Alameda Hernández, Ángela and Felices Lago, Ángel (2017) Crimes against children: an apparently terminological knowledge representation of entities in FunGramKB. Journal of Computer-Assisted Linguistic Research 1: 1-19.

Crimes against children: an apparently terminological knowledge representation of entities in FunGramKB

Ángela Alameda Hernández*, Ángel Felices Lago
University of Granada, Spain
*Corresponding author: aalameda@ugr.es

Received: 27 February 2017 / Accepted: 28 April 2017 / Published: 19 June 2017

Abstract
This article describes an example of the difficulties involved in the construction of a term-based satellite (or domain-specific) ontology integrated in FunGramKB –a lexico-conceptual knowledge base for the computational processing of natural language (Periñán-Pascual & Arcas-Túnez 2004, 2007, 2010a; Periñán-Pascual & Mairal-Usón 2009, 2010). The main hypothesis is that the multilevel model of FunGramKB Core Ontology can be connected to terminological satellite ontologies in order to minimize redundancy and maximize information (Periñán-Pascual & Arcas-Túnez 2010b). To this end we follow the four-phase COHERENT methodology (Periñán-Pascual & Mairal-Usón 2011): COnceptualization, HiErarchization, REmodelling and refinemeNT. In doing so, the paper furnishes substantial evidence on the structuring of a set of concepts borrowed from criminal law, an apparently terminological domain (cf. Breuker, Valente & Winkels 2005; Valente 2005; Breuker, Casanovas & Klein 2008). The *Globalcrimeterm* corpus has been used as an empirical foundation (Felices-Lago & Ureña-Gómez Moreno 2012, 2014). To illustrate this process, we have selected the superordinate basic concept +CRIME_00 (Alameda-Hernández & Felices-Lago 2016) and its basic and terminal subordinate concepts in the domains of organized crime and terrorism (all of them under the metaconcept #ENTITY), particularly those crimes referring predominantly to children or involving children with other vulnerable groups. The creation of specific definitions for the target concepts in this paper uses COREL (a conceptual representation language (Periñán-Pascual & Mairal-Usón 2010)) and the following upper-level conceptual path: #ENTITY> #PHYSICAL> #PROCESS> +OCCURRENCE_00> +CRIME_00. Consequently, the modelling, subsumption and hierarchisation of concepts such as $ABDUCTION_00, $CHILD_ABUSE_00, $CHILD_PORNOGRAPHY_00, $COERCE_D_00, $CHILD_TRAFFICKING_00, $MOLEST_D_00, $FORCED_LABOUR, among others, are presented.

DOI: https://doi.org/10.4995/jclr.2017.7785
PUBLISHER: UNIVERSITAT POLITÈCNICA DE VALÈNCIA.

Keywords: FunGramKB, terminology, knowledge representation, crimes against children, ontology building, legal ontology

1. INTRODUCTION

Over the last five years, a few contributions have been made in an attempt to connect the Core Ontology of the knowledge base known as FunGramKB (Periñán-Pascual & Arcas-Túnez 2010a,b) to other domain-specific ontologies in order to minimize redundancy and maximize information (Felices-Lago 2015, 2016; Felices-Lago & Ureña Gómez-Moreno 2012, 2014; Faber, Mairal-Usón & Magaña-Redondo 2011; Periñán-Pascual & Arcas-Túnez 2014; Alameda-Hernández & Felices-Lago 2016; Felices-Lago & Alameda-Hernández (forthcoming 2017)).

In the following lines we will focus our attention on the *Globalcrimeterm* Satellite Ontology (integrated in FunGramKB) and, more precisely, on the structuring of a set of concepts borrowed from the apparently terminological domain of criminal law, particularly those units included in the *Globalcrimeterm* corpus (Felices-Lago & Ureña-Gómez Moreno 2012, 2014). To illustrate this process, we have selected the superordinate basic concept +CRIME_00 (Alameda-Hernández & Felices-Lago 2016) and its basic and terminal subordinate concepts in the domain of organized crime (all of them under the metaconcept #ENTITY), particularly those crimes involving children or children in connection with other vulnerable groups. Section 2 will deal with the theoretical background of both FunGramKB and *Globalcrimeterm*. Section 3 will explain the four-phase COHERENT methodology (Periñán-Pascual & Mairal-Usón 2011): [COnceptualization, HiErarchization, REmodelling and refinemeNT] and will show how the superordinate basic concept +CRIME_00 (Alameda-Hernández & Felices-Lago 2016) has been selected. Section 4 will show the hierarchisation process stemming from +CRIME_00 and its basic and terminal subordinate concepts (all of them under the metaconcept #ENTITY), particularly those crimes referring predominantly to children or involving children with other vulnerable groups. Section 5 will offer some conclusions.

2. THEORETICAL BACKGROUND

2.1. FunGramKB

In the last few years, the comprehensive theory of constructional meaning known as the Lexical Constructional Model (Mairal-Usón & Ruiz-de-Mendoza 2008, 2009; Ruiz-de-Mendoza & Mairal-Usón 2008, among others) has incorporated as part of its architecture *FunGramKB* (FGKB), which is a multipurpose lexico-conceptual knowledge base for natural language processing (NLP) systems (Periñán-Pascual & Arcas-Túnez 2004; Periñán-Pascual & Arcas-Túnez 2005; Mairal-Usón & Periñán-Pascual 2009; Mairal-Usón & Periñán-Pascual 2010; Periñán-Pascual & Mairal-Usón 2009; Periñán-Pascual & Mairal-Usón 2010). It is multipurpose in the sense that it is both multifunctional and multilingual. In other words, FunGramKB can be reused in various NLP tasks (e.g. information retrieval and extraction, machine translation, dialogue-based systems, etc.) and with several natural languages. This knowledge base comprises three major knowledge levels, consisting of several independent but interrelated

modules: (1) Lexical level: *The Lexicon* stores morphosyntactic, pragmatic and collocational information about words. *The Morphicon* helps our system to handle cases of inflectional morphology. (2) Grammatical level: *The Grammaticon* stores the constructional schemata which take part in the bi-directional linking algorithm: semantics <-> syntax. (3) Conceptual level: *The Ontology* is presented as a hierarchical catalogue of the concepts describing semantic knowledge.[1] *The Cognicon* stores procedural knowledge by means of script-like schemata in which a sequence of stereotypical actions is organised on the basis of temporal continuity. *The Onomasticon* stores information about instances of entities and events. In FunGramKB, every lexical or grammatical module is language-dependent, whereas every conceptual module is shared by all languages. FunGramKB adopts a conceptualist approach to language, where the ontology becomes the pivotal module for the whole architecture. The concepts of FunGramKB belong to three levels. FunGramKB ontology can be split into three subontologies: #ENTITY for nouns, #EVENT for verbs, and #QUALITY for adjectives (and some adverbs). In this paper we shall concentrate on the subontology #ENTITY. Within each one of the three subontologies, FunGramKB also distinguishes three categories of concepts organized hierarchically and defined through a machine-readable metalanguage called COREL (i.e. *Conceptual Representation Language*):

(a) Metaconcepts (e.g. #ABSTRACT, #COLLECTION, #EMOTION, #TEMPORAL, etc.), They are headed by the symbol # and form the upper level in the taxonomy, as a result of the analysis of the most relevant linguistic ontologies, i.e. DOLCE (Gangemi et al. 2005), SIMPLE (Lenci 2008), SUMO (Niles & Pease 2001), etc.
(b) Basic concepts, preceded by symbol +, are used as defining units which enable the construction of Meaning Postulates (MPs) for basic concepts and terminals, as well as taking part as selection preferences in Thematic Frames (TFs): e.g. +FORGET_00, +HAND_00, +MOVE_00, +DIRTY_00, etc. They can be employed to define any word in any of the European languages that are claimed to be part of the Ontology. The starting point for the identification of basic concepts was the defining vocabulary in *Longman Dictionary of Contemporary English* (Procter 1978), though deep revision was required in order to perform the cognitive mapping into a single inventory of about 1,300 basic concepts.
(c) Terminal concepts, which are headed by symbol $. Terminals are not hierarchically structured and do not have definitory potential to take part in MPs: e.g. $GRASP_00, $ABDUCTION_00, $BOW_00, $VARNISH_00, $PERCEPTIBLE_00, etc.

Basic and terminal concepts in FunGramKB are provided with semantic properties which are captured by *thematic frames* (TFs) and *meaning postulates* (MPs). Every event in the ontology is assigned one single thematic frame, i.e. a conceptual construct which states the number and type of participants involved in the prototypical cognitive situation portrayed by the event (Periñán-Pascual & Arcas-Túnez, 2007). Moreover, a meaning postulate is a set of one or more logically connected predications (e_1, e_2, … e_n), i.e. conceptual constructs that represent the

[1] FunGramKB Core ontology is deemed as an IS-A conceptual hierarchy which allows non-monotonic multiple inheritance. This ontology is both universal and linguistically-motivated.

generic features of concepts. As stated above, the basic concepts are the main building blocks of these types of constructs in the Core Ontology.

2.2. Globalcrimeterm

This subontology can be defined as a hierarchical taxonomy of specialised concepts belonging to an expert area of knowledge: basically, an area of criminal law (Felices-Lago 2015). It thus serves the purpose of enhancing FunGramKB with specialised knowledge, as the knowledge base has been so far implemented to work with elementary common-sense concepts of human cognition. The building of this subontology involved the compilation of the *Globalcrimeterm* Corpus (GCTC), which ultimately aimed at the purpose of structuring information for easy retrieval in professional contexts and the solving of problem-oriented tasks in real situations (Bowker & Pearson 2002; Koester 2010). For this purpose, a semi-automatic population of the ontology and its hierarchical structure have proved essential, involving the building of relevant meaning postulates and the creation of terminal concepts; this process has been followed by the semi-automatic population of specialised lexica for the languages chosen (English and Spanish), which cover relevant lexical information and which can store hundreds of lexical units into each lexicon. In order to feed the Satellite Ontology (or Subontology) and the Lexica for the two languages, terminological units have been obtained from documentary and textual databases offered from reliable reference sources, such as regulations, treaties, articles, books, glossaries or previous legal ontologies provided by international agencies and institutions which work on criminal law or the fight against organized crime such as United Nations, Criminal Court of Justice, EUROPOL, EUROJUST or CDPC (*European Committee on Crime Problems*), among others. These organizations and their legal representatives are leaders in the field of the fight against organised crime and terrorism and offer a rich representation of the expert knowledge and specialised vocabulary that is officially used in a vast array of reports, acts and other documents. In total, this domain-specifc corpus comprises 621 documents and 5,698,754 tokens.

Apart from that, the design of *FunGramKB Term Extractor* (Periñán-Pascual & Arcas-Túnez 2014; Felices-Lago & Ureña Gómez-Moreno, 2014) has been the pivotal element integrated in *FunGramKB Suite* to deal with the *Globalcrimeterm* corpus (Felices-Lago & Ureña Gómez-Moreno, 2012). This is the basic instrument which has allowed not only the automatic identification of candidate terms according to their probabilistic weight, but also the technical support to terminologists to choose the relevant terms for the Satellite Ontology.

Once the whole process of automatic extraction and filtering was taken, the total number of relevant n-grams was dramatically reduced from 5,698,754 tokens to 57,502. On the other hand, identifying a close list of winning terms was not a straightforward task; rather, it raised a number of theoretical problems and complex decisions. To facilitate term identification, a four-criterion methodology was proposed to be considered by terminologists during the manual filtering process. These four criteria are presented in a sequential and logical order: i) statistical significance; ii) ontological grounding; iii) lexicological features and iv) consultation of specialised dictionaries (Felices-Lago & Ureña Gómez-Moreno 2014: 264-266). The final number of winning conceptual term candidates was 406. Among these winning units, one of the most

relevant and representative concepts in the domain-specific ontology under scrutiny was +CRIME_00.

3. THE COHERENT METHODOLOGY

The four-phase COHERENT methodology (COnceptualization + HiErarchization + REmodeling + refinemeNT) designed by Periñán & Mairal (2011) was used for the construction of the basic conceptual level of FunGramKB Core Ontology and could also serve as a methodological basis for the development of satellite ontologies linked to it (Carrión-Delgado 2012). In an initial step, the basic concepts of the Longman Defining Vocabulary (LDV) of the *Longman Dictionary of Contemporary English* (Procter 1978) were identified. Although the LDV has been proven as a benchmark in the development of a basic vocabulary of an artificial language, it was necessary to do a major review to form the conceptual map. Specifically, both the population and the basic conceptual structure of the Core Ontology were developed manually following the COHERENT methodology in the four phases . As a result of applying this method, a catalogue of approximately 1,300 basic concepts has been a stepping stone to populating the Core Ontology with terminal concepts, a process still in progress. In fact, the end node of the conceptual hierarchy has been enriched with the integration of terminal concepts of the criminal law domain from *Globalcrimetern*. If we take the basic concept +CRIME_00 as a sample, it will be a key concept to understand the diversity of an essential branch of the law.

In the following lines we are going to introduce the first two phases of the COHERENT method (conceptualization and hierarchization). The remaining two phases (remodelling and refinement) require further research. Remodelling generally affects verbal concepts under the #EVENT subontology, but not entities. As for refinement, the conceptual productivity of both, +CRIME_00 and +PERSONAL_CRIME_00, is guaranteed by a large percentage of subordinate concepts.

3.1. Conceptualisation of the concept +CRIME_00

The first impression that we get when we think of the concept "crime" is that no one needs to be an expert to know, in general terms, what it means. Thus, even if perfectly obvious, it seems relevant to point out that the concept "crime" is part of a person's general knowledge. As such, "crime" is commonly understood to be "an evil act, punishable by law". In *Black's Law Dictionary*, for instance, we find that "crime" is defined as: "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".

In addition, together with this definition, the dictionary entry for "crime" includes more than fifty subentries in order to thoroughly analyze and present what this concept entails for the expert user of the dictionary and hence, in the specialized field. Thus, the concept "crime" is not exclusive to the legal field, but it is definitely perceived differently depending on whether it is being used by a lay person or a legal expert. It follows, then, that general and specialized knowledge are not two independent worlds, rather they are deeply connected (Alameda-

Hernández & Felices-Lago 2016). However, the general features of most crimes are understood and used by the lay person and that is the reason why we have eventually included most concepts related to specific crimes in the FunGramKB Core Ontology (Felices-Lago 2016) and not in the *Globalcrimeterm* subontology.

In order to gain more insight into this complex process, in Alameda-Hernández and Felices-Lago (2016) it was stated how it is necessary to scrutinize the representation of the concept "crime" and its conceptual path in the Core Ontology. Following FunGramKB notation system, known as COREL, the concept "crime" is represented as +CRIME_00. As a basic concept, it is preceded by the symbol [+]. For +CRIME_00, the conceptual path is as follows: #ENTITY > #PHYSICAL > #PROCESS > +OCCURRENCE_00 > +CRIME_00 (see figure 1 below). Hence, it belongs to the "#ENTITY Subontology" and to the metacognitive dimensions #PHYSICAL and #PROCESS. Its immediate superordinate concept is +OCCURRENCE_00. It means that +CRIME_00 inherits all the conceptual properties of +OCCURRENCE_00. The MP of +CRIME_00, as notated in COREL interface language, is as follows:

+(e1: +BE_00 (x1: +CRIME_00)Theme (x2: +OCCURRENCE_00)Referent)
+(e2: n +BE_00 (x1)Theme (x3: +LEGAL_00)Attribute)

This MP is made up of two predications (headed by the symbol "e") whose translation into natural language can be 1. "Crime is an occurrence" and 2. "It is something which is not legal". The operator [n] in the second predication negates the concept that follows, namely, +BE_00. Moreover, this concept has other subordinates, either basic or terminal, since it has further specificity in the repertoire of concepts of a lay person, as will be explained below

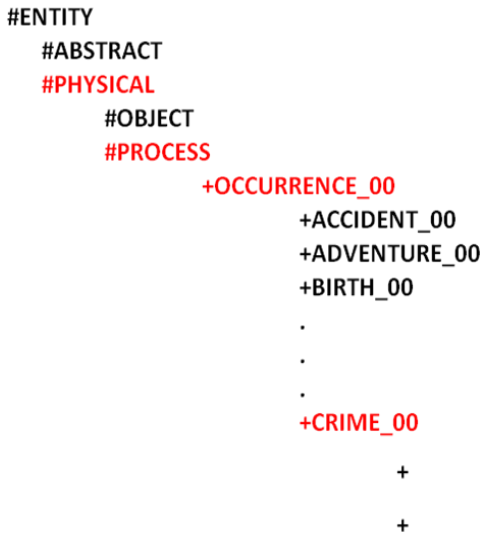

FIGURE 1. CONCEPTUAL PATH OF +CRIME_00 IN FUNGRAMKB CORE ONTOLOGY

3.2. Hierarchisation of the concept +CRIME_00

The new stage to be reached in the COHERENT methodology is the hierarchisation phase, which deals with the establishment of hierarchical meaning relations among concepts in the domain. Designing a networked hierarchy will endow FunGramKB with the capacity to derive relevant and meaningful inferences, as well as to *understand* and produce knowledge for a specific user-defined goal. The present section deals with the details of conceptual-hierarchy construction.

Hierarchisation consists of determining for each concept defined in FunGramKB its corresponding hyperordinate, subordinate(s) and sister concept(s). Hyperordinates are the most general type of units in the hierarchy and work as host concepts for the classification of one or more subordinate concepts. Each subordinate concept can in turn have one or more sister concepts, which are characterised by sharing common semantic features inherited from the hyperordinate. This arrangement of concepts is called the "IS-A" subsumption.

The basis for the development of the hierarchical organization of the concept +CRIME_00 in this work has been the *Globalcrimeterm* ontology. This repository has taken into account previous terminological and corpus-based work (Felices-Lago, Ureña Gómez-Moreno & Alameda Hernández 2011; Ureña Gómez-Moreno Alameda-Hernández & Ángel Felices-Lago 2011), retrieving a list of terms linked to crimes prototypically associated with organized crime and terrorism, such as: *assault, mayhem, fencing, smuggling, embezzlement, phishing, trafficking, slavery, chattel slavery, cuckoo smurfing, tax evasion, waterboarding, bribery, collusion* or *abduction*, among others (Alameda-Hernández & Felices-Lago 2016). On the whole, 66 terms were listed. However, for visualizing the conceptual organization of this criminal area, with special emphasis on the crimes against children, the first step was a thorough analysis of the lexicographical sources available. Various dictionaries were extensively consulted. These included both general and specialized dictionaries, as well as monolingual and English-Spanish bilingual ones, such as *Cambridge Advanced Learner's Dictionary, Longman Dictionary of Contemporary English, Oxford Advanced Learner's Dictionary, DRAE: Diccionario de la Lengua Española* (Real Academia), *CLAVE*, Alcaraz-Varó bilingual (English-Spanish) dictionaries of legal terms, *Black's Law Dictionary, Routledge Dictionary of Terrorism,* and *Oxford Law Dictionary*. Similarly, various academic sources (particularly, specialized journals and academic reference works) and the official websites of institutions concerned were consulted, not only dealing with general legal aspects but also related to questions of organized crime (such as OSCE, UN, INTERPOL, Eurojust, etc.). All this work helped us to determine and clarify the conceptual organization of the concept "crime" on this ground. As a result, two broad umbrella concepts were identified: ***personal crimes*** and ***property crimes***: those offenses which affect persons and those which deal with property.

In turn, each one of these two groups is itself divided into subgroups: ***personal crimes*** is divided into *crimes which affect physical integrity, crimes against freedom, trafficking of human beings* or *sexual content offenses*; ***property crimes*** is divided into *trafficking, theft, financial crimes and*

corruption or *cybercrimes*. Each type of crime is included in one of these subgroups. For example: *physical integrity*, which is a subgroup of ***personal crimes***, includes *assault, mayhem, torture,* and *waterboarding*; another example is *financial crimes*, which is a subgroup of ***property crime*** and includes *money laundering, cuckoo smurfing, tax evasion,* and *layering*. Thus, "crime" is itself an umbrella concept that encompasses a set of other concepts that are semantically associated.

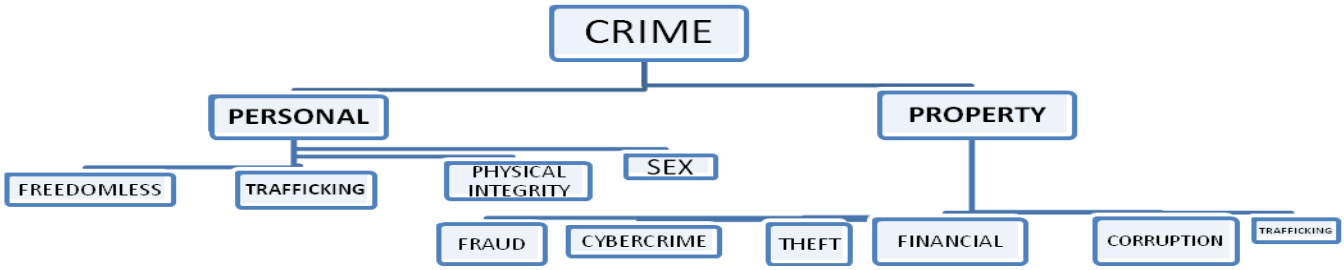

FIGURE 2: TYPOLOGY OF ORGANIZED CRIME OFFENSES UNDER +CRIME_00

As a conceptual entry in FunGramKB, the basic concept +PERSONAL_CRIME_00 can be explained as follows: its immediate hyperordinate concept is +CRIME_00. It means that +PERSONAL_CRIME_00 inherits all the conceptual properties of +CRIME_00. The MP of +PERSONAL_CRIME_00, as notated in COREL interface language, is as follows:

+(e1: +BE_00 (x1: +PERSONAL_CRIME_00)Theme (x2: +CRIME_00)Referent)
+(e2: +DO_00 (x3: +CRIMINAL_00)Theme (x4: +HUMAN_00)Referent (f1: DAMAGE_00 (x3)Theme (x4)Referent)Purpose)

This MP is made up of two predications (headed by the symbol "e") whose translation into natural language can be 1. "Personal crime is a type of crime" and 2. "It is something done by a criminal to cause harm to a human being" (see figure 3 below). Moreover, +PERSONAL_CRIME_00 remains as a basic concept because it has subordinates, either basic or terminal, since it is very productive.

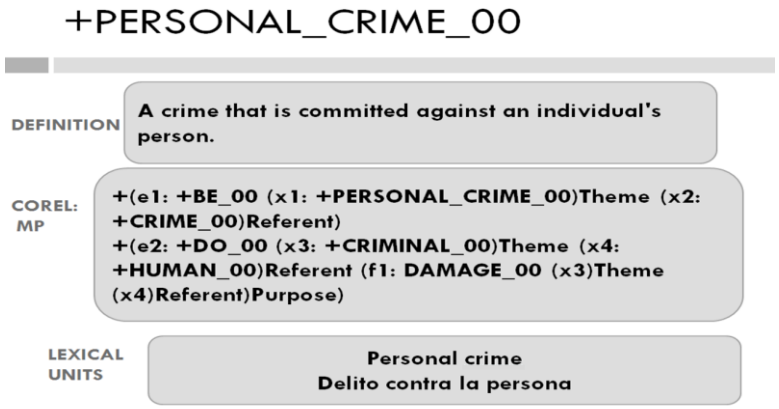

FIGURE 3: REPRESENTATION OF THE CONCEPT +PERSONAL_CRIME_00

4. CRIMES AGAINST CHILDREN

According to INTERPOL, the crimes against children tend to be local crimes with the vast majority taking place within the home or family circle. However, international organized crime groups tend to be very active in certain offenses that we have included under the umbrella concept +PERSONAL_CRIME_00 in FunGramKB Core Ontology (see figure 4) and are generally addressed to children and/or other vulnerable groups.

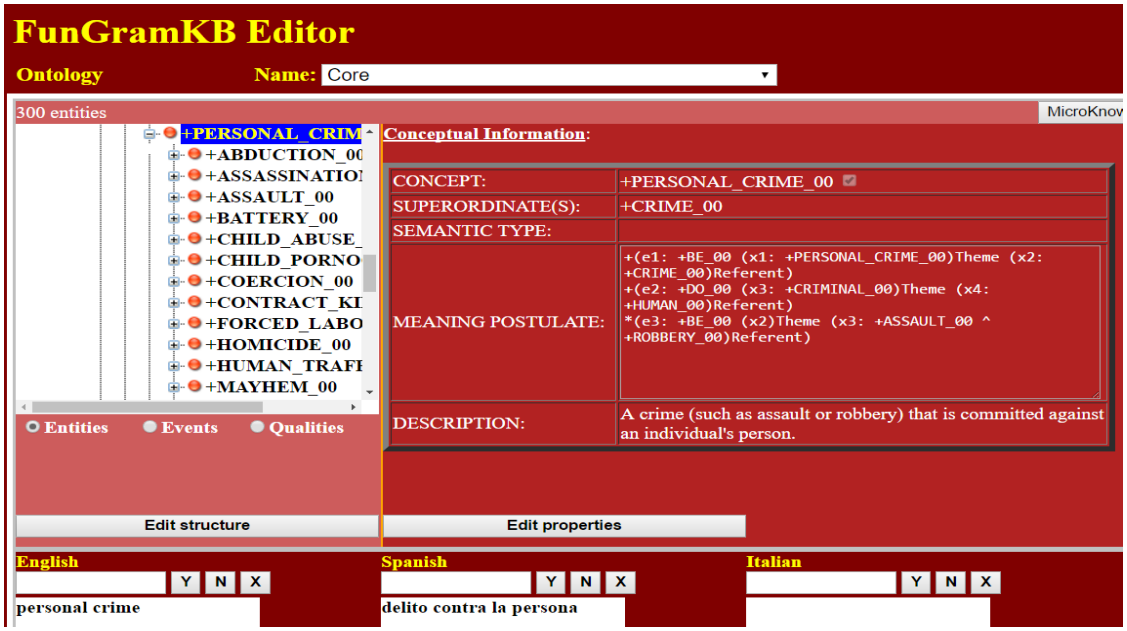

FIGURE 4: REPRESENTATION OF THE CONCEPT +PERSONAL_CRIME_00 AND ITS SUBORDINATES IN FUNGRAMKB EDITOR.

These crimes against children are distributed among the typology described in 3.2: *crimes which affect physical integrity, crimes against freedom, trafficking of human beings* or *sexual content offenses*. However, this typology is not relevant for the purpose of conceptual modeling, because the characteristics of each type of crime become explicit in the MPs of the corresponding concepts and integrated in their predications (e_1 e_2 e_3 ...) and their satellites (f_1 f_2 f_3 ...). Consequently, in the process of refinement and filtering of the original list of candidate terms extracted from the *Globalcrimeterm* corpus, the following result was obtained: $CHILD_ABUSE_00, $CHILD_PORNOGRAPHY_00 $CHILD_PROSTITUTION_00$ and $CHILD_TRAFFICKING_00 are the four sister concepts that are addressed directly to children (see the central oval circle in figure 5) . However, other criminal concepts are not linked to children only, but they are frequently associated to them (and other vulnerable groups). That is the case $ABDUCTION_00, $COERCE_D_00, $FORCED_LABOR_00, $MOLEST_D_00, $ORGAN_TRAFFICKING_00 or $SLAVERY_00. These sister concepts could share a common ground with crimes associated with children, but not so frequently linked to organized crime, as is the case of ´infanticide`, ´domestic violence` or ´bullying` (see the larger oval circle in figure 5). Finally, other crimes might be occasionally or rarely linked to minors (see the conceptual units outside the oval circles in figure 5).

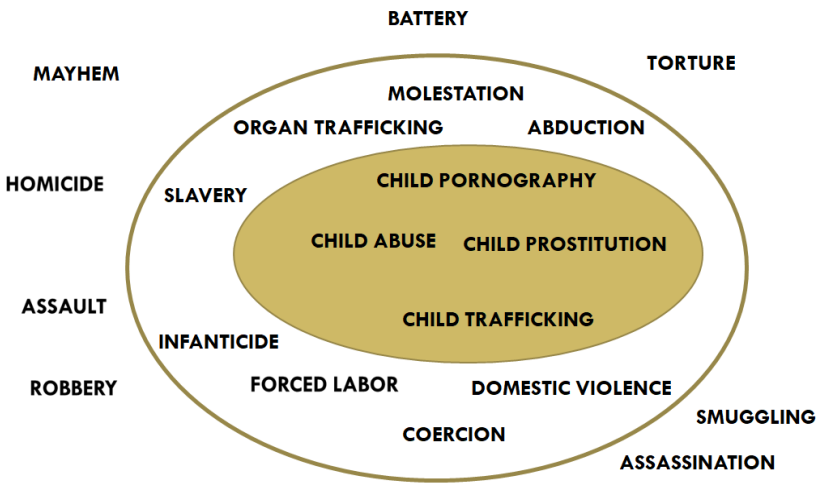

FIGURE 5: CRIMES AGAINST CHILDREN AT VARYING DEGREES

All the concepts above cannot be considered expert knowledge as they can easily be found in learner`s dictionary definitions. As Van Dijk (2003: 27) pointed out, "at first this knowledge may still be specialized, but at least part of it is assumed to be shared by others in society, such as journalists, teachers and others who distribute and popularize specialized knowledge". As a consequence, although some concepts were specialized concepts in origin, they have come to be part of a layperson's general knowledge in a process known as banalization. This is particularly true of some terms we are dealing with in this study, such as ´molestation` or ´abduction`, since the constant reference to organized crime activities in the news have popularized types of crime which were originally used by experts only.

The 10 sister concepts (using the COREL notation above) come to populate the Core Ontology as terminal concepts subordinated to the hyperordinate +PERSONAL_CRIME_00. They are considered terminal concepts because they constitute the final link in the conceptual hierarchy of the ontology and, as such, they lack definitional potential to take part in the MP of other concepts (Jiménez-Briones & Luzondo-Oyón 2011:16). Following the notation system, these concepts are preceded by the symbol [$]. In FunGramKB Suite, these concepts are alphabetically listed since the conceptual organization of the field is based on deep semantics and "IS A" relations (Periñán-Pascual & Arcas-Túnez 2007). In other words, the mechanisms of inference and inheritance that characterize the micro-conceptual knowledge spreading process (also called Microknowing) in FungramKB (Periñán-Pascual & Arcas- Túnez 2005) connect concepts and predications among the MPs of concepts. Furthermore, the concepts representing types of crime which were rejected for their inclusion in the specialized satellite ontology, on the grounds of their being part of general knowledge, can be included in the Core Ontology as terminal concepts eventually subordinated to +CRIME_00 through the umbrella concept +PERSONAL_CRIME_00. Hence, work in the specialized domain has not only been productive for the *Globalcrimeterm* itself, but it has likewise enriched the Core Ontology by providing more terminal concepts to populate it. To understand the whole process, the role of deep semantics in FunGramKb has to be clarified at this point. In surface semantics, legal ontology engineers, for example, have been producing taxonomies and have established connections among units

(or concepts) basing their assumptions on expert extra-linguistic information, for example, legal theories or deontic logic, but the reasoning capacity has been generally limited to very specific tasks. However, the way the concepts relate to each other in this proposal, as stated above, is based on deep semantics, which combines an extensive commonsense knowledge base (FunGramKB) and a reasoning engine. Consequently, the Ontology of FunGramKB (and the other two modules: Cognicon and Onomasticon) can work with two reasoning processes: MicroKnowing (Microconceptual Knowledge Spreading) and MacroKnowing (Macroconceptual-Knowing Spreading). [2] Microknowing is performed by two types of reasoning mechanisms: inheritance and inference. Inheritance, for instance, strictly involves the transfer of one or several predications from a superordinate concept to a subordinate one in the ontology. On the other hand, inference is based on the structures shared between predications linked to conceptual units which do not take part in the same subsumption relation within the ontology. The application of these two mechanisms on the MPs allowsFunGramKB to minimize redundancy and maximize the informative capacity of the knowledge base.

In the following lines, the concepts under scrutiny can be eventually included in the Core Ontology after their previous extraction from the *Globalcrimeterm* Corpus and their subsequent conceptual modelling. In alphabetical order:

1. Terminal Concept:
$ABDUCTION_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
A person (usually a child or woman) is taken away illegally, usually by force or fraud.
Meaning Postulate:
+(e1: +BE_00 (x1: +ABDUCTION_00)Theme (x2: +PERSONAL_CRIME_00)Referent)
*(e2: +TAKE_01 (x3)Agent (x4: +CHILD_00 ^ +WOMAN_00)Theme (x5)Location (x6)Origin (x7)Goal (f1: +VIOLENCE_00 | +LIE_01)Manner (f1: x1)Scene)
Lexical units: Abduction, kidnapping [English]; secuestro, abducción, rapto [Spanish]

2. Terminal Concept:
$CHILD_ABUSE_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
Child abuse is a crime. An adult harms a child in an emotional, physical and sexual way. The result that child abuse brings with it is often death.
Meaning Postulate:

[2] For an account of these two reasoning processes, see Periñán-Pascual&Arcas-Túnez (2005, 2007).

+(e1: +BE_00 (x1: +CHILD_ABUSE_00)Theme (x2: +PERSONAL_CRIME_00)Referent) (f1: +(e2: +HARM_00 (x3: +ADULT_00)Theme (x4: +CHILD_00)Referent) +EMOTION_00^ +BODY_00 ^ +SEX_00)Manner (f2: +(e3: +USE_00 (x3)Theme (x4)Referent)Purpose) (f3: +(e4: +BE_00 (x1)Theme (x6: prob+DEATH_00)Referent)Result)
Lexical units: Child abuse [English]; maltrato infantil [Spanish]

3. Terminal Concept:
$CHILD_PORNOGRAPHY_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
Sexually explicit images or the like involving children in the company of adults.
Meaning Postulate:
+(e1: +BE_00 (x1: $CHILD_PORNOGRAPHY_00)Theme (x2: +PERSONAL_CRIME_00)Referent) +(e2: +SHOW_00 (x1: s +IMAGE_00) Theme (x3: +SEX_00)Referent (f1: s +CHILD_00 & +ADULT_00)Company)
Lexical units:
Child pornography [English]; pornografía infantil [Spanish]

4. Terminal Concept:
$CHILD_PROSTITUTION_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
The act of obtaining, procuring or offering the services of a child or inducing a child to perform sexual acts for any form of compensation or reward.
Meaning Postulate:
+(e1: +BE_00 (x1: $CHILD_PROSTITUTION_00)Theme (x2: +PERSONAL_CRIME_00)Referent) +(e2: +TRANSFER_00 (x3: +CUSTOMER_00)Agent (x4: +MONEY_00)Theme (x5)Origin (x6: +CHILD_00)Goal (f1: +SEX_00)Purpose)
Lexical units: Child prostitution[English]; prostitución infantil [Spanish]

5. Terminal Concept:
$CHILD_TRAFFICKING_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
A criminal trafficks children. A criminal recruits, transports, transfer, hides, and/or receives a child. The criminal wants to exploit the child.
Meaning Postulate:
+(e1: +BE_00 (x1: +CHILD_TRAFFICKING_00)Theme (x2: +PERSONAL_CRIME_00)Referent) +(e2: +TAKE_00| +TRANSPORT_00| +TRANSFER_00| +HIDE_00 |^ +RECEIVE_00 (x3: +CRIMINAL_00)Agent (x4: s+CHILD_00)Theme) (x5)Goal)

(f1: +(e3: +USE_00 (x3)Theme (x4)Referent)Purpose)
Lexical units: Child trafficking [English]; tráfico de niños [Spanish]

6. Terminal Concept:
$COERCE_D_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
A person commits criminal coercion when he restricts a person's freedom of action and threatens a person.
Meaning Postulate:
+(e1: +BE_00 (x1:$COERCE_D_00)Theme (x2: +PERSONAL_CRIME_00)Referent)
+(e2: $COERCE_00 (x3)Theme (x4)Referent (x5)Goal (f1: x1)Scene)
(…)
Terminal concept:
$COERCE_00
Thematic frame:
(x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00)Goal
Meaning Postulate:
+(e1: +THREATEN_00 (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00)Goal (f1: +VIOLENT_00)Manner (f2: (e2: +DO_00 (x3)Theme (x4)Referent))Purpose)
Lexical units: Coercion [English]; coerción, coacción [Spanish]

7. Terminal Concept:
$FORCED_LABOR_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
Involuntary work exacted from a person under threat of penalty.
Meaning Postulate:
+(e1: +BE_00 (x1: +FORCED_LABOUR_00)Theme (x2: +PERSONAL_CRIME_00)Referent)
+(e2: +BE_00 (x1)Theme (x3: +COMMAND_01)Referent)
*(e3: +THREATEN_00 (x4: +HUMAN_00)Theme (x5: +PUNISHMENT_00)Referent (x6: +HUMAN_00)Goal (f1: x1)Scene)
Lexical units: Forced labour, bonded labour, compulsory labour [English]; trabajos forzados, trabajos forzosos [Spanish]

8. Terminal Concept:
$MOLEST_D_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:

The crime of the trade in humans, for the purpose of exploitation. The trade is generally achieved by means of violence or deceit.
Meaning Postulate:
(e1: +BE_00 (x1: $MOLEST_D_00)Theme (x2: +PERSONAL_CRIME_00)Referent)
+(e2: $MOLEST_00 (x1)Theme (x2)Referent)
(…)
Terminal concept:
$MOLEST_00
Thematic frame:
(x1: +HUMAN_00)Theme (x2: +HUMAN_00)Referent
Meaning Postulate:
+(e1: +ATTACK_00 (x1: +ADULT_00)Theme (x2: +CHILD_00 ^ +ADULT_00)Referent (f1: +SEX_00)Manner)
Lexical units: Molestation, abuse [English]; abuso sexual, acoso sexual [Spanish]

9. Terminal Concept:
$ORGAN_TRAFFICKING_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
The illegal organ trade occurs when organs are removed from the body of one person for the purpose of commercial transactions.
Meaning Postulate:
+(e1: +BE_00 (x1: $ORGAN_TRAFFICKING_00)Theme (x2: +PERSONAL_CRIME_00)Referent)
+(e2: +TAKE_00 (x3: +CRIMINAL_00)Theme (x4: +ORGAN_00)Referent
(f1: VIOLENT_00)Manner
(f2:(e3: +TRANSFER_00 (x3)Agent (x5: s +INTERNAL_ORGAN_00)Theme (x6)Origin (x7)Goal))Purpose
(f3:(e4: +SELL_00 (x3)Agent (x5)Theme (x8)Origin (x9)Goal))Purpose)
Lexical units: Organ trafficking, organ trade [English]; tráfico de órganos [Spanish]

10. Terminal Concept:
$SLAVERY_00
Conceptual path:
#ENTITY>#PHYSICAL>#PROCCESS>+OCCURRENCE_00>+CRIME_00>+PERSONAL_CRIME_00
Definition:
It is the system by which people are owned by other people as slaves. It is now illegal.
Meaning Postulate:
+(e1: +BE_00 (x1: $SLAVERY_00)Theme (x2: +PERSONAL_CRIME_00)Referent)
+(e2: +HAVE_00 (x3: +HUMAN_00)Theme (x4: +HUMAN_00)Referent (f1: +SLAVE_00)Role)
Lexical units: Slavery, chattel slavery[English]; esclavitud [Spanish]

The ten sister concepts defined above share a number of common features in the way their meaning is represented using COREL as interface language. They are all terminal concepts with a similar conceptual path and have an MP which includes a minimum of two predications (e_1, e_2, e_3 …) and at least one satellite (f_1, f_2, f_3 …). They are all subordinated to the umbrella concept +PERSONAL_CRIME_00. Concepts such as +HUMAN_00 or +CHILD_00 within the MPs are recipients (or victims) of the criminal activity depending on the thematic role. The active role in the performance of the crime also includes concepts such as +HUMAN_00, +CRIMINAL_00 or +ADULT_00 in the MP. Except for $ORGAN_TRAFFICKING_00 and $FORCED_LABOR_00, all the other terminal concepts may involve crimes directly or indirectly linked to some sort of sexual abuse or exploitation, but a common factor to all of them is the psychological or physical violence often represented in basic concepts included in the MPs such as +VIOLENT_00, +VIOLENCE_00, +SLAVE_00, +ATTACK_00, +THREATEN_00, +PUNISHMENT_00, +DEATH_00 or +HARM_00. The semantic and conceptual information inserted into the MPs undoubtedly shows the weakness and vulnerability of children when they are exposed to organized crime.

5. Conclusions

The initial list of potentially relevant terms provided by the FunGramKB Term Extractor Tool has been organized taking the *Globalcrimeterm* corpus as an empirical source: ten winning concepts and their corresponding lexical units have been identified and defined in connection with the area of organized crime, but taking children as a target. The hierarchy of the winning concepts (superordinates, subordinates, etc.) has been set up as well as their meaning postulates. It is true that in previous research (Alameda-Hernández & Felices-Lago 2016), it was claimed that "crime" is a concept that is part of both general and specialized knowledge and, in consequence, it had to be included both in the core and the satellite ontology. However, the hierarchisation phase has demonstrated how the apparently specialised crimes related to children in the role of victims replicate the same classification tenets and share the same upper conceptual level as other basic and terminal concepts of the Core Ontology. In fact, these supposedly "specific" entities (nouns) are eventually included in the Core Ontology and not in the domain-specific ontology, as previously calculated. At a more specific level, the concept +CRIME_00 has been used as a canonical instantiation of conceptual modelling and a similar process has been followed to represent the meaning of the unmbrella concept +PERSONAL_CRIME_00 and the other 10 subordinate terminal concepts collected in section 4 above. Among the reasons that could explain this unexpected result, it is worth noting the evidence provided by lexicographical sources, which show how the semantic content of the units linked to the selected concepts is not only known by legal practitioners but also shared by the average speaker of the language. Another crucial factor for the development of this research has been the possibility to use FunGramKB, which was designed to cover many of the most noticeable problems currently faced by NLP and practitioners in the area of artificial intelligence. In this respect, this contribution could do its bit in the implementation and development of FunGramKB Core Ontology and a better understanding of the high degree of complexity to establish clear-cut limits between general and specialised information in

disciplines such as criminal law or other branches of law, which are considered as social sciences.

Acknowledgements

This article is based on research carried out within the framework of the Project FFI2014-53788-C3-1-P, which is funded by the Spanish Ministry of Economy and Competitiveness, and entitled: *Development of a virtual laboratory for natural language processing from a functional paradigm.* It is also based on the previous Project code no. FFI2010-15983: *Development of a subontology in a multilingual context (English, Spanish and Italian): Using FunGramKB in the field of international cooperation in criminal matters: Terrorism and organised crime*".

References

Alameda Hernández, Ángela and Ángel Felices Lago. 2016. "The integration of the concept +CRIME_00 in FunGramKB and the conceptualization or hierarchization problems involved". In *Understanding Meaning and Knowledge Representation: From Theoretical and Cognitive Linguistics to Natural Language Processing*, edited by Carlos Periñán-Pascual and Eva Mestre-Mestre, 319-340, Newcastle: Cambridge Scholars.

Black's Law Dictionary. 2009. Saint Paul: Thomson Reuters.

Bowker, Lynne and Jennifer Pearson. 2002. *Working with Specialized Language. A Practical Guide to Using Corpora*. London and New York: Routledge. https://doi.org/10.4324/9780203469255

Breuker, Joost, André Valente and Radboud Winkels. 2005. "Use and reuse of legal ontologies in knowledge engineering and information management." In *Law and the semantic web*, edited by Richard Benjamins, Pompeu Casonovas, Joost Breuker and Aldo Gangemi, vol. 3369, 36-64. Berlin: Springer. https://doi.org/10.1007/978-3-540-32253-5_4

Breuker, Joost, Pompeu Casanovas, Michel C. A. Klein and Enrico Francesconi, eds. 2008. *Law, ontologies and the semantic web. Channelling the legal information flood, Frontiers in Artificial Intelligence and Applications*, vol. 188, Amsterdam: IOS Press.

Carrión-Delgado, Gracia. 2012. "Extracción y análisis de unidades léxico-conceptuales del dominio jurídico: un acercamiento metodológico desde Fungramkb." *RAEL: revista electrónica de lingüística aplicada* 11: 25-39.

Faber, Pamela, Ricardo Mairal-Usón, and Pedro Magaña-Redondo. 2011. "Linking a domain-specific ontology to a general ontology." *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*, edited by R. Charles Murray and Philip M. McCarthy, 564-569. Menlo Park, California: The AAAI Press.

Felices-Lago, Ángel. 2015. "Foundational considerations for the development of the *Globalcrimeterm* subontology: A research project based on FunGramKB". *Onomázein* 31: 127-144. https://doi.org/10.7764/onomazein.31.9

Felices-Lago, Ángel. 2016. "The Process of Constructing Ontological Meaning Based on Criminal Law Verbs". *Círculo de Lingüística Aplicada a la Comunicación* 65: 109-148. https://doi.org/10.5209/rev_CLAC.2016.v65.51983

Felices-Lago, Ángel, Pedro Ureña Gómez-Moreno and Ángela Alameda Hernández. 2011. "FunGramKB y la adquisición terminológica." *Anglogermánica Online* 8: 66-86.

Felices-Lago, Ángel and Pedro Ureña Gómez-Moreno. 2012. "Fundamentos metodológicos de la creación subontológica en FunGramKB", *Onomázein* 26: 49-67.

Felices-Lago, Ángel and Pedro Ureña Gómez-Moreno. 2014. "FunGramKB Term Extractor: a key instrument for building a satellite ontology based on a specialized corpus". In *Language processing and grammars: The role of functionally oriented computational models* (Studies in Language Series), edited by Brian Nolan and Carlos Periñán, 251-269, Amsterdam: John Benjamins.

Felices-Lago, Ángel and Ángela Alameda-Hernández. Forthcoming, 2017. "The process of building the upper-level hierarchy for the aircraft structure ontology to be integrated in FunGramKB." *Revista de lenguas para fines específicos* 22/2.

Gangemi, Aldo, Maria-Teresa Sagri, and Daniela Tiscornia. 2005. "A Constructive Framework for Legal Ontologies". In *Law and the Semantic Web*, edited by V.Richard Benjamins et al., 97-124. Berlin: Springer. https://doi.org/10.1007/978-3-540-32253-5_7

Jiménez-Briones, Rocío and Alba Luzondo-Oyón. 2011. "Building ontological meaning in a léxico-conceptual knowledge base". *Onomázein* 23: 11-40.

Koester, Almut. 2010. "Building small specialised corpora". In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O' Keeffe and Michael McCarthy, 66-79, London: Routledge. https://doi.org/10.4324/9780203856949.ch6

Lenci, Alessandro. 2008. "From context to meaning: distributional models of the lexicon in linguistics and cognitive science. A foreword". *Italian Journal of Linguistics* 20: 1-31.

Mairal-Usón, Ricardo and Carlos Periñán-Pascual. 2009. "The anatomy of the lexicon component within the framework of a conceptual knowledge base." *Revista Española de Lingüística Aplicada* 22: 217-244.

Mairal-Usón, Ricardo and Carlos Periñán-Pascual. 2010. "Role and Reference Grammar and Ontological Engineering." In *Los caminos de la lengua: Estudios en homenaje a Enrique Alcaraz Varó*, edited by José Luis Cifuentes et al., 649-65, Alicante: Universidad de Alicante.

Mairal-Usón, Ricardo and Francisco J. Ruiz-de-Mendoza. 2008. "New challenges for lexical representation within the Lexical-Constructional Model." *Revista Canaria de Estudios Ingleses* 57: 137-158.

Mairal-Usón, Ricardo and Francisco J. Ruiz-de-Mendoza. 2009. "Levels of description and explanation in meaning construction". In *Deconstructing Constructions*, edited by Chris Butler and Javier Martín Arista, 153–198. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.107.08lev

Niles, Ian and Adam Pease. 2001. "Towards a Standard Upper Ontology". In *Formal Ontology in Information Systems. Collected Papers from the 2nd International Conference (FOIS-01),* edited by Christopher Welty and Barry Smith, 2-9. Ogunquit, Maine. https://doi.org/10.1145/505168.505170

Oxford Dictionary of Law. 2003. Oxford: Oxford University Press.

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2004. "Meaning postulates in a lexico-conceptual knowledge base". In *15th International Workshop on Databases and Expert Systems Applications*, 38-42. Los Alamitos (California): IEEE.

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2005. "Microconceptual-Knowledge Spreading in FunGramKB". In *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*, 239-244. Anaheim-Calgary-Zurich: ACTA Press.

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2007. "Cognitive modules of an NLP knowledge base for language understanding", *Procesamiento del Lenguaje Natural* 39: 197-204.

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2010a. "Ontological commitments in FunGramKB", *Procesamiento del Lenguaje Natural* 44: 27-34.

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2010b. "The architecture of FunGramKB." In *Proceedings of the 7th International Conference on Language Resources and Evaluation, European Language Resources Association*, 2667-2674, EuropeanLanguageResourcesAssociation (ELRA).

Periñán-Pascual, Carlos and Francisco Arcas-Túnez. 2014. "La ingeniería del conocimiento en el dominio legal: La construcción de una Ontología Satélite en FunGramKB", *Revista Signos. Estudios de Lingüística* 47 (84): 113-139.

Periñán-Pascual, Carlos and Ricardo Mairal-Usón. 2009. "Bringing Role and Reference Grammar to natural language understanding", *Procesamiento del Lenguaje Natural* 43: 265-273.

Periñán-Pascual, Carlos and Ricardo Mairal-Usón. 2010. "La gramática de COREL. Un lenguaje de representación conceptual". *Onomázein* 21: 11-45.

Periñán-Pascual, Carlos and Ricardo Mairal-Usón. 2011. "The COHERENT methodology in FunGramKB". *Onomázein* 24: 13-33.

Procter, Paul, ed. 1978. *Longman Dictionary of Contemporary English.* Harlow, Essex: Longman.

Ruiz-de-Mendoza, Francisco J. and Ricardo Mairal-Usón. 2008. "Levels of description and constraining factors in meaning construction. An introduction to the Lexical Constructional Model", *Folia Linguistica* 42: 355-400. https://doi.org/10.1515/flin.2008.355

Ureña Gómez-Moreno, Pedro, Ángela Alameda-Hernández and Ángel Felices-Lago. 2011. "Towards a specialized corpus of organized crime and terrorism." In *La investigación y la enseñanza aplicadas a las lenguas de especialidad y a la tecnología*, edited by Mª Luisa Carrió et al., 301-306. Valencia: Universitat Politècnica de València.

Valente, André. 2005. "Types and roles of legal ontologies". In *Law and the semantic web*, edited by Richard Benjamins, Pompeu Casonovas, Joost Breuker and Aldo Gangemi, vol. 3369, 65-76. Berlin: Springer. https://doi.org/10.1007/978-3-540-32253-5_5

Van Dijk, Teun A. 2003. "Specialized discourse and knowledge. A case study of the discourse of modern genetics." *Cadernos de Estudos Lingüísticos* 44: 21-55.