



## **Building ontological meaning in a lexico-conceptual knowledge base**

**Rocío Jiménez-Briones**

Universidad Autónoma de Madrid  
España

**Alba Luzondo Oyón**

Universidad de La Rioja  
España

### **Abstract**

Framed within the world of Artificial Intelligence, and more precisely within the project FunGramKB, i.e. a user-friendly environment for the semiautomatic construction of a multipurpose lexico-conceptual knowledge base for Natural Language Processing systems, the aim of this paper is two-fold. Firstly, we shall provide a necessarily non-exhaustive theoretical discussion of FunGramKB in which we will introduce the main elements that make up its Ontology (i.e. Thematic Frames, Meaning Postulates, different types of concepts, etc.). Secondly, we will describe the meticulous process carried out by knowledge engineers when populating this conceptually-driven Ontology. In doing so, we shall examine various examples belonging to the domain of 'change' or #TRANSFORMATION (in the COREL notation), in an attempt to show how conceptual knowledge can be modeled in for Artificial Intelligence purposes.

**Keywords:** FunGramKB; ontological meaning; conceptual modeling; meaning postulate; thematic frame; terminal concept.

---

Afiliaciones: Rocío Jiménez-Briones: Departamento de Filología Inglesa, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid. España. — Alba Luzondo Oyón: Departamento de Filologías Modernas, Facultad de Letras y de la Educación, Universidad de La Rioja. España.  
Correos electrónicos: rocio.jimenez@uam.es; alba.luzondo@gmail.com.  
Direcciones postales: Rocío Jiménez-Briones: C/ Virgen de Lluc 29, 3ºB. Madrid 28027. España. — Alba Luzondo Oyón: C/ Sagasta 25, 2º Dcha. 26001, Logroño (La Rioja). España.

Fecha de recepción: octubre de 2010

Fecha de aceptación: marzo de 2011

## 1. Introduction<sup>1</sup>

In the last few years the comprehensive theory of constructional meaning known as the Lexical Constructional Model (hereafter LCM; Mairal and Ruiz de Mendoza, 2008, 2009; Ruiz de Mendoza and Mairal, 2008, among others) has incorporated as part of its architecture the knowledge base FunGramKB (i.e. Functional Grammar Knowledge Base at [www.fungramkb.com](http://www.fungramkb.com))<sup>2</sup>, a user-friendly online lexico-conceptual knowledge base that integrates rich semantic and syntactic information for the creation of Natural Language Processing (henceforth NLP) applications (cf. Perrián and Arcas, 2004, 2005, 2007a, 2007b). Therefore, linguists trained within the LCM and other related theories are currently working in the world of Artificial Intelligence by populating FunGramKB. In fact, such a joint collaboration has seen the consolidation and spread of job profiles such as *ontology modelers*, *knowledge engineers*, *computational lexicographers*, and, in specialized domains of knowledge, *terminographers* and *terminologists*.

Our focus in this article is on the FunGramKB Ontology, which can be regarded as one of the principal components around which the whole knowledge base revolves. Along with the Cognicon and the Onomasticon, the Ontology, which stores semantic knowledge in the form of meaning postulates (hereafter MP) by presenting a hierarchical catalogue of all the concepts that a person has in mind, is one of the non-linguistic modules that make up FunGramKB (Perrián and Arcas, 2007a: 198). It is important to stress that one of the main features that sets FunGramKB apart is the fact that it follows a conceptual approach based on deep semantics, unlike other knowledge bases grounded on surface semantics (cf. Perrián and Arcas, 2004,

<sup>1</sup> Financial support for this research has been provided by the DGI, Spanish Ministry of Science and Innovation, grants no. HUM2007-65755, no. FFI2008-05035-C02-01 (co-financed through FEDER funds), and no. FFI2010-17610/FILO. We would like to thank Prof. Mairal Usón, Prof. Ruiz de Mendoza and Prof. Perrián for their insightful comments and suggestions on a previous version of this paper. All usual disclaimers apply.

<sup>2</sup> Abbreviations employed in this article: C 'Cambridge International Dictionary of English'; CC 'Collins Cobuild English Language Dictionary'; COREL 'Conceptual Representation Language'; FunGramKB 'Functional Grammar Knowledge Base'; L 'Longman Dictionary of Contemporary English'; LCM 'Lexical Constructional Model'; MP 'meaning postulate'; NLP 'natural language processing'; TF 'thematic frame.'

2005, 2007a, 2007b). For instance, the FunGramKB Ontology offers rich conceptual descriptions to which lexical units are then associated, that is, each lexical unit is provided with a real definition (in the form of basic and terminal concepts) formalized employing what has been termed Conceptual Representation Language or COREL (Periñán and Mairal, 2009, 2010). Therefore, lexical units are always linked to one or more concepts in the Ontology and, vice versa, the same concept is lexicalized by one or more words in the several FunGramKB lexica. Such a methodological move, as will be explained in these pages, is directly connected to the degree of granularity that our knowledge base should attain, which in turn places a heavy burden on the knowledge engineer's shoulders, since s/he is the one who has to decide key issues such as how many concepts should be built, their types (i.e. terminals or subconcepts), which lexical units should be associated to which concept(s), etc.

Within this context, the goal of this paper is to account for the methodology that will help knowledge engineers to work out such issues successfully and elaborate the different types of concepts that populate the FunGramKB Ontology. This is not a minor task, since the construction of the lexical entries in the different lexica that FunGramKB consists of first involves the previous ontological modeling of their corresponding concepts (Periñán and Mairal, 2009)<sup>3</sup>. In other words, computational lexicographers will not be able to type in the morphosyntactic, pragmatic and collocational information of verbs such as English "sport" or its Spanish translations *lucir* and *ostentar*, unless some knowledge engineer has first introduced in the Ontology the terminal concept \$SPORT\_00 together with its conceptual information, that is, its thematic frame (henceforth TF) and MP.

The organization of this paper is as follows. In section 2 we review the basic theoretical tenets of FunGramKB. In section 3 we detail the laborious process necessary to create terminal concepts in FunGramKB, employing for that the examples taken

---

<sup>3</sup> Although it is commonly agreed that there is not just one single methodology for ontology development (Noy and McGuinness, 2001 *apud* Periñán and Arcas, 2007a: 199), since it tends to be quite a creative process, whatever the design one decides for an ontology, it should be founded on a solid methodology so that all the knowledge engineers working on it can employ the same criteria to model concepts consistently. The interested reader is referred to Periñán and Arcas (2010b).

from the conceptual scenario +TRANSLATE\_00. Finally, some concluding remarks are offered in section 4.

## 2. FunGramKB: An overview

FunGramKB (Periñán and Arcas, 2004, 2005, 2007a, 2010a, 2010b; Mairal and Periñán 2009a, 2009b, 2010; Periñán and Mairal 2009) is a lexico-conceptual knowledge base that is both multifunctional and multilingual, in the sense that not only has it been designed to be reused in NLP applications - mainly those requiring language understanding like machine translation, information retrieval and extraction, etc. - but it also aims to cover various Western languages. FunGramKB consists of three information levels:

- a. Lexical level = linguistic knowledge
- b. Grammatical level = linguistic knowledge
- b. Conceptual level = non-linguistic knowledge

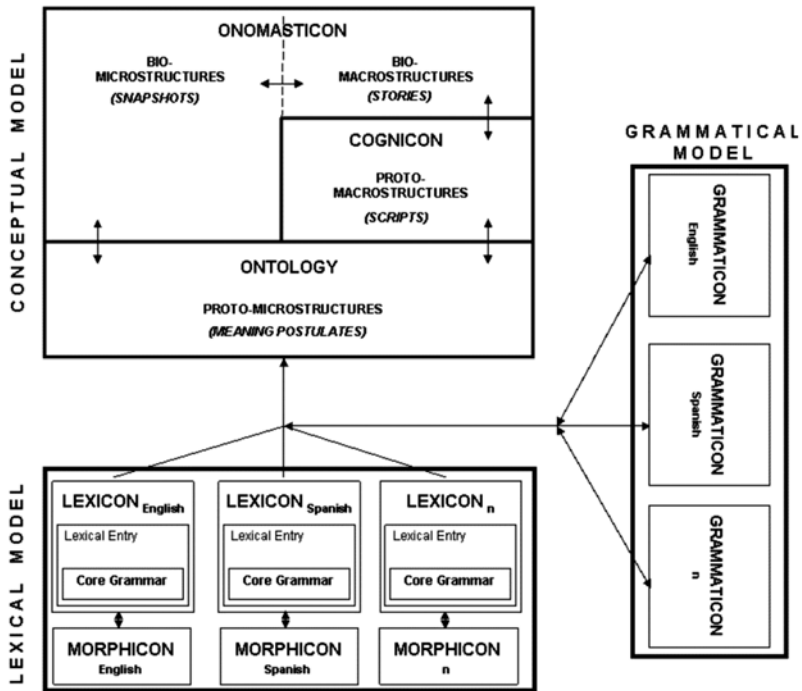
Each of these levels is in turn made up of several independent but interrelated modules, as shown in the diagram below:

The lexical level comprises: a) the various lexica (e.g. English, Spanish, Italian, German, etc., see footnote 9) which store morphosyntactic, pragmatic and collocational information about lexical units; and b) the Morphicon, which handles cases of inflectional morphology. The grammatical level, i.e. the Grammaticon, is currently being developed within the LCM<sup>4</sup>. It is formed by several Grammaticon modules (e.g. English, Spanish, Italian, German, etc.) and captures the properties that are specific to the most relevant constructional families in the languages under consideration on the basis of corpus data. Finally, the conceptual level consists of three modules: a) the Ontology or the hierarchical structure of concepts; b) the Cognicon, where procedural information is kept; and c) the Onomasticon, where information about instances of entities and events is stored.

---

<sup>4</sup> For further information on the LCM and the scholars currently working within this framework, we refer the reader to the Lexicom group research webpage: [www.lexicom.es](http://www.lexicom.es).

FIGURE 1  
**FunGramKB modules (at [www.fungramkb.com](http://www.fungramkb.com))**



As stated in Periñán and Arcas (2010a, 2010b), one must be fully aware that only the lexical and grammatical levels are language-dependent, while the modules that make up the conceptual level are all language-independent, that is to say, shared by all the languages in the knowledge base. This means that, when working within the FunGramKB context, linguists and computational lexicographers will build one Lexicon, one Morphicon and one Grammaticon for each of the languages implemented in the knowledge base (English, Spanish, Italian, etc.), whereas knowledge engineers will develop just one Ontology, one Cognicon and one Onomasticon to account for all those languages. Since the Ontology is the module where knowledge engineers will model conceptual meaning, let us now describe its main features.

As pictured in Figure 1, the FunGramKB Ontology is the key component around which the whole knowledge base pivots.

It distinguishes three different conceptual levels, each one with concepts of a different type (Periñán and Arcas, 2004)<sup>5</sup>:

- i. Metaconcepts (e.g. #ABSTRACT, #COLLECTION, #PSYCHOLOGICAL, #MOTION, #POSSESSION, etc), which 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.*, 2002; Masolo *et al.*, 2003), SIMPLE (Pedersen and Keson, 1999; Lenci *et al.*, 2000), and SUMO (Niles and Pease, 2001a, 2001b), to name just a few. The FunGramKB Ontology is actually split into three subontologies, since subsumption (IS-A) is the only taxonomic relation permitted, and therefore each subontology arranges lexical units of a different part of speech: i.e. #ENTITIES for nouns (e.g. +BIRD\_00, +SOUL\_00, +FREEDOM\_00,...), #EVENTS for verbs (+SAY\_00, +TRANSLATE\_00, etc.), and #QUALITIES for adjectives and some adverbs (i.e. +HAPPY\_00, +ALONE\_00...).
- ii. Basic concepts, preceded by symbol +, are used as defining units which enable the construction of MPs for basic concepts and terminals, as well as taking part as selection preferences in TFs: e.g. +HUMAN\_00, +ON\_00, +BE\_00, +MACHINE\_00, etc<sup>6</sup>.
- iii. Terminal concepts, the object of study of this work, 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, \$SPORT\_00, \$SUBTITLE\_00, \$BOW\_00, etc.

In FunGramKB, basic and terminal concepts are always stored with their ontological properties in the form of TFs and MPs. A TF is a conceptual construct which states the number and type of participants involved in the prototypical cognitive situation portrayed by concepts (Periñán and Arcas, 2007a: 267). Appendix 6.1 (adapted from Periñán and Mairal, 2010: 32-33) provides the semantic interpretation of the thematic roles in the metacognitive dimensions dealt with: #COGNITION,

<sup>5</sup> Periñán and Arcas (2007a, 2010b) provide a detailed explanation of the ontological commitments to which the FunGramKB Ontology is subject.

<sup>6</sup> The inventory of almost 1,300 basic concepts employed in FunGramKB stems from the defining vocabulary used in the *Longman Dictionary of Contemporary English* (Procter, 1978) and in the *Diccionario para la Enseñanza de la Lengua Española* (Alvar Ezquerro, 1995).

#POSSESSION, and #TRANSFORMATION. We should not forget that, unlike other ontologies, in FunGramKB every event and quality is assigned one TF. As way of exemplification, consider the TF of +WEAR\_00 below, to which lexical units like English *wear, have on, dress* or Spanish *llevar, llevar puesto, traer*, etc. are linked:

- (1) TF = (x1: +HUMAN\_00 ^ +PET\_00)Theme (x2: +CLOTHING\_00 ^ +ORNAMENT\_00)Referent

TFs also account for those selectional preferences typically involved in the cognitive situation under analysis (Jiménez and Pérez, 2010): +HUMAN\_00, +PET\_00, +CLOTHING\_00, and +ORNAMENT\_00 for the cognitive scenario of ‘wearing something’. Therefore, TF (1) describes this prototypical scenario as involving two participants: (i) entity x1 (Theme), being typically a human or a pet, which employing COREL is expressed with the basic concepts +HUMAN\_00 and +PET\_00 connected with the exclusion logical connector “^”; (ii) entity x2 (Referent), typically identified with clothes or some decoration, as the basic concepts +CLOTHING\_00 ^ +ORNAMENT\_00 codify. Appendix 6.2 (adapted from Perrián and Arcas, 2004: 41; Perrián and Mairal, 2010: 26-28) stores the logical connectors employed in COREL.

In turn, an MP comprises a group of one or more logically connected predications ( $e_1, e_2 \dots e_n$ ), which are conceptual constructs carrying the generic features of concepts (Perrián and Arcas, 2004: 39). It also incorporates the information stated in a TF by the co-indexation of the participants<sup>7</sup>. For example, the MP of the basic concept + WEAR\_00, which belongs to the metacognitive dimensions #POSSESSION > #RELATIONAL > #STATIVE > #EVENT, presents a semantic representation in which a human being or a pet (x1 = Theme) typically has clothes or ornaments (x2 = Referent) located on his/her body (Location), which employing COREL is formalized as follows:

- (2) MP = +(e1: +HAVE\_00 (x1)Theme (x2)Referent (f1: +BODY\_AREA\_00)Location (f2: +ON\_00)Position)

<sup>7</sup> Although the inclusion of TFs in MPs may seem redundant, it is highly necessary since it is through TFs that the mapping with the variables of the lexical templates (located in the lexical module) occurs. In other words, if TFs did not exist, the linkage between the Ontology and the different lexica would be inexistent (cf. Perrián and Mairal, 2009).

Following the Aristotelian tradition, the *genus* of the MP is the superordinate +HAVE\_00, which is modified by two *differentiae*, (f1) and (f2), which, as recorded in Appendix 6.3, express the location (+BODY\_AREA\_00) and the position of the clothes or ornaments (+ON\_00), respectively. In sum, both TFs and MPs are the notational means employed in FunGramKB to represent conceptual meaning, since they use concepts –and not words– as the building blocks for the formal description of meaning. Thus, TFs as well as MPs become language-independent semantic knowledge representations.

In relation to terminal concepts, as already pointed out, they cannot be employed as defining units in MPs and TFs. In fact, terminal concepts only arise when there is a conceptual narrowing on the MPs and TFs of a basic concept. Let us have a look at the representation of the terminal concept \$SPORT\_00, which is a further specification of the basic concept +WEAR\_00 studied in (1) and (2):

- (3) TF = (x1: +HUMAN\_00)Theme (x2: +CLOTHING\_00 ^ +ORNAMENT\_00)Referent  
 MP = +(e1: +WEAR\_00 (x1)Theme (x2)Referent (f1: (e2: +SHOW\_00 (x1)Theme (x2)Referent (f2: +PROUD\_00)Manner) Purpose)  
 ('someone (x1 = Theme) wears something (x2 = Referent) to show it proudly')

If compared with the TF and MP of +WEAR\_00, the terminal concept \$SPORT\_00 arises as a result of further specifying this basic concept, firstly by restricting the first participant to only human beings (x1: +HUMAN\_00) and, secondly, by including the parameter purpose (f1), which itself includes a manner parameter (f2) with the selectional preference +PROUD\_00. In other words, the terminal concept \$SPORT\_00 is brought about because people can prototypically wear things very proudly so that everybody can see them, which in English is lexicalized by the verb “sport” and in Spanish by *lucir* and *ostentar*.

Yet there are cases in which the conceptual specification takes place exclusively inside the TF of a basic or a terminal concept, without varying the MPs. When this occurs, ‘subconcepts’ come into existence in FunGramKB, being codified by a preceding minus symbol and in capital letters. For instance, within the cognitive situation ‘wearing something’, we have been able to distinguish the subconcept -TAKE\_SHOES:



## (4) -TAKE\_SHOES:

TF: (x1: +HUMAN\_00)Theme (x2: +SHOE\_00)Referent

MP: +(e1: +HAVE\_00 (x1)Theme (x2)Referent (f1: +BODY\_AREA\_00)Location (f2: +ON\_00)Position)

('someone has shoes on the body')

If compared with the TF of +WEAR\_00 in (1), since the MP must be exactly the same, -TAKE\_SHOES occurs when the selectional preferences of the Referent entity of the basic concept +WEAR\_00 are restricted to shoes, boots, etc., (captured by the basic concept +SHOE\_00) and those of the Theme entity to humans (i.e. +HUMAN\_00). This conceptual specification is lexicalized in Spanish in the verb *calzar* ('wear shoes or boots'), although English does not have a verb for it.

Since metaconcepts and basic concepts are already defined in FunGramKB, it is worth stressing the relevance of building terminal concepts and subconcepts appropriately for a fine-grained knowledge base which, as stated in section 1, is based on deep semantics, that is, a knowledge base where TFs and MPs offer rich conceptual descriptions to which lexical units are then associated. Therefore, all the detailed specifications done by knowledge engineers on the MPs and TFs that bring about terminal concepts and subconcepts can only but contribute to the fine-grained granularity of the FunGramKB Ontology. As a consequence, knowledge engineers have to cope with the crucial task of deciding how to model ontological meaning which, in practical terms, means deciding on, for example, whether or not to create the terminal concept \$SPORT\_00 and the subconcept -TAKE\_SHOES, how to formalize them in COREL notation, which lexical units should be linked to them, and so on and so forth. To help knowledge engineers out, the following section accounts for the protocol used to create such concepts thoroughly and methodically.

### 3. How to create terminal concepts and subconcepts: a step-by-step guide

In this section we present a detailed description of the laborious protocol carried out by knowledge engineers when creating terminal concepts within the FunGramKB Ontology.

It is essential to emphasize that we will only be dealing with #EVENTS, and thus both #QUALITIES and #ENTITIES will be disregarded.

The process of populating this Ontology can be broken down in the following necessary steps, which we will thoroughly elaborate along this section:

- a. As stated in Periñán and Arcas (2004, 2005), every predication posited is built on the basis of dictionary entries, which work as our guidance tools for the posterior definition of MPs. Likewise, dictionary definitions can be enriched with our commonsense knowledge, as this may not be reflected in the lexicographical statements of meaning. Since our Ontology is meant to be universal<sup>8</sup>, both English and Spanish dictionaries must be employed<sup>9</sup>. Hence, the underlying idea around which this project revolves is the design of an ontological module based on as many languages as possible, in which “the structuring of the ontology is guided by a process of negotiation” (Periñán and Arcas, 2010a: 2669).
- b. Thus, departing from these definitions (representing lexical meaning), the engineer is embarked on the process of deciding which terminal concepts and subconcepts (if any) may be created. It is worth emphasizing that terminal concepts should only be introduced when well-marked *differentiae*, which clearly set them apart from their immediate superordinate concepts, are present. Likewise, the minimum requirement for the generation of a terminal is that there exists at least *one* word in a language that can be linked to the terminal in question, an issue to which we will come back later on. In words of Mairal and Periñán (2009a: 222-223): “A new concept should be introduced whenever there is at least one lexical unit whose meaning

---

<sup>8</sup> FunGramKB’s ontological component takes the form of an universal concept taxonomy, where universal means that every concept we could imagine has an appropriate place in this Ontology (Periñán and Arcas, 2004: 38). For more information regarding the issue of universality and its relation to culture, please see Periñán and Mairal (fc).

<sup>9</sup> Although so far English and Spanish are fully supported in the current version of FunGramKB, in the near future other languages such as German, French, Italian, Bulgarian and Catalan will also be contemplated (cf. Periñán and Mairal, 2009: 266).

does not match any of the MPs stored in the knowledge base provided that the values of the ontological properties of that new concept are shared by all lexical units which are linked to it<sup>9</sup>. In this sense, lexical gaps among different languages should not impede the creation of a terminal or a subconcept – cf. -TAKE\_SHOES in (4). Nevertheless, as advanced before, the tendency sought in FunGramKB is to relate words in English and Spanish (Italian, German, etc.), so that we avoid modeling an Ontology which is solely based on one single language. As mentioned in section 2, recall that the FunGramKB Ontology is to be defined as ‘linguistically-motivated’ although ‘language independent’, since the Ontology is involved with the semantics of lexical units, but the knowledge stored is not specific to any particular language (Periñán and Arcas, 2010a: 2668).

- c. Finally, once the knowledge engineer decides which terminal concepts might be inserted in the hierarchy, the meanings of the lexical units involved or examined need to be transformed into COREL notation, thus becoming conceptual units<sup>10</sup>.

Hence, in order to exemplify this process, we shall provide the reader with instances from the cognitive dimension #TRANSFORMATION (see Figure 2 and Appendix 6.1), whose TF, besides any possible selectional preferences, contains two obligatory participants or thematic roles: (a) a Theme (i.e. entity that transforms another entity = x1), and (b) a Referent (i.e. entity transformed by another entity = x2).

In the following subsections we will deal with one basic concept within this metaconcept, namely +TRANSLATE\_00, as well as its corresponding terminals.

---

<sup>10</sup> As stated elsewhere, FunGramKB possesses a total of about 1,300 basic concepts on the basis of which MPs and TF’s (potential) selectional preferences are constructed. In this sense, the knowledge engineer not only faces the difficulty of adapting dictionary definitions to a machine-readable syntax, but also of modeling knowledge with a limited set of basic concepts.

FIGURE 2  
**The metaconcept #TRANSFORMATION in FunGramKB**



### 3.1. From dictionaries to COREL

As advanced previously, the first, most crucial step in the elaboration of terminal concepts strongly relies on dictionaries, thesauri, and databases, which serve as the linguistic input for the development of a conceptualist taxonomy. For that matter, the FunGramKB website provides a list of English, Spanish, as well as other multilingual resources with which the engineer has to work. The following are the available online resources, although, suffice it to say, any other (written) tool can and should also be employed if necessary<sup>11</sup>:

- (i) English: *Longman Dictionary of Contemporary English*, *English Collins Dictionary and Thesaurus*, *Cambridge Advanced Learner's Dictionary*, *Merriam-Webster Dictionary*, *OneLook Dictionary*, BNC: *British National Corpus*, and COCA: *Corpus of Contemporary American English*.
- (ii) Spanish: DRAE: *Diccionario de la Lengua Española* (Real Academia), CLAVE, *Diccionario de Sinónimos y Antónimos* (Espasa Calpe), *Diccionari de sentits verbals*, ADESSE:

<sup>11</sup> Among those paper dictionaries, one can mention as extremely helpful the ones written by Casares and by María Moliner, or the *The New Oxford Thesaurus of English* (2000). Likewise, Levin (1993) and Faber and Mairal's (1999) work are also worth consulting. Full references for these resources are given in subsection 5.1 of this paper.

*Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español, Base de datos sintácticos del español actual, CREA: Corpus de Referencia del Español Actual, Corpus del Español* (Mark Davies), *Corpus Sensem*.

- (iii) Multilingual resources: *Collins Bilingual Dictionaries*, *MultiWordNet*, *Word Magic Dictionary*, *Woxikon*, *WordReference Dictionaries* and *WebCorp: linguist's search engine*.

A word of caution is required here: this list of resources is meant to be employed by knowledge engineers as well as linguists or computational lexicographers working on the lexical and grammatical modules. Among all the aforementioned tools, the most utilized ones for our purposes here are: the *Longman*, *Collins Cobuild* and *Cambridge* dictionaries (for English), and *CLAVE*, *Vox*, *María Moliner* and *Diccionario de Salamanca* (in the case of Spanish). Owing to the fact that we are developing what we may call ‘a nuclear ontology’, which aims to gather those concepts possessed by an average cultivated speaker, we recommend avoiding highly academic dictionaries such as *DRAE* or *Merriam-Webster*, since the amount of senses or information offered often exceeds the knowledge of a prototypical well-educated speaker. Being this the case, learner’s dictionaries, which are far more intuitive, are more useful when it comes to modeling conceptual knowledge<sup>12</sup>.

With these resources at hand, and, departing from a given basic concept or superordinate, the engineer, first of all, has to solely enlist those possible subordinate concepts related to the conceptual domain under scrutiny. In the case at work, i.e. +TRANSLATE\_00, it means that the knowledge engineer needs to look up the word “translate” in the available dictionaries, thesauri, corpora, etc., and write a list of the lexical items related to it, such as its synonyms, hypernyms, hyponyms, etc.:

---

<sup>12</sup> We should not forget that as a usage-base model, the use of corpora is also crucial in FunGramKB, particularly in the development of selectional preferences and the various lexica and grammaticons. For a detailed account on how to benefit from corpora when building the definitions of words in terminological ontologies, the interested reader is referred to Meyer (2001), Marshman, Morgan and Meyer (2002), and Malaisé, Zweigenbaum and Bachimont (2005), among others.

- (5) “translate”:
- (5.a) English: “construe”, “convert”, “elucidate”, “explicate”, “gloss”, “transpose”, “interpret”, “render”, “transcribe”, “transliterate”, “crack”, “decipher”, “decode”, “decrypt”, “unscramble”, “cipher”, “code”, “dub”, “subtitle”, “paraphrase”, “alter”, “transmute”, “expound”.
- (5.b) Spanish: *traducir, descifrar, descodificar, descriptar, interpretar, transcribir, transliterar, verter, trasladar, cifrar, codificar, doblar, subtitular*.

The next crucial step consists of selecting which of the items listed above shares the *genus* of “translate”, discarding those that do not. According to Faber & Mairal (1999: 59), the *genus* is the superordinate term of a domain by means of which other lexemes are directly or indirectly defined, thus, it ought to guide the engineer in the selection of the right lexical items pertaining to the domain under analysis. Out of the list in (5), predicates such as “construe”, “convert”, “elucidate”, “explicate”, “gloss”, “transpose”, “paraphrase”, “alter”, “transmute”, and “expound” will be disregarded, since they belong to a different cognitive territory, i.e. #COMMUNICATION, and thus do not share the same or closely related *genus* as that of “translate”: ‘to change written or spoken words into another language’. These lexical items refer to completely different senses: “construe” means ‘to understand a remark or action in a particular way’ (*Longman*); “elucidate”, “expound” and “gloss” could be paraphrased as ‘to explain or make clear’ (*Longman, Cambridge*); whereas “convert”, “alter” or “transmute” express ‘to change something’ (*Longman*).

We would like to insist on the fact that ontological modeling is a fairly creative, subjective process in which constant decisions need to be made. For this matter, our common sense knowledge is of vital importance in such a practice. Hence, although we need to pay attention to the *genus* of a given lexical item in order to decide whether to include such a lexical piece as a subordinate concept or not, one should not take it as a clear-cut law, but consider too those definitions that contain “the closest genus, i.e., the genus immediately above the concept in its generic-specific hierarchy” (Meyer, 2001: 286). Since more often than not, dictionary entries tend to vary in their choice of superordinates, not only various lexicographical tools must be thoroughly consulted, but also an on-going negotiation process

must be followed, on the basis of which the whole structure of this Ontology will finally be built. Below are the final subordinates extracted for the basic concept considered here:

- (6) +TRANSLATE\_00
- (6.a) English: “translate”, “interpret”, “render”, “transcribe”, “transliterate”, “crack”, “decipher”, “decode”, “decrypt”, “unscramble”, “cipher”, “code”, “dub”, “subtitle”.
- (6.b) Spanish: *traducir, descifrar, descodificar, descriptar, interpretar, transcribir, transliterar, verter, trasladar, cifrar, codificar, doblar, subtítular*.

Once the list of lexical items that will be lexicalizing the basic concept +TRANSLATE\_00 is gathered, it is time to deeply explore their semantics in order to look for differentiating parameters that could help us to conceptually narrow down the basic concept and thus elaborate possible terminal concepts. At this point, the knowledge engineer needs to separately look up the meaning of each of these lexical units and write down the definitions provided by at least three or four different dictionaries, always bearing in mind that such predicates should share the same or the closest *genus*. This is a methodological step that is somehow similarly carried out by terminologists and terminographers identifying and extracting definition-like contexts, or knowledge rich-contexts, from corpus texts (Meyer, 2001: 281-289; Marshman, Morgan and Meyer, 2002: 1-2; Malaisé, Zweigenbaum and Bachimont, 2005: 22).

After mature consideration of the meanings of each of these predicates, the decision making process begins. In this concern, there are two possible alternatives:

- (a) Whenever the semantics of a given verbal predicate does not show a clearly distinct feature which ultimately distinguishes the item in question from its basic concept, then such a predicate will be simply agglutinated as a *lexical unit* and not as a concept. This is parallel to “the ontology siblings” described in Malaisé, Zweigenbaum and Bachimont (2005).
- (b) If a given lexical piece contains semantic information which is not present in the superordinate term (e.g., +TRANSLATE\_00, in our analysis), the knowledge engineer will create a new terminal concept which will inherit those features contained in the “parent” concept.

Let us now clarify these alternatives with some examples.

### 3.2. The case of +TRANSLATE\_00

+TRANSLATE\_00 is stored with the following ontological properties in the form of its TF and MP:

- (7) TF = (x1: +HUMAN\_00 ^ +COMPUTER\_00)Theme (x2) Referent  
 MP = +(e1: +CHANGE\_00 (x1)Theme (x2)Referent (f1: (e2: +BECOME\_00 (x2)Theme (x3: +LANGUAGE\_00)Attribute)) Result (f2: (e3: past +BE\_01 (x2)Theme (x4: +LANGUAGE\_00) Attribute))Condition)

In this TF two are the obligatory participants: (i) a Theme (x1), which, codified by the basic concepts +HUMAN\_00 or +COMPUTER\_00, expresses the selectional preferences that the entity that translates is logically a human being or a computer but not, say, an animal; and (ii) a Referent (x2), which would be, broadly speaking, an entity being changed. In turn, the MP reads as follows: a human or a computer (x1) change something (x2) with the condition (i.e. satellite f1) that the quality ascribed to that entity or Attribute (i.e. x4 being in the past one language) becomes another language (x3) (i.e. the satellite f2 Result)<sup>13</sup>. A word of caution about the syntax of COREL: +BECOME\_00 is employed whenever there is a change in one of the properties that the object has (e.g. its shape, color, form, etc.), but not whenever we perceive the entity in question as experiencing a conspicuous change or as transforming into something else, in which case +BE\_01 should be utilized<sup>14</sup>. For instance, within the scenario of

---

<sup>13</sup> The manner in which MPs are interpreted is the following. We shall begin by the first event or 'e<sub>1</sub>' +CHANGE\_00, mapping each thematic role to the corresponding categories of Subject, Verb, Object in the paraphrase, e.g. a human (x1) changes something (x2). We will then move on to the satellites (f1 and f2) and the events described within them: e2 or Result and e3 or Condition. Since conditions must be logically prior to results, satellite f2 is interpreted before f1 in (7). In Appendix 6.4 the reader can find the meaning of COREL event operators, such as 'past' in (7) or 'ing' in (9). Recall that an MP is a set of one or more logically connected predications in which various concatenated cognitive situations are expressed (e<sub>1</sub> and/or e<sub>2</sub> and/or e<sub>3</sub> ... e<sub>n</sub>).

<sup>14</sup> In the syntax of COREL +BE\_00, +BE\_01, and +BE\_02 are semantic primitives or undefinable concepts. +BE\_00 means 'to belong to a class' or 'to have identity with'; +BE\_01 is employed with adjectives to express inalienable properties (e.g. +BE\_01 plus the attribute +DIFFERENT\_00); and finally, +BE\_02 should be utilized for locations. Notice that, whenever we have a concept displaying more than one meaning, and thus associated to different



+DRY\_01, the terminal \$EVAPORATE\_00 conceptualizes the fact that a liquid typically turns into gas by means of +BE\_01 in its MP, whereas +BECOME\_00 is the basic concept required in the MP of \$WITHER\_00 to accurately codify the fact that plants or fruit can gradually become weaker/smaller and start to die:

(8) \$EVAPORATE\_00: +(e1: +DRY\_01 (x1)Theme (x2: +LIQUID\_00)Referent (f1: (e2: +BE\_01 (x2)Theme (x3: +GAS\_00)Attribute))Result

(9) \$WITHER\_00: +(e1: +DRY\_01 (x1)Theme (x2: +PLANT\_00 ^ +FRUIT\_00)Referent (f1: (e2: +BECOME\_00 (x2)Theme (x3: +SMALL\_00 | +WEAK\_00)Attribute))Result (f2: (e3: ing +DIE\_00 (x2)Theme))Result)

We shall now analyze some definitions of the lexical units in (6)<sup>15</sup>:

(10) “interpret”

- a. *Longman* (L): ‘to translate spoken words from one language into another’
- b. *Collins Cobuild* (CC): ‘if you interpret what someone is saying, you translate it immediately into another language’
- c. *Cambridge* (C): ‘to change what someone is saying into another language’

(11) “render”

- a. (L): ‘to translate something into English, Russian etc.’
- b. (CC): ‘to render something in a particular language or in a particular way means to translate it into that language or in that way’
- c. (C): ‘to change words into a different language or form’

(12) “transcribe”<sup>16</sup>

- a. (L): ‘to change a piece of writing into the alphabet of another language’

---

conceptual scenarios (i.e. #IDENTIFICATION and #LOCATION), numbers 00, 01, 02 are simply employed to distinguish them.

<sup>15</sup> Due to space limitations we will only offer the English definitions for some of these lexical units, whose corresponding verbs in Spanish are traducir, interpretar, transcribir, translitar, verter and trasladar.

<sup>16</sup> Since definitions (a) and (b) were misleading in whether this predicate should or should not be included within this conceptual domain, more dictionaries were checked.

- b. (CC): ‘if you transcribe a speech or text, you write it out in a different form from the one in which it exists, for example by writing it out in full from notes or from a tape recording’
  - c. (C): ‘to change a piece of writing or music into another form, for example into a different writing system or into music for different instruments’
  - d. *Encarta World English Dictionary*: ‘to translate or transliterate something’
  - e. *Webster’s New World College Dictionary*: ‘to translate or transliterate’
- (13) “transliterate”
- a. (L): ‘to write a word, sentence etc in the alphabet of a different language or writing system’
  - b. (CC): ‘to write a word or letter in a different alphabet’
  - c. (C): ‘to write a word or letter in a different alphabet’

As pointed out in section 2, it should be highlighted that the granularity of MPs in the Ontology is not as exhaustively detailed as that in human-oriented lexicographical definitions, in the sense that there is yet no machine or NLP system capable of capturing the astonishing complexity of a human being’s world knowledge. Note, however, that the issue of granularity does not depend so much on the capability or incapability of the machine to build such a degree of semantic specificity, but rather on what is the usefulness in reaching it. In other words, the FunGramKB Ontology aims 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. Nonetheless, although MPs in FunGramKB are coarse-grained when compared with standard lexicography, such an attempt is more fine-grained than other existing ontologies (cf. section 2).

Having said this, it is our belief that the verbal predicates from (10) to (13), whether in English or Spanish, do not contain sufficient features as to construct a different terminal concept and thus, they should all be agglutinated as lexical units related to one single basic concept, i.e. +TRANSLATE\_00<sup>17</sup>. However,

---

<sup>17</sup> Since FunGramKB lexica are populated in a top-down fashion according to which “the construction of lexical entries involves the previous ontological modeling of their corresponding concepts” (Mairal and Periñán, 2009a:

this is not the case with the rest of lexical items: “decipher”, “decode”, “decrypt”, “crack”, “unscramble”, *descifrar*, *descodificar*, *desencriptar*, etc. which, from our point of view, display some characteristics which are not present in the *genus* of the superordinate term (i.e. ‘to change written or spoken words into another language’). According to various dictionaries, these verbs express the idea that:

(14) (L): ‘the message we are trying to understand or read is written in *code*’

(CC): ‘the message is *difficult to read or understand*’

(C): ‘the message is written *badly or in a difficult or hidden way*’

Since such differences (in italics for your convenience) are outstanding enough, the knowledge engineer will create a new terminal concept, which within the taxonomy would ‘hang’ under its most immediate superordinate, i.e. +TRANSLATE\_00. Thus, the first predication of any subordinate concept of +TRANSLATE\_00 will specify the superordinate concept to which it is linked and its corresponding TF. In addition to this, we shall add the *differantiae* expressed in the form of satellites (f1, f2...fn). Therefore, at this stage, we basically have to transform dictionary entries into COREL. This process is carried out as follows:

- (a) After naming the newly added terminal, say \$DECIPHER\_00<sup>18</sup>, we proceed to fill in the information that concerns the number and type of participants involved in the TF, i.e. (x1)Theme and (x2)Referent, which in this example are restricted by the following selectional preferences:

---

221), the task of computational lexicographers or linguists is to specify the morphosyntactic, pragmatic and collocational information of those predicates which are agglutinated as lexical units in the Ontology (cf. section 2). Thus, although “interpret”, “render”, “transliterate”, “transcribe”, *traducir*, *interpretar*, *transcribir*, *translitarar*, *verter* and *trasladar* all depend on the same basic concept, i.e. +TRANSLATE\_00, their individual differences will be examined in each lexicon.

<sup>18</sup> Terminal concepts and subconcepts are usually named after the most prototypical item or short paraphrase that lexicalizes them (cf. \$SPORT\_00 or -TAKE\_SHOES\_00). Notice, however, that English is used as a metalanguage for the convenience of engineers, linguists, users, etc., and that a label such as \$8X92D\_00 would be equally meaningful for the machine as \$DECIPHER\_00.

- (15) TF = (x1: +HUMAN\_00 ^ +COMPUTER\_00)Theme  
 ('the entity that translates is either a human or a computer')  
 (x2: +INFORMATION\_OBJECT)Referent  
 ('the entity being translated is a symbol')
- (b) Once this is done, we move on to the construction of the MP. As advanced before, the superordinate concept is expressed or referred in the first predication of its corresponding subordinates: +TRANSLATE\_00:
- (16) *Genus*: (e1: +TRANSLATE\_00 (x1)Theme (x2)Referent  
 ('a human or a computer translate a symbol')
- (c) Adding to this, we now need to formalize in COREL the purpose of the deciphering action through the appropriate satellite(s), that is to say, the fact that you want to understand a message that is difficult and/or hidden. In practical terms, it means that the engineer should check if this dictionary definition can somehow be *translated* using basic concepts; in other words, whether key words such as "understand", "difficult" or "hidden" find a correlate among the 1,300 concepts of FunGramKB. This has certainly been achieved in the satellite f1 below:
- (17) *Differentiae*: (f1: (e2: +UNDERSTAND\_00 (x1)Theme (x2)Referent))Purpose)(e3: +BE\_01 (x2)Theme (x3: +SECRET\_00 | +DIFFICULT\_00)Attribute)  
 ('the purpose of this action is to understand something (x2) which is secret and/or (codified with the disjunction logical connector "|") difficult')

Accordingly, the resulting MP definition of \$DECIPHER\_00 is:

- (18) \*((e1: +TRANSLATE\_00 (x1)Theme (x2)Referent (f1: (e2: +UNDERSTAND\_00 (x1)Theme (x2)Referent))Purpose)(e3: +BE\_01 (x2)Theme (x3: +SECRET\_00 | +DIFFICULT\_00)Attribute))

At this point, there are some important ideas worth signaling (cf. Periñán and Mairal, 2010 for more details):

- Every predication (e1, e2, etc.) must be followed by an event plus its corresponding TF. For instance, the basic concept +UNDERSTAND\_00 above belongs to the metaconcept #COGNITION, which, as shown in Appendix 6.1, implies

that all the concepts linked to it are always associated with three possible participants: (i) an Agent, i.e. the entity that makes another entity undergo a cognitive process and that may be optional in some scenarios; (ii) a Theme, i.e. the entity that undergoes a cognitive process; and (iii) a Referent, which is the entity present in the consciousness of an entity that undergoes a cognitive process. The indexation of these three participants will be adapted to the MP in question, that is, the one of \$DECIPHER\_00, which explains the omission of the Agent in (18).

- The 'fs' or satellites (e.g. Manner, Purpose, Reason, Means...), stored in Appendix 6.3, can be followed by either a basic concept (cf. (f1: +BODY\_AREA\_00)Location in (2)) or by another predication (f1: (e2: +UNDERSTAND and its thematic roles)).
- As stated in Periñán and Arcas (2004), every predication can be preceded by two different *reasoning operators*: strict (+(e1...)) or defeasible (\*(e1...)). The former can be argued to exhibit law-like rules according to which the event defined in that predication contains no exceptions (e.g. 'whales are mammals'). As opposed to this, the latter, which covers fuzzy concepts, may be overridden in the light of some contradictory information (e.g. 'birds typically fly. A penguin is a bird, but penguins do not fly'). FunGramKB's inference engine, which handles predications as rules, allows *monotonic* reasoning with strict predications, while *non-monotonic* is employed with defeasible predications. In the case of \$DECIPHER\_00, we selected a defeasible operator since not everything that is deciphered obligatorily implies the attribute +SECRET\_00 and/or +DIFFICULT\_00. On the contrary, the MP of +TRANSLATE\_00 in (7) is headed by a + symbol or strict predication because, whenever one translates a (written or spoken) text, it always implies changing from one language to another.
- It should be noticed that the MP of \$DECIPHER\_00 includes two parentheses before the first predication to represent what is known as 'conceptual binding'. Consistent with this, whenever two predications are enclosed between two parentheses, it is signaled that there exists a direct connection between two of the participants involved in the predications, so, through co-indexation, one of the participants can modify the other. That is to say, conceptual

binding in the example discussed here implies that what the Theme is trying to understand (i.e. (x2)Referent) *is* or *has the attribute* of being secret and/or difficult. Following this rational, (x3: +SECRET\_00 | +DIFFICULT\_00)Attribute) is represented as modifying (x2). To further clarify this idea, let us provide another example from the terminal \$MOP\_00, located under the basic concept +WASH\_00, whose TF and MP are, respectively:

- 
- (19) TF = (x1: +HUMAN\_00)Theme (x2: +FLOOR\_00)Referent  
 ('a human entity washes a floor')  
 MP = +((e1: +WASH\_00 (x1)Theme (x2)Referent (f1:  
 +CLOTH\_00)Instrument)(e2: +BE\_01 (x3: f1)Theme (x4:  
 +WET\_01)Attribute))  
 ('a human washes a floor using as an instrument (f1) a  
 cloth that *is* wet (e2)')

Hence, in the syntax of COREL conceptual binding is employed in order to link two events by means of the use of double parenthesis. Such a procedure allows the machine to understand that there is a specific type of relation between two participants expressed in separate, but logically connected satellites.

- Finally, once the new terminal concept, \$DECIPHER\_00, is created, we will have to associate those lexical units expressing the same or at least similar lexical meanings. Thus, “decipher”, “crack”, “decode”, “decrypt”, “unscramble”, *descifrar*, *descodificar* and *desencriptar* depend on this same terminal concept, whereas “translate”, “render”, “transcribe”, “interpret”, “transliterate”, *traducir*, *interpretar*, *transcribir*, *transliterar*, *trasladar* and *verter* are all linked to the basic concept +TRANSLATE\_00.

The following figure shows how this whole process is finally depicted in the FunGramKB Ontology:

In closing this sub-section, there are still three more terminal concepts we should mention, i.e. \$CIPHER\_00, \$DUB\_00 and \$SUBTITLE, each of which, as signaled elsewhere, obviously displays well-marked *differentiae* and whose creation has thoroughly followed the already mentioned steps:

- (20) \$CIPHER\_00: 'to convert ordinary language into code'  
 TF = (x1: +HUMAN\_00 ^ +COMPUTER\_00)Theme (x2)  
 Referent

FIGURE 3  
**Conceptual information of the basic concept +TRANSLATE\_00**



MP = \*(e1: +TRANSLATE\_00 (x1)Theme (x2)Referent (f1: (e2: n +UNDERSTAND\_00 (x3: +HUMAN\_00)Theme (x2)Referent))Purpose)

(‘a human entity or a computer (x1) translate something (x2) so that another human entity (x3) does not understand (n +UNDERSTAND\_00) it (x2)’).

- Related lexical units: “code”, “cipher”, *cifrar*, *codificar*.

(21) \$DUB\_00: ‘to change the original spoken language of a film or television program into another language’

TF = (x1: +HUMAN\_00)Theme (x2: +SPEECH\_00)Referent

MP = +(e1: +TRANSLATE\_00 (x1)Theme (x2)Referent (f1: (e2: +LISTEN\_00 (x3: +HUMAN\_00)Theme (x2)Referent (f2: +MOVIE\_00)Location (f3: +IN\_00)Position))Purpose)

(‘a human entity (x1) translates speech (x2) so that (f2: Purpose) another human (x3) listens to the translated speech (x2) now in a movie (f2: Location)’)

- Related lexical units: “dub”, *doblar*.

(22) \$SUBTITLE: ‘to translate a foreign dialogue of a movie or TV program, usually displayed at the bottom of the screen’

TF = (x1: +HUMAN\_00)Theme (x2: +SPEECH\_00)Referent

MP = +((e1: +TRANSLATE\_00 (x1)Theme (x2)Referent (f1: (e2: +BECOME\_00 (x2)Theme (x3: +WRITING\_00)Attribute))Result) (e3: +SHOW\_00 (x4)Theme (x3)Referent (f2: +MOVIE\_00)Location (f3: +IN\_00)Position))

(‘someone translates speech into writing and that appears located in a movie’)

- Related lexical units: “subtitle”, *subtitular*.

Let us conclude by expounding on this rather complex MP. The first predication contains a theme role or human entity who translates (a piece of) speech. As a result of this action (i.e. satellite ‘f1’), what was spoken (i.e. the Referent role in the first predication which tallies with the Theme role of the second event) becomes written instead (+BECOME\_00 (x2)Theme (x3: +WRITING\_00)Attribute). Now, if we want to express that the writing appears located in a movie, we need to make use of conceptual binding, so that through co-indexation we are able to specify a direct connection between (x3: +WRITING\_00) and the (x3)Referent of the third event or e3. In other words, through the Location and Position satellites (f2: +MOVIE\_00 and f3: +IN\_00) inside the e3 that uses the basic concept +SHOW\_00, we can neatly capture that the writing *is shown* in a movie.

#### 4. Concluding remarks

This paper has made explicit the detailed methodology necessary for knowledge engineers to elaborate the terminal concepts that populate the FunGramKB Ontology, specifically, to build conceptual meaning within the ontological domain +TRANSLATE\_00. It is precisely now, at this stage, that linguists and computational lexicographers, guided by the theoretical underpinnings of the LCM, will be able to type in all the morphosyntactic, pragmatic and collocational information of verbs such as English “decipher”, “crack”, “decode”, or their Spanish correspondences *desencriptar*, *descifrar*, *descodificar* in the corresponding lexica. Consequently, a steady collaboration between these two solid approaches to language, i.e. FunGramKB and the LCM, can only but bring advantages to both the world of Artificial Intelligence and that of Linguistics.

#### 5. Bibliographic references

FABER, Pamela and Ricardo MAIRAL, 1999: *Constructing a Lexicon of English Verbs*, Berlin; New York: Mouton de Gruyter.



- GANGEMI, Aldo *et al.*, 2002: "Sweetening ontologies with DOLCE" in Asunción GÓMEZ-PÉREZ, and Richard BENJAMINS (eds.): *Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web: 13th International Conference, EKAW 2002*, Sigüenza, Spain, October 1-4.
- JIMÉNEZ, Rocío and Beatriz PÉREZ, 2010: "An account of selection restrictions within a conceptual framework: Its relevance for Role and Reference Grammar" in *Proceedings of the 2009 International Conference on Role and Reference Grammar* [available from [http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/ProceedingsofRRG2009\\_02.pdf](http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/ProceedingsofRRG2009_02.pdf)], 131-146.
- LENCI, Alessandro *et al.*, 2000: "SIMPLE: A general framework for the development of multilingual lexicons", *International Journal of Lexicography* 13-4, 249-263.
- LEVIN, Beth, 1993: *English Verb Classes and Alternations*, Chicago: University of Chicago Press.
- MAIRAL, Ricardo and Carlos PERIÑÁN, 2009a: "The anatomy of the lexicon component within the framework of a conceptual knowledge base", *Revista Española de Lingüística Aplicada* 22, 217-244.
- , 2009b, in press: "Role and Reference Grammar and ontological engineering" in *Volumen Homenaje a Enrique Alcaraz*, Alicante: Universidad de Alicante,
- , 2010: "Teoría lingüística y representación del conocimiento: una discusión preliminar" in Dolores GARCÍA PADRÓN and María del Carmen FUMERO PÉREZ (eds.): *Tendencias en lingüística general y aplicada*, Berlin: Peter Lang, 155-168.
- MAIRAL, Ricardo and Francisco José RUIZ DE MENDOZA, 2008: "New challenges for lexical representation within the Lexical-Constructional Model", *Revista Canaria de Estudios Ingleses* 57, 137-158.
- , 2009: "Levels of description and explanation in meaning construction" in Christopher BUTLER and Javier MARTÍN ARISTA (eds.): *Deconstructing Constructions*, Amsterdam/Philadelphia: John Benjamins, 153-198.
- MALAISE, Véronique, Pierre ZWEIGENBAUM and Bruno BACHIMONT, 2005: "Mining defining contexts to help structuring differential ontologies", *Terminology* 11-1, 21-53.
- MARSHMAN, Elizabeth, Tricia MORGAN and Ingrid MEYER, 2002: "French patterns for expressing concept relations", *Terminology* 8-1: 1-29.
- MASOLO, Claudio *et al.*, 2003: "WonderWeb Deliverable D18: Ontology Library". Laboratory for Applied Ontology, ISTC-CNR.
- MEYER, Ingrid, 2001: "Extracting a knowledge-rich contexts for terminology: A conceptual and methodological framework" in Didier BOURIGAULT, Christian JAQUEMIN and Marie-Claude L'HOMME (eds.): *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia: John Benjamins, 279-302.

- NILES, Ian and Adam PEASE, 2001a: "Origins of the Standard Upper Merged Ontology: A Proposal for the IEEE Standard Upper Ontology" in *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology*. Seattle.
- , 2001b: "Towards a standard Upper Ontology" in *Proceedings of the Second International Conference on Formal Ontology in Information Systems*. Ogunquit.
- NOY, Natalya and Deborah MCGUINNESS, 2001: "Ontology Development 101: A Guide to Creating Your First Ontology", technical report KSL-01-05, Stanford Knowledge Systems Laboratory, Stanford University.
- PEDERSEN, Bolette Sandford and Britt KESON, 1999: "SIMPLE – Semantic Information for Multifunctional Plurilingual Lexica: Some examples of Danish concrete nouns" in *Proceedings of the SIGLEX-99 Workshop*. Maryland.
- PERIÑÁN, Carlos and Francisco ARCAS, 2004: "Meaning postulates in a lexico-conceptual knowledge base" in *15th International Workshop on Databases and Expert Systems Applications*, IEEE, Los Alamitos (California), 38-42.
- , 2005: "Microconceptual-Knowledge Spreading in FunGramKB" in *9th IASTED International Conference on Artificial Intelligence and Soft Computing*, Anaheim-Calgary-Zurich: ACTA Press, 239-244.
- , 2007a: "Cognitive modules of an NLP knowledge base for language understanding". *Procesamiento del Lenguaje Natural* 39, 197-204.
- , 2007b: "Deep semantics in an NLP knowledge base" in Daniel BORRAJO, Luis CASTILLO and Juan Manuel CORCHADO (eds.): *12th Conference of the Spanish Association for Artificial Intelligence*, Salamanca: Universidad de Salamanca, 279-288.
- , 2010a: "The Architecture of FunGramKB" in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Malta, 2667-2674.
- , 2010b: "Ontological commitments in FunGramKB", *Procesamiento del Lenguaje Natural* 44, 27-34.
- PERIÑÁN, Carlos and Ricardo 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", *Onomázein* 21, 11-45.
- , (fc) "Prototypicality, universality and culturality in an NLP knowledge base".
- RUIZ DE MENDOZA, Francisco José and Ricardo MAIRAL, 2008: "Levels of description and constraining factors in meaning construction: An introduction to the Lexical Constructional Model". *Folia Linguistica* 42-2, 355-400.

## 5.1. Dictionaries, data bases and corpora

- ADESSE: *Base de datos de verbos, alternancias de diátesis y esquemas sintáctico-semánticos del español* [available online from <http://adesse.uvigo.es/index.html>, June-July 2009].
- BDS: *Base de datos sintácticos del español actual* [available online from <http://www.bds.usc.es/>]
- BYU-BNC: DAVIES, Mark *British National Corpus* [available online from <http://corpus.byu.edu/bnc/>, September-December 2009].
- CAMBRIDGE: *Cambridge Advanced Learners Dictionary* [available online from <http://dictionary.cambridge.org/default.asp?dict=CALD>, June-July 2009].
- CASARES: CASARES, Julio, 2004: *Diccionario ideológico de la lengua española*. Barcelona: Editorial Gustavo Gili, SA.
- CLAVE: MALDONADO GONZALEZ, Concepción (ed.), 2005: *Clave. Diccionario de Uso del Español Actual*. Madrid: SM.
- COCA: DAVIES, Mark *The Corpus of Contemporary American English* [available online from <http://www.americancorpus.org/>, September-December 2009].
- COLLINS BILINGUAL: *Collins Bilingual Dictionaries* [available online from <http://dictionary.reverso.net/>, September-December 2009].
- COLLINS THESAURUS [available online from <http://dictionary.reverso.net/>, June-July 2009].
- CREA: REAL ACADEMIA ESPAÑOLA. *Corpus de referencia del español actual* [available online from <http://corpus.rae.es/creanet.html>, June-July 2009].
- DRAE: REAL ACADEMIA ESPAÑOLA. *Diccionario de la Lengua Española* [available online from <http://www.rae.es>, June-July 2009].
- ENCARTA: *Encarta World English Dictionary* [available online from <http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx>, September-December 2009].
- ESPASA: *Diccionario de Sinónimos y Antónimos* [available online from <http://www.wordreference.com/sinonimos/>, September-December 2009].
- GRIAL: *Diccionari de sentits verbals* [available online from <http://grial.uab.es/sentits/>].
- GRIAL: *Corpus SenSem* [available online from <http://grial.uab.es/search/index.php>].
- LONGMAN: PROCTER, Paul (ed.), 1978: *Longman Dictionary of Contemporary English*. Harlow (Essex): Longman.
- LONGMAN: *Longman Dictionary of Contemporary English* [available online from <http://www.ldoceonline.com/>, June-July 2009].
- MERRIAM: *Merriam-Webster Dictionary* [available online from <http://www.merriam-webster.com/>, September-December 2009].
- MOLINER: MOLINER, María, 2001: *Diccionario de Uso del Español. Edición Electrónica*. Madrid: Gredos.

- MULTIWORDNET: *MultiWordNet* [available online from <http://multiwordnet.fbk.eu/english/home.php>, June-July 2009].
- ONELOOK: *OneLook Dictionary* [available online from <http://www.onelook.com/>, September-December 2009].
- OXFORD THESAURUS: HANKS, Patrick (ed.), 2000: *The New Oxford Thesaurus of English*. Oxford: OUP.
- SPANISH CORPUS: DAVIES, Mark. *Corpus del Español* [available online from <http://www.corpusdelespanol.org/>, September-December 2009].
- VOX: ALVAR EZQUERRA, Manuel, 1995: *Diccionario para la enseñanza de la lengua española*. Barcelona: Bibliograf.
- WEBCORP: *WebCorp Linguist's Search Engine* [available from <http://www.webcorp.org.uk/>].
- WEBSTER: AGNES, Michael (ed.), 2004: *Webster's New World College Dictionary*. Webster's New World.
- WORDMAGIC: *Word Magic Dictionary* [available online from <http://www.wordmagicsoft.com/diccionario/en-es/>, September-December 2009)].
- WORDREFERENCE: *WordReference Online Language Dictionaries* [available online from <http://www.wordreference.com/>, June-July 2009].
- WOXIKON: *Woxikon Free Online Dictionary-Translation* [available online from <http://www.woxikon.com/>, June-July 2009].

## 6. Appendixes

### 6.1. Semantic interpretation of the thematic roles in the dimensions under analysis

Metaconcept	Role	Definition
#COGNITION	[Agent] Theme Referent	Entity that makes another entity undergo a cognitive process. Entity that undergoes a cognitive process. Entity present in the consciousness of an entity that undergoes a cognitive process.
#POSSESSION	Theme Referent	Entity that owns another entity Entity that is owned
#TRANSFORMATION	Theme Referent	Entity that transforms another entity Entity that is transformed by another entity

### 6.2. COREL logical operators

Type	Connector	Scope
Conjunction	&	Between predications, satellites, or selectional preferences.
Disjunction		Between predications, arguments, or satellites.
Exclusion	^	Between satellites or selectional preferences.

### 6.3. Semantic interpretation of satellites

(Adapted from Periñán and Mairal, 2010: 34)

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.

Role	Definition
Location	Entity that describes the place of an argument.
Position	Quality that describes the position of Theme with respect to Location, Goal or Origin.
Purpose	Predication that describes the aim of the event.
Result	Predication or entity that describes the consequence of the occurrence of the event.

### 6.4. Event operators

(Adapted from Periñán and Arcas, 2004: 40; Periñán and Mairal, 2010: 22-24)

Feature		Value	Example
Aspectuality		ing	John started crying
		pro	John was crying
		egr	John stopped crying
Temporality		rpast	Mary had sung
		past	Mary sang
		npast	Mary has just sung
		pres	Mary is singing
		nfut	Mary is about to sing
		fut	Mary will sing
		rfut	Mary will sing
Modality	Epistemic	cert   prob   pos	Mary will   should   might sing
	Non-epistemic	obl   adv   perm	Mary must   ought to   can sing
Polarity		n	Mary is not singing