ y otro de un término especializado.

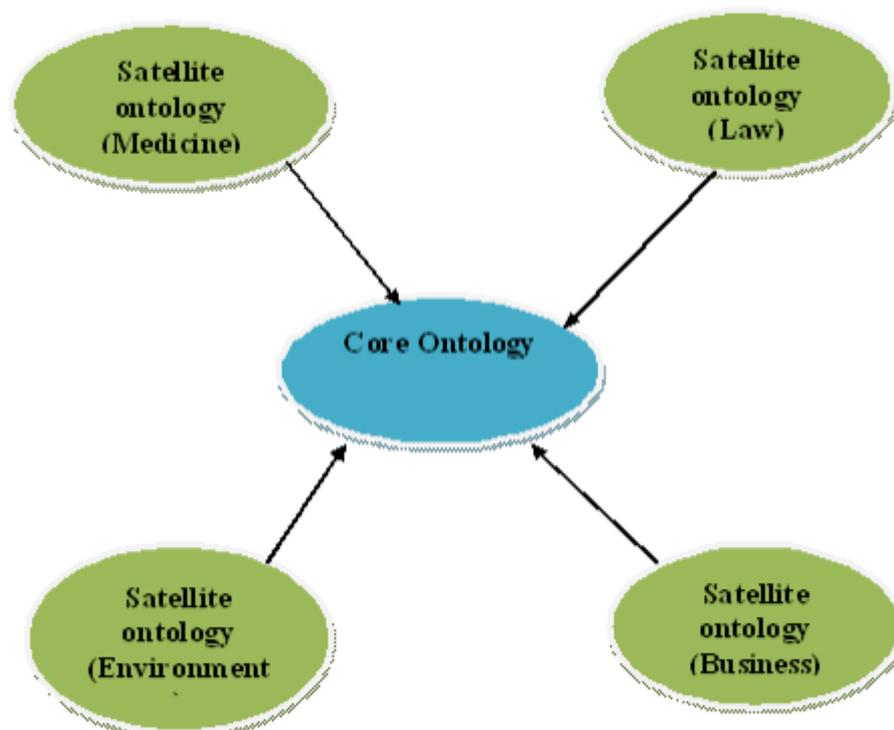


Figura 2: La ontología nuclear y las ontologías satélites (Mairal-Usón, Periñán-Pascual y Samaniego 2011: 5).

2. FunGramKB Term Extractor

FunGramKB Term Extractor es una herramienta semiautomática que facilita la labor del terminólogo en la selección de términos ganadores⁴. 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 (bigramas y trigramas) no terminológicas. Después, 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 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 candidatos especializados 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 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’, puesto que son los relativos a la extracción terminológica.

⁴ Véase la sección 2.1 para una definición de los conceptos ‘término ganador’ y ‘término falso’.



Figura 3: Pantalla principal del extractor de FunGramKB.

La herramienta ‘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. Sólo para el inglés se han subido al extractor 621 textos.

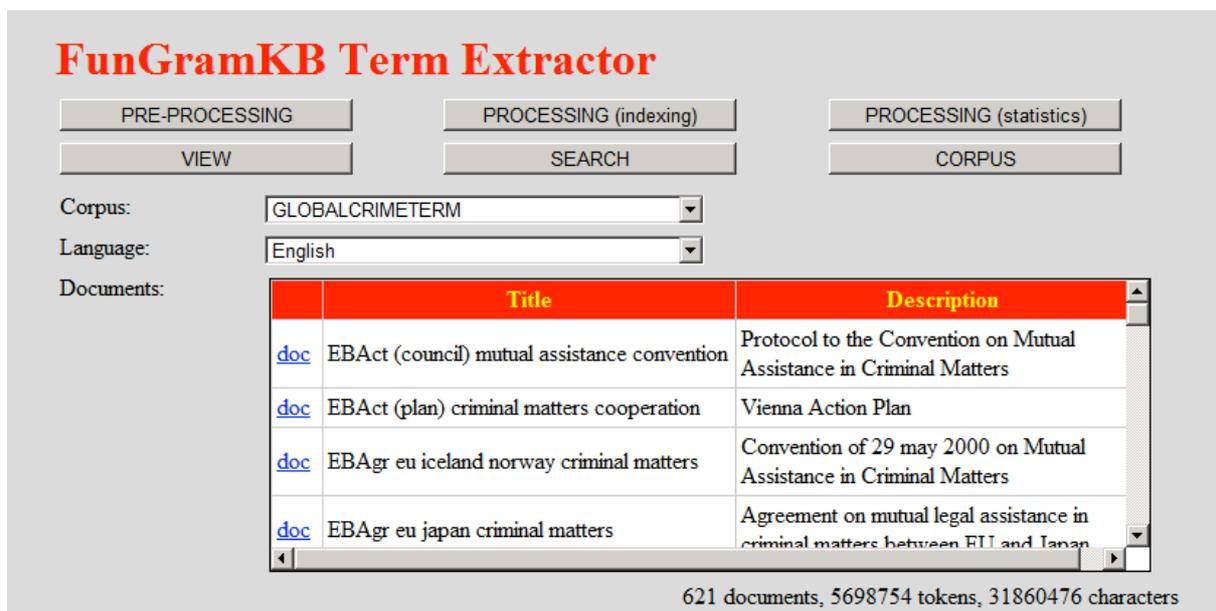


Figura 4: Pantalla de visualización del corpus.

El elemento ‘view’ nos permite ver la lista de términos candidatos. Podemos ver ejemplos de trigramas, bigramas y unigramas candidatos del dominio del crimen organizado y terrorismo en las Figuras 5, 6 y 7 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

	view	edit	<input type="checkbox"/>	Term	f	tf-idf	<input type="checkbox"/>	remove	nesting
	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 5: Tabla de trigramas 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

	view	edit	<input type="checkbox"/>	Term	f	tf-idf	<input type="checkbox"/>	remove	nesting
	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 6: Tabla de bigramas candidatos.

[Log out](#)

FunGramKB Term Extractor

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"/>	laundry	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 7: Tabla de unigramas candidatos.

2.1 El análisis terminológico

El terminólogo considerará término candidato aquel que, con la ayuda del índice de frecuencia del extractor, potencialmente sea relevante para un dominio léxico concreto. En el análisis de los términos candidatos el terminólogo ha de decidir cuáles son ganadores y cuáles falsos candidatos. Así, considerará, también con la ayuda del extractor, 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, ya que pertenece al lenguaje general (Felices-Lago, Ureña y Alameda-Hernández, 2011). En este caso serán de gran utilidad los Diccionarios para estudiantes (*Learners' Dictionaries*), ya que ayudarán al terminólogo a identificar las unidades terminológicas.

2.1.1. Criterios para desechar términos candidatos y selección de términos ganadores

El uso de diccionarios especializados es de gran utilidad para discernir las unidades léxicas (unigramas) 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 de los corpus de textos especializados subidos al extractor.

Para desechar un término candidato, además de analizar el índice de frecuencia tf-idf y ver si el término ya existe en la Ontología Nuclear, tendremos en cuenta los siguientes criterios.

- 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*).
- c) Para decidir si estamos ante un bigrama o trigrana propiamente dicho, nos aseguraremos de que su significado no sea una extensión de cada una de sus partes como ocurre en *suicide attack*, sino que se trata de un término con un significado completamente nuevo, como es el caso de *carousel fraud*:

- (1) *Suicide attack*: A **suicide attack** is a type of attack in which the attacker expects or intends to die in the process. (http://en.wikipedia.org/wiki/Suicide_Attack).
- (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).

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).

Como se trata de un término relativamente reciente *carousel fraud* no aparece aún en los diccionarios. Por otro lado, podemos encontrarnos con casos en los que el término aparece tanto en un *Learner's dictionary* como en los diccionarios especializados. Es 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 podría determinarse si hay fundamento o no para que esté en la Ontología Nuclear o en la Subontología Terminológica.

Hay casos especiales como *launder*, que, al consultar los *Learner's Dictionaries* como el *Cambridge* o el *Longman*, vemos que es ampliamente significativo que ya lo asocian a *money*. Eso nos va a indicar que se trata de un término especializado, 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 encontrar ya que sólo puede unir el significado de dos conceptos por separado, como es el caso de *hostage taking*, donde sí podría definir el bigrama a partir de *hostage* y *take*, ya que se trata de una mera colocación.

En el proceso de filtrado, el terminólogo desecha los términos que considera falsos candidatos. Estos permanecen en la papelera terminológica para su posterior recuperación si fuera necesario, como podemos ver en la Figura 8. Cuando se ha terminado el segundo y definitivo filtrado, entonces el terminólogo procederá a la limpieza de la papelera de reciclaje.



Figura 8: Papelera de reciclaje terminológica.

2.2. Conceptualización de términos ganadores

Una vez elaborada la selección de términos ganadores, el terminólogo ha de trabajar en la jerarquización y conceptualización de los mismos. Se trata de identificar las palabras definitorias del dominio temático que servirán de base para la definición de otros conceptos más específicos.

La Ontología Nuclear utiliza el lenguaje de representación COREL (COnceptual Representation Language) para modelar el significado de sus conceptos.⁵ 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 ‘#’ como resultado del análisis de las ontologías lingüísticas más relevantes como SUMO (Niles y Pease, 2001a, 2001b) o DOLCE (Gangemi et al., 2002; Masolo et al., 2003). Ejemplos de metaconceptos son #COMMUNICATION, #MATERIAL, #PSYCHOLOGICAL, etc).

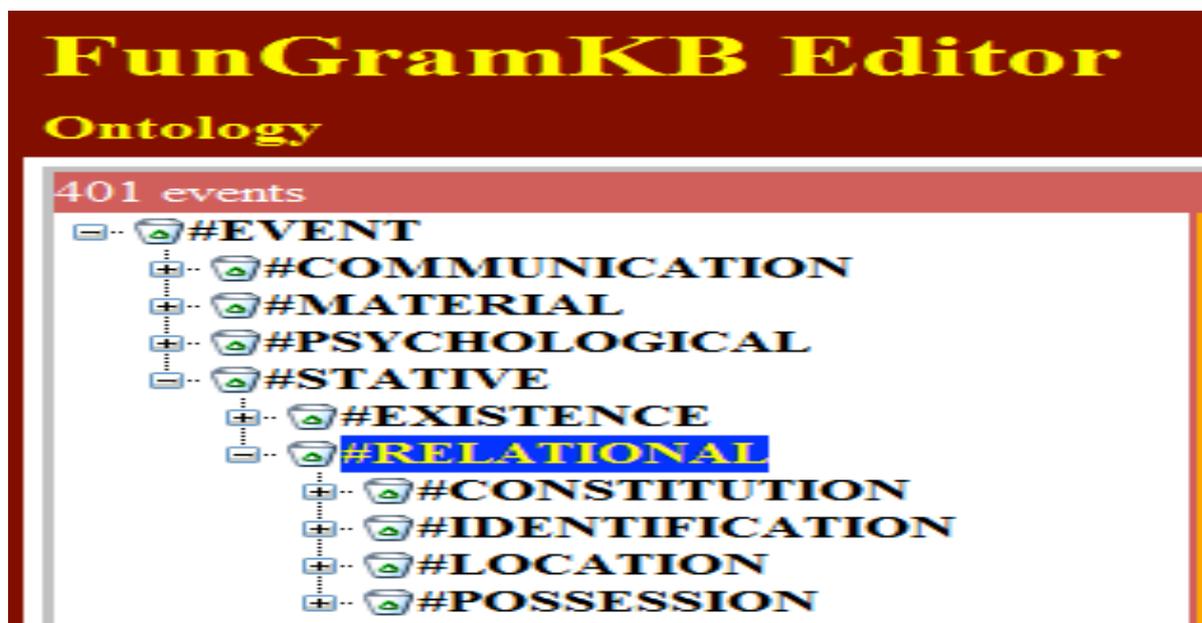


Figura 9: Ejemplos de metaconceptos en la ontología de FunGramKB.

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

Los conceptos básicos están en un nivel intermedio y se representan con el signo + y seguidos de un guión bajo y dos ceros (por ejemplo +JUDGE_00, +LAWYER_00, +TRIAL_00, etc.); y por último están los conceptos terminales, representados por el signo \$ y también seguidos de un guión bajo y dos ceros (por ejemplo \$MANKIND_00, \$PENALTY_00, \$SYNDICATE_00, etc.).⁶

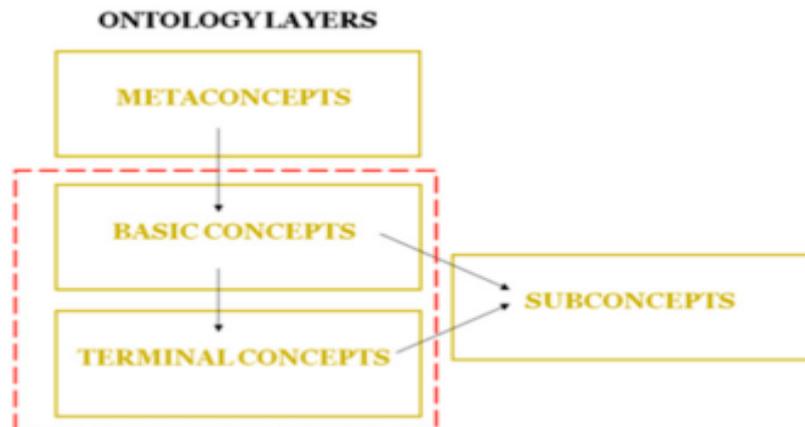


Figura 10: Jerarquía conceptual en FunGramKB (Jiménez-Briones, Luzondo-Oyón y Cabello de Alba, 2011b: 19).

Un ejemplo de definición en COREL es el de +PUNISHMENT_00, mencionado más abajo:
 +(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).

Tras organizar la estructura conceptual del dominio en cuestión formalizaremos la conceptualización de los términos. Para realizar esa tarea, el extractor dispone de una herramienta de edición de términos (Figura 11).

⁶ Cada unidad conceptual lleva asignada un índice numérico que junto con el lema en mayúsculas sirve para su identificación en el módulo conceptual de FunGramKB (cfr. Mairal-Usón y Perrián-Pascual, 2009: 227). Es decir, diferentes sentidos de un mismo concepto son codificados con distintos índices numéricos para evitar ambigüedades: 00, 01, 02, etc.

The image shows a software interface for editing terminology. It features a top section with three main areas: 'Senses:', 'CONCEPT:', and 'METACONCEPT:'. The 'CONCEPT:' section includes a 'DESCRIPTION:' field and a 'Save' button. The 'METACONCEPT:' section has radio buttons for 'Entities', 'Events', and 'Qualities'. Below these are language-specific sections for English, Spanish, Italian, French, German, Bulgarian, and Catalan, each with a text input field and 'Y'/'N' buttons. At the bottom left, there are 'Rename' and 'Delete' buttons, and checkboxes for 'Done' and 'Duplicate'. The word 'absorb' is displayed in the center.

Figura 11: Herramienta de edición de términos ganadores.

Mediante la selección manual de los términos ganadores el terminólogo elabora una jerarquía de conceptos básicos que conformarán lo que será su vocabulario defnitorio que le ayudará a definir los conceptos terminales, pero no podremos establecer la relación jerárquica de los términos ganadores hasta que no hayamos completado esta fase. En este proceso encontramos términos espejo de otros términos pertenecientes a la Ontología Nuclear, pero que en la Subontología Satélite del crimen organizado y el terrorismo adquirirán un significado distinto (como es el caso de +*PUNISHMENT_00* mencionado más abajo). Al consultar la Ontología Nuclear, vemos que *punishment* aparece ya recogido en ella, como podemos ver en la Figura 12.

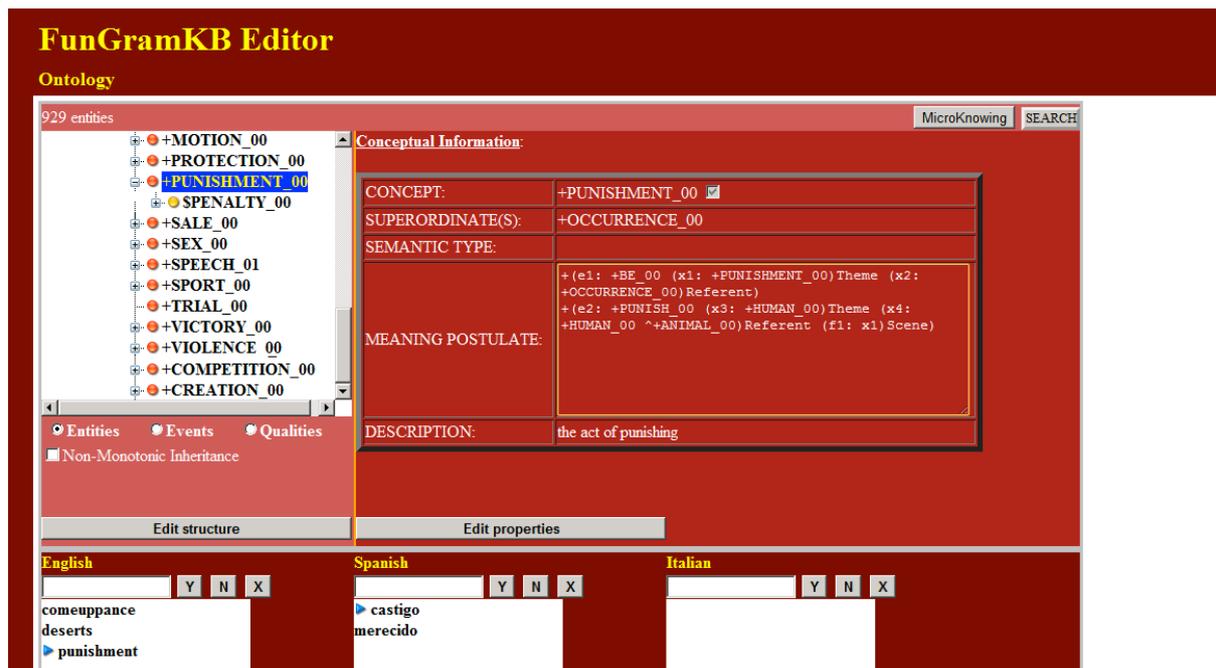


Figura 12: Información conceptual de +PUNISHMENT_00 en la Ontología Nuclear de FunGramKB.

En este caso, el significado de *punishment* será general o especializado según el diccionario que consultemos. El ejemplo (4) muestra una definición general de la palabra *punishment* obtenida del diccionario *Logman Dictionary of Contemporary English*, mientras que el ejemplo (5) muestra una definición especializada extraída del diccionario jurídico *Oxford Dictionary of Law*:

- (4) *Punishment*: something that is done in order to punish someone, or the act of punishing them.
- (5) *Punishment*: A penalty imposed on a defendant duly convicted of a crime by an authorized court. The punishment is declared in the *sentence of the court. The two basic principles governing punishment are *nullum crimen sine lege and *nulla poena sine lege. The powers of the court to punish offenders also depend on whether the crime is an *indictable offence or a *summary offence.

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, además de ser cuantitativamente más amplia. Podríamos decir, por tanto, que nos encontramos ante un caso de 'término espejo' ya que se trata de un término que está en la Ontología Nuclear, pero que en la Ontología Satélite del crimen organizado y terrorismo adquiriría un matiz más especializado. Por tanto, en el proceso de población de la ontología satélite del crimen organizado y el terrorismo tendríamos que crear el concepto +PUNISHMENT_00 que sería espejo de su equivalente en la ontología nuclear, pero con una formalización en COREL más específica que reflejaría su contenido especializado. Mediante el modelado de los conceptos espejo a través de los postulados de significado se desambiguan los distintos significados de un término.

Por otro lado, un ejemplo de término especializado sería *syndicate*. El *Black's Law Dictionary* lo define así:

- (6) *Syndicate*: 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.

Al consultar la Ontología Nuclear de *FunGramKB* vemos que *syndicate* aparece como concepto terminal \$SYNDICATE_00:

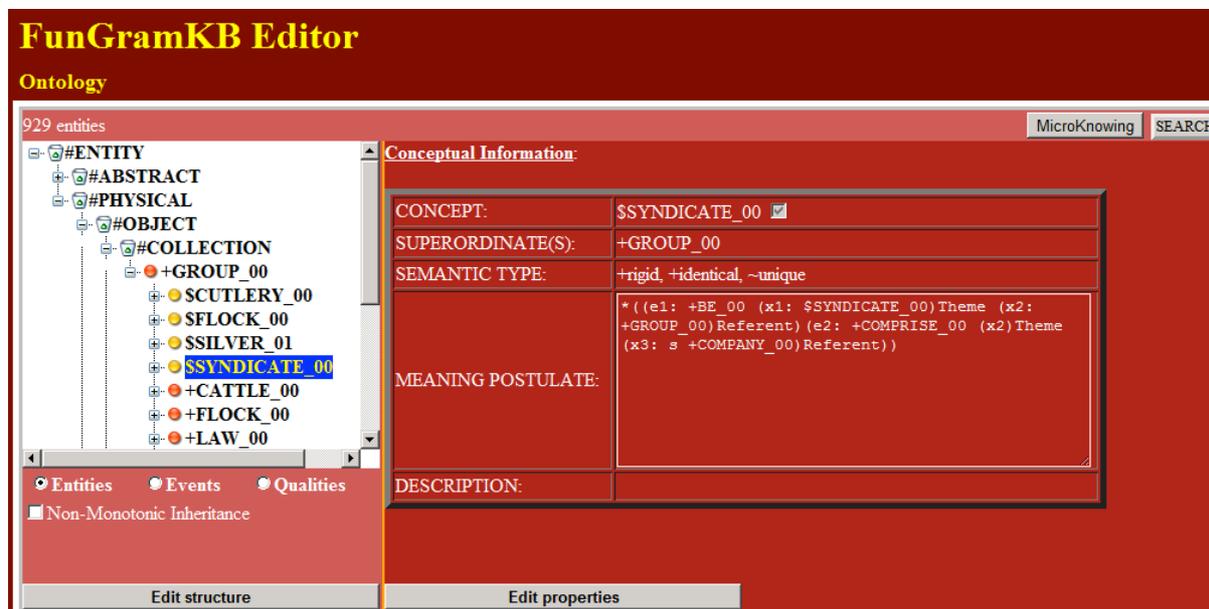


Figura 13: Información conceptual de \$SYNDICATE_00 en la Ontología Nuclear de FunGramKB.

Por su significado especializado, \$SYNDICATE_00 podría pasar a ser concepto básico en una Ontología Satélite del crimen organizado y el terrorismo. Este sería 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.

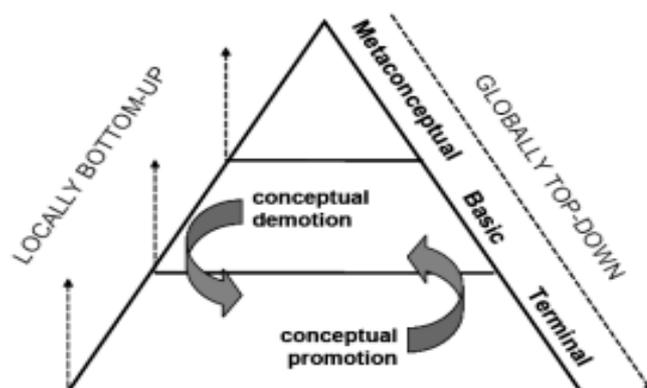


Figura 14: El diseño de la Ontología en FunGramKB (Periñán Pascual y Arcas Túnez, 2010b: 30).

3. Conclusiones

En este artículo hemos presentado *FunGramKB* como base de conocimiento léxico-conceptual multipropósito diseñada para la realización de tareas de procesamiento del lenguaje natural que puede ser reutilizada en diversas tareas (recuperación y extracción de información, traducción automática, etc) y en diversas lenguas. En concreto nos hemos centrado en el nivel conceptual de *FunGramKB*, y en la relación de la Ontología Nuclear con una posible Ontología Satélite del crimen organizado y el terrorismo.

Asimismo, hemos demostrado que *FunGramKB Term Extractor* es una 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. Además, hemos visto que en el proceso de modelado ontológico los ‘conceptos espejo’ y el modelo en espiral por el que se ha desarrollado la ontología nuclear nos sirven tanto para la reutilización de la información conceptual ya existente en la misma 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.

Referencias Bibliográficas

- Allen, James F. 1983. Maintaining knowledge about temporal intervals. In P. Adarraga and J.L. Zaccagnini (eds.), *Psicología e Inteligencia Artificial*. 141-186. Madrid: Trotta.
- Cambridge Learner's Dictionary. [<http://dictionary.cambridge.org/dictionary/learner-english/>].
- Faber, P., R. Mairal and 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). [Disponible en <http://aaai.org/ocs/index.php/FLAIRS/FLAIRS11/paper/view/2523/3107>]
- Felices-Lago, A. and A. Marín-Rubiales. (En prensa): La semántica profunda como fundamento para desarrollar una subontología jurídica en el contexto de las ontologías de ámbito legal. *Actas del XXIX Congreso Internacional de AESLA*.
- Felices-Lago, A., P. Ureña Gómez-Moreno and A. Alameda-Hernández. 2011. FunGramKB y la adquisición terminológica. *Anglogermánica Online* 2011: 66-86.
- Gangemi, A. et al. 2002. Sweetening ontologies with DOLCE. In: A. Gómez-Pérez and R. Benjamins (eds.), *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*: 13th International Conference, EKAW 2002, Sigüenza, 1-4.
- Garner, B.A., ed. 2009. Black's Law Dictionary. 9ª edición. USA.
- Jiménez-Briones, R. and A. Luzondo-Oyón. 2011. Building ontological meaning in a lexico-conceptual knowledge base. *Onomázein* 23: 11-40.
- Jiménez-Briones, R., A. Luzondo-Oyón and M. Beatriz Cabello de Alba. 2011. FunGramKB y la organización ontológica. *Anglogermánica Online* 2011: 16-36.
- Longman Dictionary of Contemporary English. [<http://www.ldoceonline.com/>].
- Mairal-Usón, R., C. Perrián-Pascual and E. Samaniego 2011. Using ontologies for terminological knowledge representation: a preliminary discussion. In N. Talaván, E. Martín Monje and F. Palazón (eds.), *Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10*. 267-280. UNED: Madrid.
- Mairal-Usón, R. and C. Perrián-Pascual. 2010. Role and Reference Grammar and Ontological Engineering. In J.L. Cifuentes, A. Gómez, A. Lillo, J. Mateo and F. Yus (eds.) *Los caminos de la lengua. Estudios en homenaje a Enrique Alcaraz Varó*. 649-665. Alicante: Universidad de Alicante.
- Mairal-Usón, R. and C. Perrián-Pascual. 2009. The anatomy of the lexicon component within the framework of a conceptual knowledge base. *Revista Española de Lingüística Aplicada* 22: 217-244.

- Niles, I. and A. Pease . 2001a. Origins of the Standard Upper Merged Ontology: a proposal for the IEEE Standard Upper Ontology. *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology*. Seattle.
- Niles, I. and A. Pease 2001b. Towards a Standard Upper Ontology. *Proceedings of the Second International Conference on Formal Ontology in Information Systems*. Ogunquit.
- Oxford Dictionary of Law*. 2009. 9ª edición. Oxford: Oxford University Press.
- Periñán-Pascual, C. and F. Arcas-Túnez. 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-Pascual, C. and F. Arcas-Túnez. 2005. Microconceptual-Knowledge Spreading in FunGramKB. *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*. 239-244. Anaheim-Calgary-Zurich: ACTA Press
- Periñán-Pascual, C. and F. Arcas-Túnez. 2007. Deep semantics in an NLP knowledge base. *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence held at the Universidad de Salamanca*. 279-288.
- Periñán-Pascual, C. and F. Arcas-Túnez. 2010a. The architecture of FunGramKB. *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, European Language Resources Association (ELRA). 2667-2674.
- Periñán-Pascual, C. and F. Arcas-Túnez 2010b. Ontological Commitments in FunGramKB. *Procesamiento del Lenguaje Natural* 44: 27-34.
- Periñán Pascual, C. and R. Mairal Usón. 2011 The COHERENT methodology in FunGramKB. *Onomázein* 24: 13-33.
- Periñán-Pascual, C. and F. Arcas-Túnez. 2011. Introducción a FunGramKB. *Anglogermánica Online* 2011: 1-15.
- Periñán-Pascual, C. and R. Mairal-Usón. 2009. Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural* 43: 265-273.
- Periñán-Pascual, C. and R. Mairal Usón. 2010a. Enhancing UniArab with FunGramKB. *Procesamiento del Lenguaje Natural* 44: 19-26.
- Periñán-Pascual, C. and R. Mairal-Usón. 2010b. La gramática de COREL: un lenguaje de representación conceptual. *Onomázein* 21: 11-45.
- Van Valin, R. 2005. *Exploring the Syntax-Semantic Interface*. Cambridge: Cambridge University Press.
- Van Valin, R. and R. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.