Basics of Ontology modeling in FunGramKB. The case of *burn**

Andreea Rosca
University of La Rioja
deia_nira7@yahoo.com

Abstract

The aim of the present study is two-fold. Firstly, we discuss the advantages of a conceptual approach to meaning representation within the framework of a multipurpose Natural Language Processing (NLP) system known as FunGramKB (Periñán and Arcas, 2004, 2005, 2006; Periñán and Mairal, 2009ab, 2010, to name a few). FunGramKB solves some of the problems encountered in relational databases in that it provides morphosyntactic and pragmatic information about lexical units, it avoids language dependency by working with concepts and not words, and it minimizes redundancy by cognitive clustering. Secondly, we offer an outline of the ontological modeling of concepts related to the change-of-state verb *burn*. FunGramKB is an invaluable knowledge base that can be later used for the development of numerous NLP applications, such as intelligent question-answer systems or cross-linguistic information retrieval applications.

Keywords: FunGramKB, Natural Language Processing, relational databases, lexico-conceptual knowledge base, the Ontology.

Resumen

El objetivo de este trabajo de investigación es doble. En primer lugar, se analizan las ventajas de un enfoque conceptualista para la representación de significado dentro del marco del sistema de procesamiento de lenguaje natural (PLN) denominado FunGramKB (Periñán y Arcas 2004, 2005, 2006; Periñán y Mairal, 2009ab, 2010, entre muchos otros). FunGramKB soluciona muchos de los problemas encontrados en las bases de datos relacionales en el sentido de que proporciona información morfosintáctica y pragmática sobre las unidades léxicas, evita la dependencia del lenguaje mediante el uso de conceptos y minimiza la redundancia mediante el agrupamiento cognitivo. En segundo lugar, se ofrece una descripción de la modelación ontológica de conceptos relacionados con el verbo de cambio de estado 'quemar'. FunGramKB es una base de conocimiento inestimable que podría servir para el desarrollo de numerosas aplicaciones de procesamiento de lenguaje natural, a saber sistemas inteligentes de pregunta-respuesta o recuperación de información entre lenguas.

Palabras clave: FunGramKB, Procesamiento de Lenguaje Natural, bases de datos relacionales, base de conocimiento léxico-conceptual, Ontología.

1. Introduction

The main goal of this study is to describe how linguistic information related to the change-of-state verb *burn* can be implemented in a Natural Language Processing (henceforth NLP) system known as FunGramKB. FunGramKB or Functional-Grammar Knowledge Base is, in the words of Periñán and Arcas (2005: 239), "a user-friendly online environment for the semiautomatic construction of a multipurpose lexico-conceptual knowledge base for NLP systems". This knowledge engineering project is multipurpose since it is both multifunctional and multilingual. On the one hand, FunGramKB can be employed in several natural language processing (NLP) tasks, such as information retrieval and extraction, machine translation, dialogue-based systems, text categorization, data mining, etc. On the other hand, it provides information from many natural languages. English and Spanish are fully supported in the current version of this knowledge base whereas information from German, French, Italian, Bulgarian and Catalan is in the process of being included. Periñán and Arcas (2007) argue that although FunGramKB follows Dik's Functional Grammar (1997), it differs from this model of semantic representation in two important aspects: (i) Dik's model was devised for a single NLP task, i.e. machine translation whereas FunGramKB can be used in various NLP tasks, including machine translation; (ii) Contrary to Dik (1997), who proposes the use of words for the formal description of meaning postulates, FunGramKB describes words using universal concepts, thus, avoiding the problem of language dependency and lexical ambiguity caused by the polysemic nature of lexical units. The construction of such a project demands the collaboration of both linguists and knowledge engineers.

The selection of this particular knowledge base is motivated by several factors. First, the formalism of FunGramKB relies on solid linguistic models, such as Dik's Functional Grammar (1997) or Van Valin and La Polla's (1997) Role and Reference Grammar (RRG henceforth). It should be stated that the influence of Dik's Functional Grammar (1997) initially lies just in the knowledge representation language (or COREL), although it has been progressively enhanced in order to meet the requirements of natural language understanding. However, FunGramKB surpasses these models in that (i) it proposes a conceptual orientation by dealing with universal concepts and not language-dependent words; (ii) it replaces RRG logical structures with Conceptual Logical Structures or CLSs which do not express redundant information and incorporate not only the syntactically relevant aspects; (iii) CLSs are enriched with cultural and encyclopedic knowledge via inference mechanisms. Second, the description of meaning in FunGramKB goes beyond the relational approach adopted by lexicographical databases such as SIMPLE or EuroWordNet since it embraces a conceptual perspective which is more parsimonious (e.g. the minimization of redundancy through the agglutination of various lexical units to a single concept, the clustering of words encoding the same cognitive scenario, etc.) and it allows for a greater expressive power (e.g. it codes quantification, temporality, modality; it is non-monotonic, thus, permitting the withdrawal of predications).

This article is structured as follows. Section 2 introduces the reader to the basic theoretical tenets and architecture of FunGramKB. In section 3 we provide a bird's-eye view of the organization of the Ontology in this knowledge base. In section 4, we will show how knowledge engineers create terminal concepts in FunGramKB and make decisions about the number of concepts to be built, the lexical units associated to these concepts, etc. The final section summarizes all the findings of the present study.

2. The architecture of FunGramKB

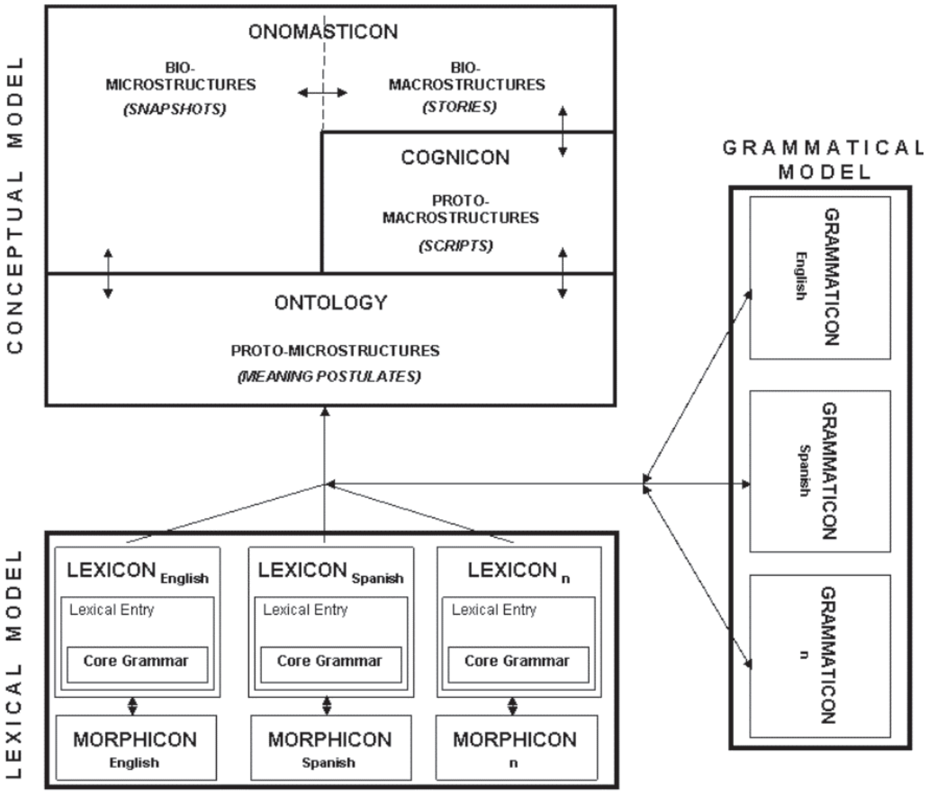
FunGramKB makes a neat distinction between the linguistic and the conceptual level:

(i) The linguistic level comprises a lexical and a grammatical module. The lexical realm captures specific properties of the languages of the world. This component can be further divided into: (a) a Lexicon, and (b) a Morphicon. The first contains morphosyntactic, pragmatic and collocational information about a given lexical unit. Mairal and Periñán (2009b: 220) claim that the lexical component is not a literal implementation of the lexical information in RRG. Despite maintaining the fundamental assumptions postulated by RRG, i.e. logical structures, macroroles, and the linking algorithm, FunGramKB is a more robust knowledge base. The Morphicon deals with cases of inflectional morphology. The grammatical level also known as the Grammaticon has four Constructicon modules: (a) L1-Constructicon or the argument structure layer; (b) L2-Constructicon or the implicational layer; (c) L3-Constructicon or the illocutionary level; and (d) L4-Constructicon or the discourse-structure level.

(ii) The conceptual level is composed of three language-independent knowledge schemata. The Cognicon stores procedural knowledge (e.g. how to fry an egg or how to buy a product), the Onomasticon deals with episodic knowledge (i.e. instances of entities and events like the Beatles, Taj Mahal, or 9/11), whereas the Ontology is organized as a hierarchical catalogue of universal concepts.

Figure 1 below offers a panoramic view of the architecture of this knowledge base and the way the six modules interact:

Figure 1. The FunGramKB architecture (extracted from Periñán and Mairal, 2012: 335)



As pictured in this figure, there is a clear-cut division between the linguistic level (the Lexicon, the Morphicon, and the Grammaticon) and the conceptual level (the Ontology, the Cognicon, and the Onomasticon). This figure also reflects the typology of conceptual schemata according to the parameters of prototypicality and temporality. The conceptual representations that store prototypical knowledge are called proto-structures while those that describe instances of entities or events are labeled bio-structures. Thus, if we want to describe the meaning of the word *building* we have to construct a proto-structure. By contrast, if we want to depict a particular building, e.g. the Guggenheim Museum, then we have to use bio-structures. Also, knowledge can be presented atemporally (i.e. microstructures, e.g. the description of the profession of an architect) or within a temporal framework (i.e. macrostructures, e.g. the biography of Frank Gehry). It should be stated that knowledge schemata stored in FunGramKB have already been described in full detail in other papers, such as Periñán (2013, on semantic knowledge), Periñán (2012, on procedural knowledge), and Periñán and Carrión (2011, on encyclopedic knowledge).

Our main focus of attention will be on the Ontology. Before examining this module in detail several observations should be made. First, the lexical and the grammatical modules contain language-specific information whilst the conceptual module is shared by all the languages included in FunGramKB. There are two main consequences deriving from this. On the one hand, computational lexicographers will create one Lexicon, one Morphicon, and one Grammaticon for English, one Lexicon, one Morphicon, and one Grammaticon for Spanish, and so on. On the other hand, knowledge engineers will develop one Ontology, one Cognicon and one Onomasticon that are sufficient to process any linguistic input conceptually. Second, this distinction between the linguistic and the non-linguistic levels leads to the use of two different metalanguages, i.e. the *conceptual logical structures* (CLS henceforth), and the *Conceptual Representation Language* (hereafter, COREL).

Finally, the Ontology is the pivot around which the different lexica revolve since this knowledge base is conceptually-driven. The Lexicon is populated in a top-down fashion, i.e. the description of a lexical entry must be preceded by the creation of a corresponding concept in the Ontology. For example, a computational lexicographer can fill in the morphosyntactic information related to lexical units such as *transfer* (Eng.) and *transferir* (Sp.) only if a knowledge engineer has previously created in the Ontology the concept +TRANSFER_00 together with its thematic frame and meaning postulate.

3. The FunGramKB Ontology

The Ontology of this knowledge engineering project differentiates between three conceptual levels with different degrees of genericity/specificity: metaconcepts, basic concepts, and terminal concepts. Metaconcepts represent the upper level and the most abstract conceptual level. They are preceded by the symbol # (e.g. #ABSTRACT, #MOTION, #POSSESSION, etc.). FunGramKB has forty-two metaconcepts which are divided into three subontologies: #ENTITY for nouns, #EVENT for verbs, and #QUALITY for adjectives. In their turn, basic concepts, which are headed by the symbol + (e.g. +MOVE_00, +HUMAN_00, +DIRTY_00, etc.), are defining units which allow knowledge engineers to create meaning postulates for both basic and terminal concepts. The latter are preceded by the symbol $ and they lack defining potential so they cannot participate in meaning postulates (e.g. $EXCHANGE_00, $CONGRATULATE_00, $HUM_00, etc.).

Basic and terminal concepts are not atomic symbols but they are characterized by semantic properties, i.e. the thematic frame (TF) and the meaning postulate (MP), which serve as building blocks for the formal description of meaning. It is also worth pointing out that both conceptual schemata are language-independent semantic knowledge representations. In the Ontology events (i.e. verbs) are provided with one thematic frame, which is conceived as "a conceptual construct which states the number and types of participants involved in the prototypical cognitive situation portrayed by the concept" (cf. Periñán and Mairal, 2009: 267). For the sake of illustration, consider the thematic frame of the basic concept +PAY_00, to which lexical units like *pay* [Eng], *pagar* [Spa], or *payer* [Fre] are connected:

(1) (x1: +HUMAN_00)Agent (x2: +MONEY_00)Theme (x3)Origin (x4: +HUMAN_00)Goal

Therefore, this basic concept takes four participant roles: (i) an Agent who makes another entity move (x1); (ii) a Theme that changes its place or position (x2); (iii) an Origin, which is the location from which the Theme moves (x3), and (iv) a Goal, which is the location to which the Theme moves (x4). The participant roles can be further specified through the addition of selectional preferences which can exert some predictive power on the thematic roles, e.g. +HUMAN_00, +MONEY_00. These selectional preferences show that the Agent and the Goal are necessarily human beings whereas the Theme (i.e. what is transferred) is always money.

Finally, we should mention the case of subconcepts which are codified by a preceding minus symbol and in capital letters. They come into existence when the conceptual narrowing or specification takes place exclusively in one or all of the participants of the TF of a basic or terminal concept, without varying the MPs. For example, the subconcept –PREEN is a conceptual specification of the basic concept +CLEAN_01, i.e. the Theme role is always a bird and the Referent is the feathers.

4. Ontology modeling

According to Jiménez-Briones and Luzondo (2011), the meticulous process of creation of new terminal concepts can be summarized in three main steps:

(i) Knowledge engineers must consult several lexicographical dictionaries prior to the introduction of any new predication in the form of meaning postulates (MPs). To preserve the universal status of the Ontology, knowledge engineers must use both English and Spanish dictionaries. When necessary, dictionary definitions will be complemented with our common sense, which may not be mirrored in the lexicographical entries.

(ii) New terminal concepts or subconcepts must be inserted in the Ontology only when these concepts are characterized by well-marked differentiae, which separate them neatly from their immediate superordinate concept. Since the Ontology is language independent, no lexical gaps between different languages should prevent the creation of terminal concepts. Also, as noted by Mairal and Periñán (2009a: 222-223), a new concept must be created whenever we encounter at least one lexical item whose meaning does not match any of the MPs stored in the knowledge base "provided that the values of the ontological properties of that concept are shared by all lexical units which are linked to it".

(iii) Finally, the meaning of the new terminal concept will be coded into the Ontology using the COREL notation. For the insertion of MPs and the potential selectional preferences in the thematic frames (TFs), knowledge engineers can choose from a limited set of concepts (e.g. 1,300 basic concepts) which can sometimes lead to coarse-grained implementations.

In order to illustrate how terminal concepts are created in FunGramKB, we will depart from the basic concept +BURN_00, whose conceptual route is: #EVENT >> #MATERIAL >> #TRANSFORMATION >> +CHANGE_00 >> +DAMAGE_00 >> +BURN_00.[1] As can be observed, the most immediate superordinate of +BURN_00 is the basic concept +DAMAGE_00, which is assigned the TF and MP illustrated in (1):

(1) TF: (x1)Theme (x2: +CORPUSCULAR_00)Referent

MP: +(e1: +CHANGE_00 (x1)Theme (x2)Referent (f1: (e2: +BECOME_00 (x2)Theme (x3: +UGLY_00)Attribute))Result)

These TF and MP provide information about the number and type of participants involved in the prototypical cognitive situation of damaging something. Thus, an unspecified Theme (x1) changes a three dimensional countable entity (x2; Referent) and, as a result, the Referent becomes ugly (f1). The 'x' and 'f' slots in the MPs differ in that the former indicate thematic roles of different cognitive dimensions whereas the latter are used to express relevant *differentiae* between basic or terminal concepts. The concept +DAMAGE_00 is lexicalized in four languages, namely English, Spanish, Italian, and French (e.g. *damage, harm, dañar, estropear, danneggiare, ledere, rovinare, sciupare, abîmer,* and *endommager*).

Coming back to +BURN_00, we notice that the selectional preferences that appear in its TF differ from the ones in +DAMAGE_00, as can be observed in (2):

(2) TF: (x1: +HUMAN_00)Theme (x2)Referent

Thus, the first participant role (x1) is delimited by the selectional preference +HUMAN_00, which tells us that the entity that performs the action can only be a human being. As for the Referent (x2), this refers to any entity that can be set on fire. The structure of the basic concept +BURN_00 complies with the similarity principle according to which all subordinate concepts must share the MP of their superordinate concept (see Periñán and Arcas 2007). In our case, the inheritance relationship is marked by the presence of the superordinate +DAMAGE_00 in the first predication of the subordinate concept +BURN_00. The MP of +BURN_00 is mapped into the COREL representation in (3), whose natural language equivalent is reproduced in (4):

(3) +(e1: +DAMAGE_00 (x1)Theme (x2)Referent (f1: +FUEL_00)Instrument).

(4) A person (x1) damages an unspecified entity (x2) using fuel as in instrument (satellite f1).

The MP of +BURN_00 also obeys the specificity principle (Periñán and Arcas 2007) which stipulates that the MP of a subordinate concept must comprise a distinctive feature (or *differentia*) not present in the MP of its superordinate concept. Hence, the satellite f1 (Instrument) is what separates the subordinate concept +BURN_00 from its superordinate concept +DAMAGE_00 (see Appendix 1.2 for the semantic interpretation of satellites). Also, the differentia in the MP of +BURN_00 has an exclusive value within the metaconcept established by the superordinate concept +DAMAGE_00, i.e. #TRANSFORMATION (cf. the opposition principle). Periñán and Arcas (2007) claim that MPs of ontological concepts in FunGramKB observe the opposition principle which stipulates that the *differentiae* in the MPs of sibling concepts must be incompatible with one another.

4.1. From dictionary definitions to COREL formalizations

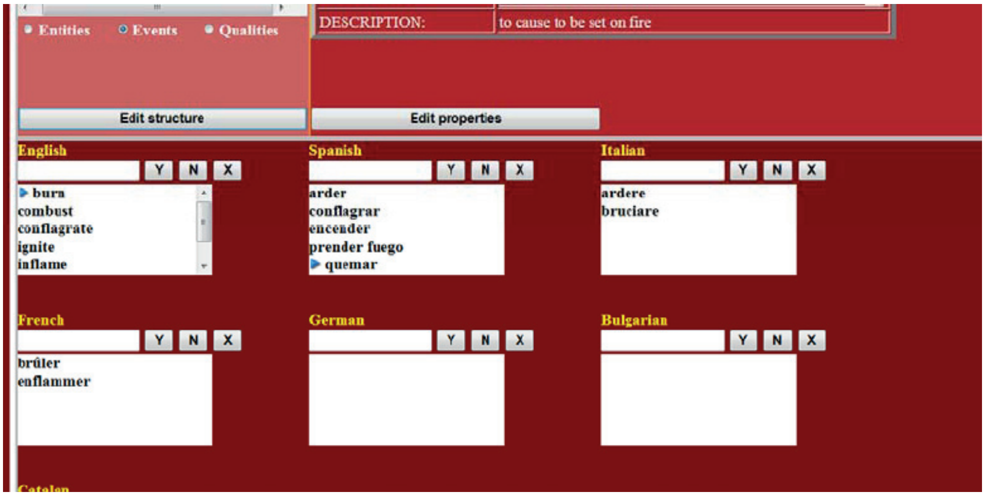
The first stage in the process of creation of new terminals consists of gathering all the possible synonyms for the basic concept +BURN_00, in English as well as in Spanish. Among the most commonly used dictionaries, we can mention on the one hand, *Longman Dictionary of Contemporary English, English Collins Dictionary and Thesaurus, Cambridge Advanced Learner's Dictionary, Merriam-Webster Dictionary, OneLook Dictionary* for English and on the other hand, DRAE: *Diccionario de la Lengua Española* (Real Academia), *CLAVE*, or *Diccionario de Sinónimos y Antónimos* (Espasa Calpe) for Spanish. After consulting these dictionaries, we collected the following English and Spanish synonyms:

English: *cauterize, carbonize, char, combust, conflagrate, cremate, ignite, incinerate, inflame, kindle, light, scorch, singe, torch.*

Spanish: *arder* ('burn'), *abrasar* ('sear'), *cauterizar* ('cauterize'), *carbonizar* ('carbonize'), *chamuscar* ('scorch'), *conflagrar* ('conflagrate'), *encender* ('light'), *incinerar* ('incinerate'), *inflamar* ('inflame'), *prender* ('light').

We have discarded the words *scorch*, *parch*, and *agostar* ('parch') because the Theme is non-human. Also, these words have already been agglutinated as lexical units under the terminal concept $WITHER_00, since they are more related to drying something than to burning it. Lastly, we have noticed that the verbs *roast, toast, sear* and their Spanish counterparts *achicharrar* ('sear'), *asar* ('roast'), *tostar* ('toast'), *abrasar* ('sear') have been agglutinated under the basic concept +COOK_00. Nevertheless, since some of the meaning extensions of *achicharrar* ('burn') and *abrasar* ('burn') are related to the basic concept +BURN_00, these verbs can be linked as lexical units to this basic concept. After discarding all the verbs that do not share the same *genus* as their superordinate basic concept, we started looking up the definitions of the verbs listed above. A closer inspection of the definitions of the verbs *combust, conflagrate, ignite, inflame, kindle, light, arder* ('burn'), *conflagrar* ('conflagrate'), *encender* ('light'), and *prender* ('light') reveals that they do not add any new features to the basic concept +BURN_00. These verbs will be simply connected as lexical units to the basic concept +BURN_00. Figure 2 below shows how the FunGramKB Ontology establishes the link between words "which are language-dependent" and the concept +BURN_00, which is language independent:

Figure 2. Lexical units linked to the concept +BURN_00



Other words that are associated with the basic concept +BURN_00 are the Italian verbs *ardere* ('burn'), *bruciare* ('burn'), and the French verbs *brûler* ('burn'), and *enflammer* ('inflame'). It should also be mentioned that the differences in syntactic patterns between verbs are not handled in the Ontology, but in the Lexicon where lexicographers specify syntactic information about a given lexical unit.

The rest of the verbs do exhibit features that differentiate them from their *genus*, i.e. +BURN_00. Consider the verbs *char, scorch, singe,* and *chamuscar* ('scorch') which share similar meanings. Regarding the verb *scorch,* the meaning listed here is different from the one which refers to the process of drying undergone by plants under the influence of strong heat or wind. These verbs add new information related to the specificity of the burning entity (i.e. a surface), the manner in which the event occurs (i.e. slightly) and the outcome of the event (i.e. the surface becomes black). Since these features cannot be overlooked, we must create a new terminal concept that will depend hierarchically on its most immediate superordinate concept, viz. +BURN_00. We will first label the new terminal concept $SINGE_00 and we will then continue to store the information concerning the type of participants involved in the burning event. The specificity of the burning entity will be reflected in the selectional preferences of the TF:

(5) TF: (x1: +HUMAN_00)Theme (x2: +SURFACE_00)Referent

Therefore, the entity that burns is human whereas the entity being burnt is a surface of another entity. Once the construction of the TF is accomplished, we proceed to create the MP whose structure can be divided into two main parts: (i) the first predication (e1), which is identical to the meaning expressed by the superordinate concept +BURN_00, and (ii) the distinct features coded in the form of satellites (f1 and f2):

(6) +(e1: +BURN_00 (x1)Theme (x2)Referent (f1: +LITTLE_00)Manner) (f2: (e2: +BECOME_00 (x2)Theme (x3: +BLACK_00)Attribute))Result)

The COREL representation can be translated into natural language in the following way: a Theme (x1) burns a Referent (x2) slightly (f1) and as a result its surface (x2) acquires a black color (f2). At this point it is important to mention that the basic concept +BECOME_00 is used to indicate a change in one of the properties of an entity, such as form, shape, or color. By contrast, the basic concept +BE_01 is employed to express a conspicuous change in an entity. Furthermore, COREL differentiates between +BE_00, which means 'to belong to a class' or 'to have identity with', +BE_01, which highlights inalienable properties, and +BE_02, which is used for locations.

At the final stage we move on to link to this new terminal concept those lexical units codifying the same or related lexical meanings. Thus, the verbs *char, scorch, singe,* and *chamuscar* ('scorch') will be associated to $SINGE_00. It should be borne in mind that the granularity of MPs is not as detailed as that in human-oriented lexicographical definitions since we have yet to see a machine or NLP system capable of rendering the complexity of human reasoning. Also, as pointed by Jiménez and Luzondo (2011: 28), the aim of the FunGramKB Ontology is "to capture the knowledge of a cultivated average speaker who in most cases will be unable to distinguish more than probably three different senses of a word".

4.2. The case of $CAUTERIZE_00

Another verb which deserves further consideration is *cauterize*, whose meaning is illustrated below:

(Collins Cobuild): If a doctor cauterizes a wound, he or she burns it with heat or with a chemical in order to close it up and prevent it from becoming infected.

(Cambridge): To burn an injury to stop bleeding and prevent infection.

(Macmillan Dictionary): to close a cut by using a hot instrument in order to prevent infection or to stop blood from flowing out.

As can be remarked, the semantic make-up of this verb is too fine-grained to be encoded by its superordinate +BURN_00. It is thus necessary to create a new terminal concept which will be placed immediately under its parent concept. The specificity of the entity that is being burnt will be codified in the selectional preferences of the TF of this new terminal concept, which will be labeled $CAUTERIZE_00:

(7) (x1: +HUMAN_00)Theme
(x2: +WOUND_00)Referent

Therefore, $CAUTERIZE_00 indicates that a human being burns a wound. There are two distinguishing properties of the verb *cauterize* that will have to be coded in the form of satellites: (i) the instrument used to perform the action is either heat or a chemical, and (ii) the purpose of the burning action is to cure the injury. With this in mind, the knowledge engineer sets out the task of editing the MP of the new terminal concept, which would look like this:

(8) +(e1: +BURN_00 (x1)Theme (x2)Referent (f1: +HEAT_00 ^ +CHEMICAL_00)Instrument) (f2: (e2: +CURE_00 (x1)Theme (x2)Referent) Purpose)

('A human being (x1) burns an entity (x2) using heat or a chemical as instruments (f1) in order (f2) to cure that entity (x2)')

The first predication of $CAUTERIZE_00 (i.e. e1: +BURN_00 (x1)Theme (x2)Referent) is inherited from its superordinate basic concept. Also, we can notice that satellites can be immediately followed by a basic concept (f1) or by another predication and its thematic roles (f2). These satellites add new information related to the cauterization process: the instruments used are heat or a chemical (f1) and the burning event has curative purposes (f2).

The next step consists of connecting semantically similar words to this terminal concept. There are only four lexical units that express the same meaning as this terminal, namely the verbs *cauterize* [Eng], *cauterizar* [Spa], *cautériser* [Fre], and *cauterizzare* [Ita].The same procedure will be followed for the introduction of three other terminal concepts related to the 'burning' scenario, viz. $INCINERATE_00, $CREMATE_00, and $TORCH_00:

(9) $INCINERATE_00: to burn something completely

TF: (x1: +HUMAN_00)Theme (x2)Referent

MP: +(e1: +BURN_00 (x1)Theme (x2)Referent (f1: (e2: +BECOME_00 (x2) Theme (x3: +ASH_00)Attribute))Result)
('A human being (x1) burns something (x2) and that entity (x2) turns to ashes (f1: Result)').

Agglutinated lexical units: *incinerate, carbonize, calcine, incinerar* ('incinerate'), *calcinar* ('calcine'), *carbonizar* ('carbonize'), *carbonizzare* ('carbonize'), *incenerire* ('incinerate'), *incinérer* ('incinerate'), *carboniser* ('carbonize'), *calciner* ('calcine').

(10) $CREMATE_00: to burn the body of a dead person

TF: (x1: +HUMAN_00)Theme (x2: +HUMAN_00)Referent

MP: +((e1: +BURN_00 (x1)Theme (x2)Referent (f1: (e2: n +BE_01 (x2) Theme (x3: +ALIVE_00)Attribute))Condition (f2: (e3: +BE_01 (x2) Theme (x4: +ASH_00)Attribute))Result))[8]

('A human being (x1) burns another human being (x2) with the condition (f1) that the Referent (x2) is not alive and as a result (f2) the Referent (x2) turns to ashes (x4)').

Agglutinated lexical units: *cremate, cremar* ('cremate'), *cremare* ('cremate'), *incinérer* ('cremate').

(11) $TORCH_00: to burn a building or other large thing, intentionally and usually illegally

TF: (x1: +HUMAN_00)Theme (x2)Referent

MP: +(e1: +BURN_00 (x1)Theme (x2)Referent (f1: $LEGAL_N_00)Manner)

(‘A human being (x1) burns something (x2) in an illegal manner (f1)’)

Agglutinated lexical units: *torch*, *incendiar* (‘torch’), *dare fuoco a* (‘set fire to something’), *mettre le feu à* (‘set fire to something’).

In the case of $CREMATE_00, the two brackets before the first predication (e1) mark a ‘conceptual binding’ phenomenon (cf. Periñán and Mairal, 2010), which establishes a direct correlation between the participants engaged in the two predications that are surrounded by the brackets. In our case, what is being burned (x2) is characterized by the attribute of being dead, i.e. not alive. Also, the *n* polarity operator, which is usually employed in negative statements, modifies here the basic concept +ALIVE_00, thus, indicating the condition of the Referent. It is also worth pointing out that although FunGramKB is an invaluable conceptual knowledge base it cannot equal the intricate reasoning carried out by the human brain.

5. Conclusions

This paper has highlighted the main differences between FunGramKB and other lexicographical databases (e.g. SIMPLE, EuroWordNet). We have also presented a brief overview of the main theoretical postulates and architecture of this knowledge base. The reader has been familiarized with the different concepts knowledge engineers must deal with, such as metaconcepts, basic concepts and terminal concepts. In the second part of this article we have detailed the laborious process carried out by knowledge engineers for the elaboration of terminal concepts. We have illustrated this protocol by focusing on the ontological domain +BURN_00. At this point new terminal concepts have been created, such as $SINGE_00, $CAUTERIZE_00, $INCINERATE_00, $CREMATE_00, and $TORCH_00. The construction of FunGramKB requires the joint collaboration of theoretical linguists with computational linguists.

6. References

Dik, S. C. 1997. *The Theory of Functional Grammar. Part 1: The Structure of the Clause.* 2nd ed. by K. Hengeveld. Berlin & New York: Mouton de Gruyer.

Jiménez, R. and A. Luzondo. 2011. "Building Ontological Meaning in a Lexico-Conceptual Knowledge Base". *Onomázein,* 23(1): 11-40.

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

____. 2009b. "Role and Reference Grammar and Ontological Engineering". *Volumen Homenaje a Enrique Alcaraz.* Universidad de Alicante.

Periñán-Pascual, C. 2012. "The situated common-sense knowledge in FunGramKB". *Review of Cognitive Linguistics* 10(1): 184-214.

Periñán-Pascual, C. 2013. "A knowledge-engineering approach to the cognitive categorization of lexical meaning". *VIAL: Vigo International Journal of Applied Linguistics* 10: 85-104.

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.

____. 2005. "Microconceptual-Knowledge Spreading in FunGramKB". *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing* (pp. 239-244). Anaheim-Calgary-Zürich: ACTA Press.

____.2006. "Reusing computer-oriented lexica as foreign-language electronic dictionaries". *Anglogermánica Online,* 4: 69-93.

____.2007. "Cognitive modules of an NLP knowledge base for language understanding". *Procesamiento del Lenguaje Natural,* 39: 197-204.

Periñán-Pascual, C. and M. Carrión Varela. 2011. "FunGramKB y el conocimiento cultural". *Anglogermánica Online* 8: 87-105.

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

____.2010. "La Gramática de COREL: un lenguaje de representación conceptual". *Onomazein,* 21(1): 11-45.

____.2012. "La dimensión computacional de la RRG: la estructura lógica conceptual y su aplicación en el procesamiento del lenguaje natural". In R. Mairal, L. Guerrero and C. González Vergara (eds.) *El Funcionalismo en la Teoría Lingüística: La*

Gramática del Papel y la Referencia. Madrid, Akal, 333-348.

Van Valin, R. and R. LaPolla. 1997. *Syntax, structure, meaning and function.* Cambridge: Cambridge University Press.

Appendices (adapted from Periñán and Mairal, 2010: 32-34)

1.1. Metaconcepts and the semantic interpretation of thematic roles

Metaconcept	Thematic roles	Definition
#TRANSFER	Agent	Entity that transfers another entity to a third entity
	Theme	Entity that is transferred
	Origin	Entity from which another entity is transferred
	Goal	Entity to which another entity is transferred
#TRANSFORMATION	Theme	Entity that transforms another entity
	Referent	Entity that is transformed by another entity

1.2. Semantic interpretation of satellites

Role	Definition
Attribute	Entity or quality that describes a feature of another entity
Condition	Predication that describes under which condition the event should occur
Instrument	Entity that is used to perform the event
Manner	Entity or quality that describes the way in which the event occurs
Purpose	Predication that describes the aim of the event
Result	Predication or entity that describes the consequence of the occurrence of the event