Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 212 (2015) 166 – 173

Procedia
Social and Behavioral Sciences

MULTIMODAL COMMUNICATION IN THE 21ST CENTURY: PROFESSIONAL AND ACADEMIC CHALLENGES. 33rd Conference of the Spanish Association of Applied Linguistics (AESLA), XXXIII AESLA CONFERENCE, 16-18 April 2015, Madrid, Spain

Conceptual representation of expert knowledge in FunGramKB: the derivation process in several typical crimes of criminal law

María de Gracia Carrión Delgado*

Universidad Nacional de Educación a Distancia (UNED), Madrid 28040, Spain

Abstract

FunGramKB is a multilingual and multipurpose lexico-conceptual knowledge-base designed for its use in various tasks in Natural Language Processing (NLP), like information extraction and retrieval, machine translation or artificial reasoning (Periñán and Arcas, 2004; Mairal and Periñán, 2009; 2010). Its modular structure reflects three levels of knowledge—lexical, grammatical and ontological— which, though independent, are interrelated through the conceptual module and divided into three others: the Ontology, the Cognicon and the Onomasticon. Moreover, the Ontology represents a hierarchical catalogue of concepts that describe semantic knowledge organized in three subontologies, whose metaconcepts correspond to #ENTITY, #EVENT and #QUALITY, which permit the internal organization of nouns, verbs and adjectives respectively. Each of these subontologies is divided as well into three groups: metaconcepts (they represent cognitive dimensions), basic concepts (which comprise common sense knowledge) and terminal concepts (that provide expert knowledge). In this contribution we analyze some criminal offences prototypical from the legal domain, such as "corruption", "extortion" and "forgery", derived from their corresponding verbs. We have chosen these as examples of terminal concepts that attest that it is possible to integrate expert knowledge in FunGramKB, thanks to the conceptual representation language COREL, common to the three main modules of the conceptual level. The detailed analysis of the above mentioned entities will ascertain not only that it is possible to reuse the information from the Meaning Postulates (MPs) of the events from which they derive, but also that the information already included in the knowledge base is maximized.

Peer-review under responsibility of the Scientific Committee of the XXXIII AESLA CONFERENCE

* Tel.: +34-618824677
E-mail address: mdgcarrion@madrid.uned.es

Peer-review under responsibility of the Scientific Committee of the XXXIII AESLA CONFERENCE
doi:10.1016/j.sbspro.2015.11.315

Keywords: basic concept, COREL, FunGramKB, Meaning Postulate (MP), metaconcept, Natural Language Processing (NLP), Ontology, terminal concept.

1. Introduction

FunGramKB is a multilingual and multipurpose lexico-conceptual knowledge base designed for performing tasks of natural language processing in different languages (Bulgarian, Catalan, Spanish, French, English and Italian). It can be reused for information extraction and retrieval, machine translation and dialogue-based systems.

Furthermore, in FunGramKB the three main levels of knowledge are represented (lexical, grammatical and conceptual) whose modules are independent but interrelated (Periñán & Arcas, 2011: 2-3). Thus, lexical and grammatical modules depend on each language, while the conceptual module is shared by all the languages integrated in the knowledge base (Fig.1).

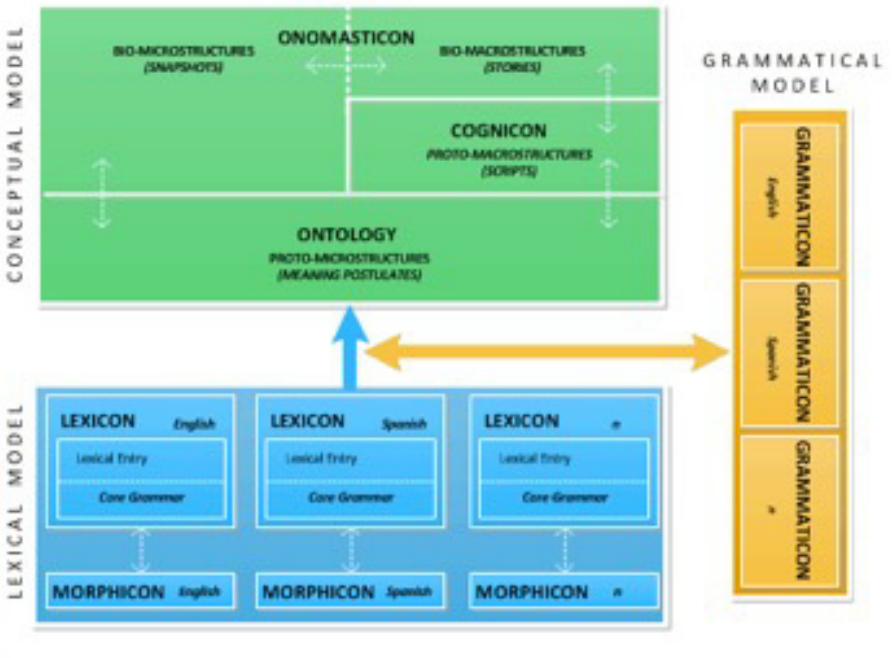


Fig. 1. FunGramKB Suite

Therefore, compared to other lexical bases where meaning is expressed through superficial relationships between lexical units (FrameNet or MultiWordnet), we can say that FunGramKB has a conceptual basis, since the Core Ontology supports all the structure of the knowledge base. Furthermore, the multilevel structure of the Core Ontology, divided into metaconcepts, basic concepts and terminal concepts allows the conceptual representation of the human cognitive structure.

Consequently, in this study we focus on the #ENTITY subontology, particularly in the definition of several prototypical crimes of the criminal law domain, which are derived from their corresponding verbs, as examples of terminal concepts integrated into FunGramKB Core Ontology.

Accordingly, the content of this paper is structured as follows: First, we present the structure of FunGramKB (Periñán & Arcas, 2010). Then, we apply the methodology COHERENT to the definition of terms; and, finally, we provide the detailed analysis of the terminal concepts $CORRUPT_D_00, $EXTORT_D_00 and $FORGE_D_00.

2. FunGramKB Conceptual Module

2.1. The Core Ontology

In FunGramKB concepts fall into three levels in the conceptual hierarchy. The top level is made up of 42 metaconcepts capitalized and preceded by the "#" sign, representing cognitive dimensions on which the concepts are organized. The Core Ontology contains three subontologies whose metaconcepts are #ENTITY, #QUALITY and #EVENT, which allow the internal organization of nouns, verbs and adjectives respectively.

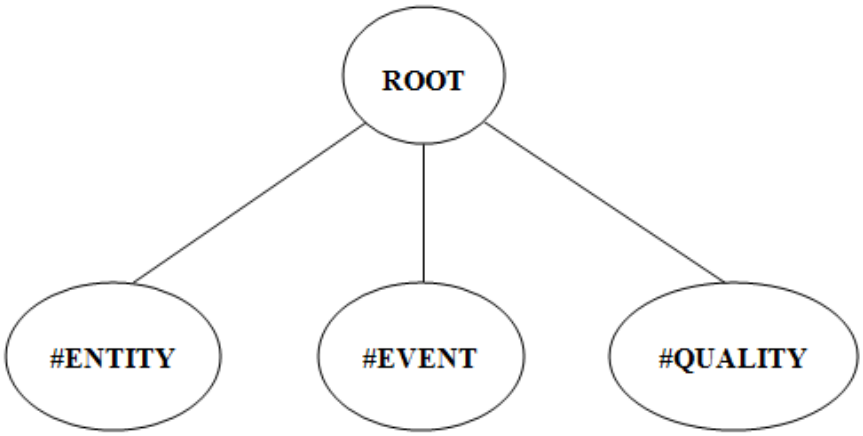
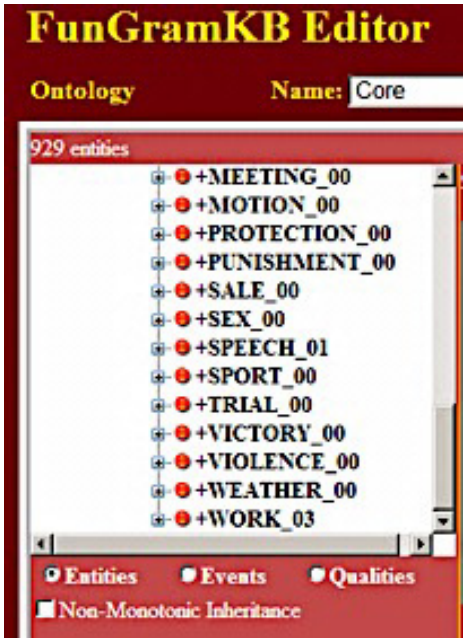


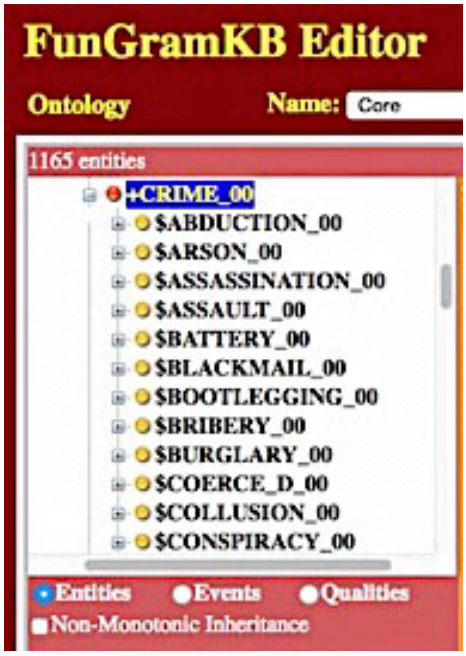
Fig. 2. Types of concepts.

Then, at an intermediate level, the basic concepts are represented with the "+" sign and followed by an underscore and a numeric index (eg. + PUNISHMENT_00, + VIOLENCE_00, etc), as we can see in Fig. 3. These are used as defining units that allow the construction of Meaning Postulates (MPs) of the basic concepts and terminals, and they are selection preferences in the Thematic Frames (ThFrs) of qualities and events as well.



Fig, 3. Partial representation of entities in FunGramKB.

Finally, in the end node of the hierarchy, terminal concepts are preceded by the "$" sign and followed by an underscore and a numeric index (eg. $ASSAULT_00, $BURGLARY_00, etc, in Fig. 4). This is where our study will focus, specifically on the analysis of the terms corruption, extortion and forgery, typical crimes of the domain of criminal law.



Fig, 4. Partial representation of +CRIME_00 in FunGramKB.

3. The COHERENT Methodology

The 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 serves as a methodological basis for the development of satellite ontologies linked to it (Carrion, 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 shown in Fig. 5.

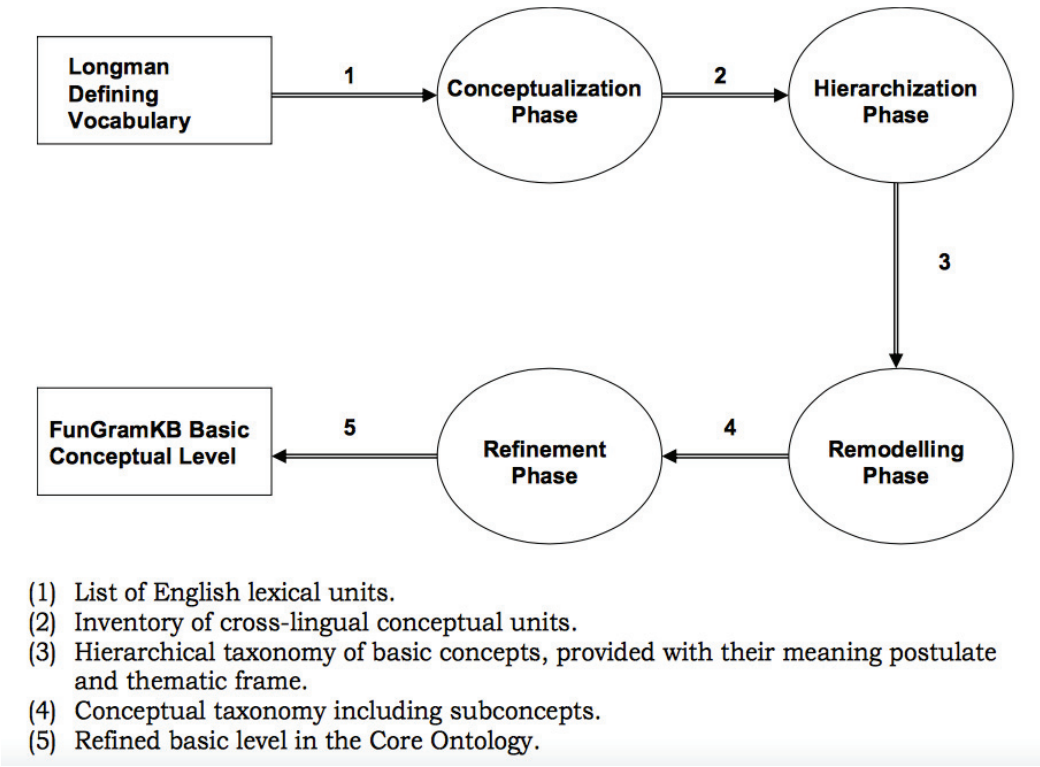


Fig. 5. The COHERENT methodology (Periñán & Mairal, 2011: 20).

As a result of these four phases, we obtained a catalogue of approximately 1,300 basic concepts that have been the basis to populate the Core Ontology with terminal concepts, 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.

3.1. Terminological conceptualization

Using a corpus of texts, *FunGramKB Term Extractor* can automatically obtain a list of candidate terms representative of a specific domain, that serve the terminologist to manually develop the filtering of terms and the definition of concepts.

As a result, the terminologist performs tasks of hierarchization and conceptualization from a selection of the winner terms obtained from the extractor. Then, the first step is to identify the basic concepts, that is, the defining words of the subject domain, as a basis for defining terminal concepts, i.e., the more specific ones.

At a later stage, the meaning of the concepts of the Core Ontology, which belong to the three levels of the conceptual hierarchy of FunGramKB mentioned above, is modeled by using COREL (COnceptual REpresentation Language, Periñán & Mairal, 2010).

In the case of derived entities, the MPs are constructed from the events, which, in turn, are also terminal concepts. Through the subsumption System (IS-A) of non-monotonic multiple inheritance, the meaning of the event is inferred in the MP of the derived entity. As the event is also a terminal concept itself, a conflict arises for the reasoner of the knowledge base, because there is more than one terminal concept in the same MP. This produces an error that prevents recording the information of the concept in the knowledge base. In the following section we analyse several typical crimes of the criminal law domain to show how we have solved this problem with the reasoner.

3.2. The derivation process in several typical crimes of criminal law

The derivation process is a common phenomenon in the terminology of criminal law, where a significant number of concepts are derived from their corresponding verb. In this study we present and analyze the derivation process in three prototypical terms of criminal law (corruption, extortion and forgery). As a matter of fact, we start from the analysis of the event for further study of the MP of the derived entity. The natural language definitions presented here are the result of the synthesis of the various lexicographical sources examined to develop the definition in COREL.

(1) *Corrupt*: To encourage someone to start behaving in an immoral or dishonest way.

+CORRUPT_00:
ThFr: (x1)Theme (x2)Referent

MP: +((e1: +DO_00 (x1)Theme (x3: (e2: ing +DO_00 (x4: +HUMAN_00)Theme (x2)Referent))Referent)(e3: +BE_01 (x2)Theme (x5: +DISHONEST_00)Attribute))

Lexical units (English, Spanish and Italian): *corrupt, corromper, corrompere.*

In the hierarchization process we begin at the primary metaconcepts to then reach the most immediate hipernym basic concept. Thus, the proposed hierarchy of this term is described as follows:

#EVENT>#MATERIAL>+DO_00>**$CORRUPT_00**

We have already mentioned that the basic concepts are used to build the MP. But how is the conflict of having two terminal concepts in the same MP solved? In the following examples we show that this problem is tackled by using the index **_D_00**. In the following lines we offer a detailed analysis of the terminal concepts $CORRUPT_D_00, $EXTORT_00, $EXTORT_D_00, $FORGE_00 and $FORGE_D_00, where we provide their definition in natural language first, next their formalization in COREL and, finally, their proposal of hierarchization.

(2) *Corruption:* Dishonest, illegal, or immoral behaviour, especially from someone with power.

$CORRUPT_D_00:

MP: +(e1: +BE_00 (x1: $CORRUPT_D_00)Theme (x2: +CRIME_00)Referent)
+(e2: **$CORRUPT_00** (x3)Theme (x4)Referent)

Lexical units (English, Spanish and Italian): *corruption, corrupción, corruzione.*

The hierarchization proposal of this term is as follows:

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE_00>+CRIME_00>**$CORRUPTION_D_00**

Then, we do the same analysis with the terms extort - extortion, forge - forgery.

(3) *Extort:* To illegally force someone to give you something especially money, by threatening them.

$EXTORT_00:

ThFr: (x1)Theme (x2)Referent

MP: +(e1: +THREATEN_00 (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00)Goal (f1: (e2: +GIVE_00 (x3)Agent (x4: +MONEY_00)Theme (x3)Origin (x1)Goal))Purpose)

Lexical units (English, Spanish and Italian): *extort, extorsionar, estorcere.*

And its hierarchization proposal:

#EVENT>#COMMUNICATION>+SAY_00>+THREATEN_00>**$EXTORT_00**

(4) *Extortion:* A public officer uses his position to take money or any other benefit that is not due to him.

+EXTORT_D_00:

MP: +(e1: +BE_00 (x1: $EXTORT_D_00)Theme (x2: +CRIME_00)Referent)
+(e2: **$EXTORT_00** (x3)Theme (x4)Referent)

Lexical units (English, Spanish and Italian): *extortion, extorsión, estorsione.*

The hierarchization proposal of this term is as follows:

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE_00>+CRIME_00>**$EXTORT_D_00**

(5) *Forge:* To illegally copy something, especially something printed or written, to make people think that it is real.

+FORGE_00:

ThFr: (x1)Theme (x2)Referent

MP: +(e1: +CREATE_00 (x1: +CRIMINAL_00)Theme (x2: +DOCUMENT_00)Referent (f1)Instrument (f2:(e2: n +BE_01 (x2)Theme (x3: +TRUE_00)Attribute))Result)
+(e3: +CHANGE_00 (x1)Theme (x2)Referent (f3: (e4: +BECOME_00 (x2)Theme (x4: +REAL_00)Attribute))Result)

Lexical units (English, Spanish and Italian): *forge, falsificar, contraffare, falsificare.*

The hierarchization proposal that we offer for this term is the following:

#EVENT>#MATERIAL>+DO_00>+CREATE_00>**$FORGE_00**

(6) *Forgery:* A document, painting, or piece of paper money that has been copied illegally. The crime of copying official documents, money, etc.

+FORGE_D_00:

MP: +(e1: +BE_00 (x1: $FORGE_D_00)Theme (x2: +CRIME_00)Referent)
*(e2: **$FORGE_00** (x3)Theme (x4)Referent)

Lexical units (English, Spanish and Italian): *forgery, falsificación, contraffazione.*

The hierarchization we propose for this term is as follows:

#ENTITY>#PHYSICAL>#PROCESS>+OCCURRENCE_00>+CRIME_00>$**FORGE_D_00**

4. Conclusions

The examples of terminal concepts presented here have served to demonstrate that FunGramKB Core Ontology may reflect derivation cases without altering the structure of the knowledge base. That is, the Ontology accepts derivation without causing functional problems to the reasoner, while the redundancy of information is minimized and the informativeness is maximized in the semantic knowledge repository of FunGramKB (ibid. Periñán and Arcas, 2004). For this purpose, we have used COREL, the metalanguage of the conceptual module of the knowledge base, which allows the cognitive modeling of specialized terms as terminal concepts of the Core Ontology.

Acknowledgements

This paper is part of the research project entitled "Development of a terminological ontology in a multilingual context (Spanish, English and Italian) from FunGramKB knowledge base in the field of international cooperation in criminal matters: terrorism and organized crime", funded by the Ministry of Economy and Competitiveness. Code: FI2010-15983.

References

Carrión, M. (2012). "Extracción y análisis de unidades léxico-conceptuales del dominio jurídico: un acercamiento metodológico desde FunGramKB", in *Revista Electrónica de Lingüística Aplicada*.

Felices, A., Ureña, P. & Alameda, A. (2011). FunGramKB y la adquisición terminológica. *Anglogermánica Online* 2011: 66-86.

Jiménez, R., Luzondo, A. & Pérez, M. (2011). FunGramKB y la organización ontológica. *Anglogermánica Online*: 16-36.

Mairal, R. & Periñán, C. (2009). The anatomy of the lexicon component within the framework of a conceptual knowledge base. *Revista Española de Lingüística Aplicada* 22: 217-244.

Periñán, C. & Arcas, F. (2004). Meaning postulates in a lexico-conceptual knowledge base, 15th International Workshop on Databases and Expert Systems Applications, IEEE, Los Alamitos (California): 38-42.

Periñán, C. & Arcas, F. (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.

Periñán, C. & Arcas, F. (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, C. & Mairal. R. (2010). La gramática de COREL: un lenguaje de representación conceptual. *Onomázein* 21: 11-45.

Periñán, C. & Mairal. R. (2011). The COHERENT methodology in FunGramKB. *Onomázein* 24: 13-33.

Procter, P. (1978). *Longman dictionary of contemporary English.* Harlow [England]: Longman.

Ureña, P., Alameda, A. & Felices, A. (2011). Towards a specialised corpus of organized crime and terrorism. María Luisa Carrió y otros (Eds.), *La investigación y la enseñanza aplicadas a las lenguas de especialidad y a la tecnología*. Universitat Politècnica de Valencia, Valencia, págs. 301-306.