ONOMÁZEIN 24 (2011/2): 13-33

The COHERENT Methodology in FunGramKB

Carlos Periñán-Pascual
Universidad Católica San Antonio
España

Ricardo Mairal-Usón
Universidad Nacional de Educación a Distancia
España

Abstract

Recent research has been done synergistically between FunGramKB, a lexical-conceptual knowledge base, and the Lexical Constructional Model, a linguistic meaning construction model. Since concepts are claimed to play an important role in the design of the cognitive-linguistic interface, this paper discusses the methodology adopted in structuring the basic conceptual level in the FunGramKB Core Ontology. More particularly, we describe our four-phase COHERENT methodology (i.e. COnceptualization + HiErarchization + REmodelling + refinemeNT), which guided the cognitive mapping of the defining vocabulary in Longman Dictionary of Contemporary English.

Keywords: FunGramKB; ontology; concept; natural language processing.

Afiliaciones: Carlos Periñán-Pascual: Departamento de Idiomas, Universidad Católica San Antonio, Murcia. España. — Ricardo Mairal-Usón: Departamento de Filologías Extranjeras y sus Lingüísticas, Universidad Nacional de Educación a Distancia. Madrid, España.
Correos electrónicos: jcperinan@pdi.ucam.edu; rmairal@flog.uned.es.
Dirección postal: Carlos Periñán-Pascual: Unidad Central de Idiomas. Universidad Católica San Antonio. Campus de los Jerónimos. E - 30107 Guadalupe (Murcia). España.

Fecha de recepción: abril de 2011
Fecha de aceptación: septiembre de 2011

1. Introduction[1]

As widely shown in recent research (Mairal-Usón and Periñán-Pascual, 2009; Periñán-Pascual & Mairal-Usón, 2009, 2010), the design of a multipurpose lexical-conceptual knowledge base like FunGramKB[2] (Periñán-Pascual & Arcas-Túnez 2004, 2007, 2010b) provides a rich explanatory framework where to anchor a broad meaning construction model of language like the Lexical Constructional Model[3] (LCM) –cf. Mairal-Usón & Ruiz de Mendoza (2009), Ruiz de Mendoza and Mairal-Usón (2008, 2011). As a result, a conceptual approach to meaning construction is advocated, a methodological strand that has also been central in both formal and functional linguistic models, e.g. Jackendoff (1990), Levin and Rappaport (2005), Pustejovsky (1995), Reinhart (2006), or Van Valin (2005). However, to the best of our knowledge, none of these models have explicitly developed a knowledge base that fully interacts with the linguistic module, which includes both a lexicon and the syntactic apparatus. Hence, the methodological claim that meaning should be seen as lying at the interface of grammar, communication and cognition has been taken far enough in FunGramKB so as to make it a strong methodological dogma.

The overall architecture of the model establishes a clear-cut demarcation between the linguistic and the conceptual levels. This division of labour between what goes in the conceptual level and what goes in the linguistic level is also indicative of a further distinction that concerns those theoretical aspects that are universal and language independent versus those aspects that are language specific. Thus, the linguistic level is connected up with a repository of conceptual knowledge, whose linkage is actually represented by means of what we have called *conceptual logical structures*[4] (hereafter, CLS), i.e. a semantic syntax-motivated

1 Financial support for this research has been provided by the DGI, Spanish Ministry of Education and Science, grant FFI2008-05035-C02-01/FILO. The research has been co-financed through FEDER funds.
2 www.fungramkb.com
3 www.lexicom.es
4 CLSs are inspired on the logical structures in Role and Reference Grammar (Van Valin & LaPolla, 1997; Van Valin, 2005). For an account of the motivation of CLSs within the framework of RRG, we refer the reader to Mairal-Usón, Periñán-Pascual & Pérez (in press).

formalism. As advanced above, although most lexical representation approaches posit primitives, which are said to have an ontological status as part of a predicate's lexical entry (i.e. the Role and Reference Grammar *logical structures*, Levin and Rappaport's *event structure templates*, or Pustejovsky's lexical entries within a generative lexicon), CLSs are proved to have a clear ontological grounding, since they are made of concepts that stem from the FunGramKB Ontology. Hence, the role of a CLS is to serve as a bridge between the more abstract level as represented in the Ontology and the particular idiosyncrasies as coded in a given linguistic expression. Therefore, CLSs are used as the interface between the semantic structure and the syntactic representation of sentences (cf. Periñán-Pascual & Mairal-Usón, 2009).

Consequently, if concepts are the building blocks for the linguistic-conceptual interface, a solid methodology for the structuring and modelling of this conceptual knowledge should be mandatory in FunGramKB. In this respect, Periñán-Pascual & Arcas-Túnez (2010a) described seven ontological commitments to which the FunGramKB Ontology is subject, i.e. ontology development guidelines concerning the structuring of the ontological model as well as the elements to be included and their ontological properties. This paper portrays the identification process of the basic concepts in the FunGramKB Core Ontology by means of the four-phase COHERENT methodology: COnceptualization, HiErarchization, REmodelling and refinemeNT. However, before doing that in Section 3 and 4, Section 2 presents a brief theoretical context as to the architecture of this knowledge base.

2. The scientific framework

FunGramKB is viewed as a multipurpose lexico-conceptual knowledge base for natural language processing systems and natural language understanding. The knowledge base is made up of three major knowledge levels, consisting in turn of several independent but interrelated modules. As shown in Periñán-Pascual & Arcas-Túnez (2010b), these are:

a) The linguistic level (linguistic knowledge):

a.1) Lexical level:

- The *Lexicon* stores morphosyntactic, pragmatic[5] and collocational information about lexical units.
- The *Morphicon* handles cases of inflectional morphology.

a.2) Grammatical level[6]:

- The *Grammaticon* stores the constructional schemata which help Role and Reference Grammar to construct the semantics-to-syntax linking algorithm (Van Valin & LaPolla, 1997; Van Valin, 2005). The Grammaticon is composed of several *Constructicon* modules that are inspired in the four levels of meaning construction formulated in the LCM:

  (i) an argument structure layer, which contains CLSs and argument structure constructions;

  (ii) an implicational level, with constructional configurations, based on low-level situational models (or scenarios), which contain fixed and variable elements where the default meaning interpretation carries a heavily conventionalized implication;

5 Brian Nolan (personal communication) questions our assumption of including pragmatic information within the lexicon since typically, in his view, pragmatics is the domain of meaning use in a discourse context and consequently should be outside the scope of the lexical module. He goes on to suggest that this information should be located at a metalevel. He is right and in fact the LCM provides the exact locus to deal with this type of pragmatic information, i.e. levels 2, 3 and 4 in the Grammaticon. However, the type of pragmatic information we include as part of a lexical entry concerns cultural distinctive features which happen to differentiate conceptual and lexical information. The actual treatment of this theoretical issue (i.e. “cultural distinctiveness”) in a knowledge base is in fact a future topic of research we would like to deal with in a different paper.

6 An important advantage of the LCM is that it clearly distinguishes amongst different dimensions of meaning construction other than the lexical and the argument structure dimensions. It does this by recognizing four representational layers, each of which can encompass lower-level layers, if licensed to do so by a number of explicit constraints. The LCM provides explanatory tools to explain the pervasive nature of implicational, illocutionary and discursive layers of meaning. For a description of the knowledge representation in the Grammaticon, we refer the reader to Mairal-Usón, Ruiz de Mendoza & Periñán-Pascual (in press).

(iii) an illocutionary level, which features illocutionary constructions, with fixed and variable elements based on high-level situational models;

(iv) a discourse level, which deals with cohesion and coherence phenomena from the point of view of the activity of discourse constructions based on high-level non-situational cognitive models like reason-result, cause-effect or condition-consequence.

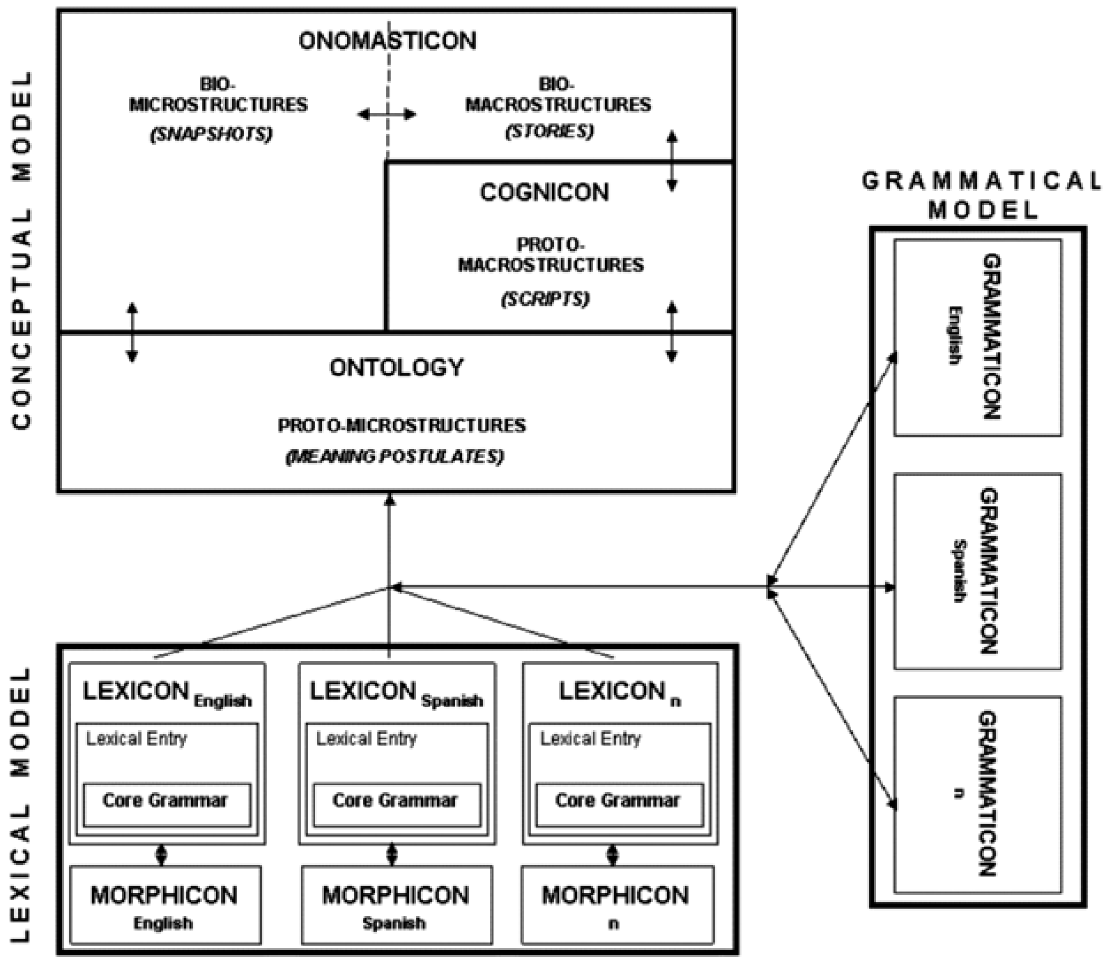
b) The conceptual level (non-linguistic knowledge)

- The *Ontology* is presented as a hierarchical catalogue of the concepts that a person has in mind, so here is where semantic knowledge is stored in the form of meaning postulates. The Ontology consists of a general-purpose module (i.e. Core Ontology) and several domain-specific terminological modules (i.e. Satellite Ontologies).
- The *Cognicon* stores procedural knowledge by means of scripts, that is, conceptual schemata in which a sequence of stereotypical actions is organised on the basis of temporal continuity, and more particularly on Allen's temporal model (Allen, 1983; Allen and Ferguson, 1994).
- The *Onomasticon* stores information about instances of entities and events such as the Beatles or La Alhambra de Granada. This module stores two different types of schemata (i.e. snapshots and stories), since instances can be portrayed synchronically or diachronically[7].

[7] Unlike other FunGramKB modules, the population of the Onomasticon is taking place semi-automatically, by exploiting the DBpedia knowledge base (Bizer *et al.*, 2009). The DBpedia project is intended to extract structured information from Wikipedia, turn this information into a rich knowledge base, which currently describes more than 2.6 million entities, and make this knowledge base accessible on the Web. The population process of the Onomasticon is being performed by means of template-based rules which can map the knowledge stored in the DBpedia ontology into COREL-formatted schemata. To illustrate this mapping process, we refer the reader to García Carrión (2010), which includes the inventory of mapping rules for those entities in the category PLACE.

Figure 1 offers a view of the whole architecture and the way the three levels are interconnected.

FIGURE 1
The FunGramKB architecture



2.1. Concepts and conceptual properties

The FunGramKB Ontology distinguishes three different conceptual levels, each one of them with concepts of a different type:

(i) Metaconcepts, preceded by symbol # (e.g. #ABSTRACT, #COMMUNICATION, #MATERIAL, #PHYSICAL, #PSYCHOLOGICAL, #QUANTITATIVE, #SOCIAL, etc), constitute the upper level in the taxonomy. The result amounts to forty-two metaconcepts distributed in three subontologies: #ENTITY, #EVENT and #QUALITY.

(ii) Basic concepts, preceded by symbol + (e.g. +READY_00, +DIRTY_00, +BALL_00, +BARRIER_00, +BLADE_00, +THINK_00, +DREAM_00, +HAVE_00, etc), are used in FunGramKB as defining units which enable the construction of meaning postulates for basic concepts and terminals, as well as taking part as selectional preferences in thematic frames.

(iii) Terminals (e.g. $AUCTION_00, $WATCH_00, $HOSE_00, $SKYLIGHT_00, $RECONSIDER_00 etc) are headed by the symbol $. The borderline between basic concepts and terminals is based on their definitory potential to take part in meaning postulates. Hierarchical structuring of the terminal level is practically non-existent.

Basic and terminal concepts in FunGramKB are provided with semantic properties which are captured by *thematic frames* and *meaning postulates*. Every event in the Ontology is assigned one single thematic frame, i.e. a conceptual construct which states the number and type of participants involved in the prototypical cognitive situation portrayed by the event (Periñán-Pascual & Arcas-Túnez, 2007). Moreover, a meaning postulate is a set of one or more logically connected predications (e_1, e_2, ...e_n), i.e. conceptual constructs that represent the generic features of concepts[8]. As stated above, the basic concepts are the main building blocks of these types of constructs in the Core Ontology. Hence, a further question is to ascertain how we actually arrived at these conceptual units, i.e. if there is any standardized procedure used by the FunGramKB knowledge engineer. In connection with this, we present the COHERENT methodology.

3. The COHERENT methodology

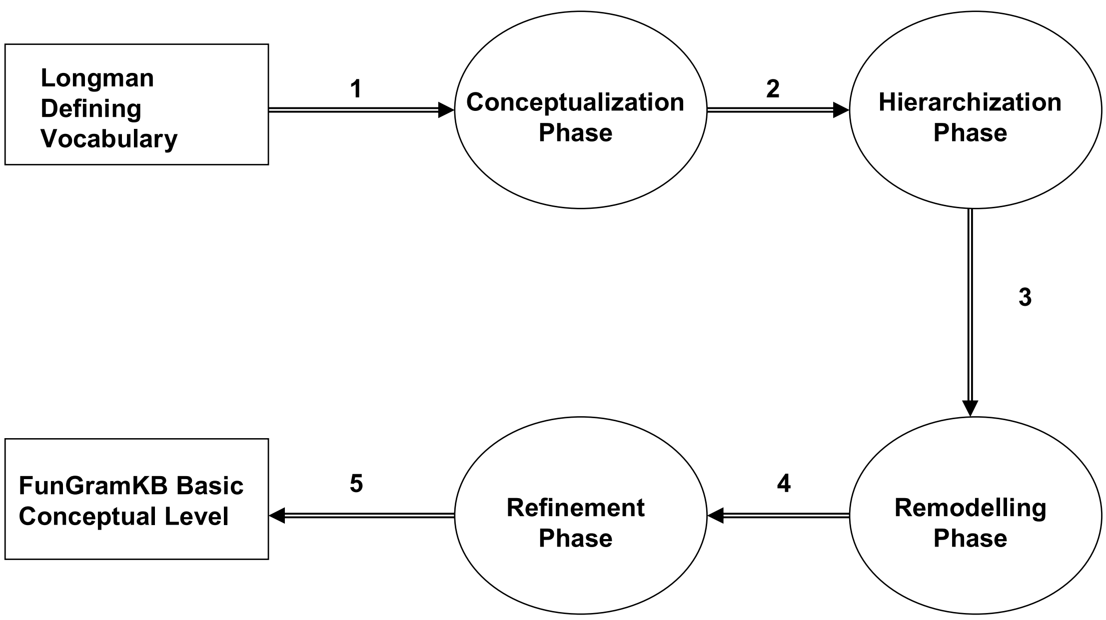
Instead of adopting a strong approach like that represented by the Natural Semantic Metalanguage (cf. Goddard & Wierzbicka, 1994, 2002; Goddard, 2008), which identifies a reduced inventory of semantic primitives that are used to represent meaning, FunGramKB posits an inventory of basic concepts which can

8 We refer the reader to Periñán-Pascual & Mairal-Usón (2010) for examples of conceptual representation in the form of thematic frames and meaning postulates.

be used to define any word in any of the European languages that are claimed to be part of the knowledge base[9]. In what follows, we shall like to focus on the methodology used for the construction of the basic conceptual level in the Core Ontology.

The FunGramKB basic concepts were identified by means of the Longman Defining Vocabulary (LDV) from *Longman Dictionary of Contemporary English* (Procter, 1978), which has been deemed to be a useful source of basic vocabulary for an artificial language. However, deep revision was required in order to perform the conceptual mapping. More particularly, both the population and the structuring of the basic conceptual level in the Core Ontology were handcrafted following our four-phase COHERENT methodology. Figure 2 illustrates the whole process of construction of this basic conceptual level.

FIGURE 2
The COHERENT methodology



(1) List of English lexical units.
(2) Inventory of cross-lingual conceptual units.
(3) Hierarchical taxonomy of basic concepts, provided with their meaning postulate and thematic frame.
(4) Conceptual taxonomy including subconcepts.
(5) Refined basic level in the Core Ontology.

---

9 English and Spanish are fully supported in the current version of FunGramKB, although we have just begun to work with other languages, such as German, French, Italian, Bulgarian and Catalan.

3.1. The conceptualization phase

The starting point of the whole process was the LDV, i.e. an inventory of about 2,197 English lexical units which facilitate the semantic description of any type of word. Our motivation was to perform a conceptual mapping of the LDV, i.e. the list of English words had to be converted into an inventory of interlingual conceptual units. From the very beginning, it was evident that this was not a one-to-one mapping, so a set of tasks were carried out in order to apply (1) lexical rejection (i.e. some LDV words were not mapped into basic concepts but terminal ones) and (2) cognitive clustering (i.e. some LDV words were grouped into the same basic concept). As far as lexical rejection is concerned, the following tasks were performed:

Task 1.1. Not only were functional words rejected, i.e. conjunctions, prepositions, determiners and pronouns, but also partitive nouns[10], modal verbs, and numerals. The lexical instanciation of quantification, aspectuality, temporality and modality in the LDV was also ignored in this conceptual mapping, since this type of meanings is expressed by means of COREL operators (cf. Periñán-Pascual & Mairal-Usón, 2010).

Task 1.2. Full-content words belonging to the lexicographical metalanguage, e.g. words such as *adjective, article, grammar, noun, verb,* etc, were also rejected. Unlike dictionary definitions, where some usage and grammatical remarks are also included, the FunGramKB meaning postulates are aimed to provide just semantic knowledge.

Task 1.3. When two or more lexical units in the LDV are morphologically-related by derivation, *a priori* all of them except for one are rejected according to the following priority criterion: verb > noun > adjective. That is, if we have to choose between a noun and a verb, the latter is selected (e.g. *advice-advise, agreement-agree, appearance-appear, arrival-arrive, sale-sell*, etc). On the contrary, if the relation takes place between an adjective and a verb/noun, the adjective is rejected (e.g. *asleep-sleep, successful-success*, etc). Finally, when the three types of words are involved, then the verb is selected (e.g. *obedience-obedient-obey*, etc). In this way, redundancy is dramatically minimized, since there is no point to have two basic concepts which can

10 Some examples are *absence, piece, amount, bunch, pair, set, variety,* etc.

serve to represent the same state of affairs, as can be noted in sentences such as *arsenic is a poison* and *arsenic is poisonous*. This priority criterion is grounded on the descriptive power of concepts in COREL predications, where events are able to introduce their whole cognitive schemes in the form of thematic frames, participants are typically represented by entities, and qualities are practically restricted to the Attribute argument.

Task 1.4. FunGramKB describes meaning oppositions between qualities by locating them in cognitive spaces, where positive and negative focal concepts are determined (Periñán-Pascual & Arcas-Túnez, 2008). Here terms such as "positive" and "negative" are not applied to refer to a kind of meaning connotation, but to the presence or not of the negation operator in the meaning representation. In other words, the negative focal concept is defined as the negation of the positive one: e.g. *false* means *not true*. Evidently, if A is the opposing concept of B, then there is no need to state that B is the opposing concept of A. Any of the two focal concepts in a semantic dimension is liable to be deemed as positive. However, FunGramKB knowledge engineers follow the arbitrary criterion of taking as positive the concept to which the lexical unit with the highest frequency index is linked[11]. If there is gradation within a semantic dimension[12], all concepts involved are described around the two focal concepts, which are determined in turn by comparing the frequency indices of the lexical units linked to all those concepts belonging

---

[11] This frequency index is obtained from WordNet. However, for the sake of clarity in meaning representations, this index-based criterion can be violated when standard dictionaries typically use a less frequent concept to define the opposing one. This is the case of *alive-dead*, for example, where the second adjective is more frequent but the first one is preferred as defining word. Thus, (i) *alive* is mapped into the positive focal concept, and (ii) *dead* into the negative one.

(i) Alive: <u>still living</u> and not dead. (freq: 14)
+ALIVE_00
*(e1: +BE_01 (x1: +HUMAN_00 ^ +ANIMAL_00)Theme (x2: +ALIVE_00) Attribute)
+(e2: +LIVE_00 (x1)Theme)

(ii) Dead: no longer alive. (freq: 72)
$ALIVE_N_00
*(e1: +BE_01 (x1: +HUMAN_00 ^ +ANIMAL_00)Theme (x2: $ALIVE_N_00) Attribute)
+(e2: n +BE_01 (x1)Theme (x3: +ALIVE_00)Attribute)

[12] A quality is gradable (e.g. +EXPENSIVE_00) when, for the same instance of the entity, the quality can take varying degrees of intensity along the time. Otherwise, the quality is non-gradable (e.g. +ALIVE_00).

to the semantic dimension[13]. For the remaining concepts in the semantic dimension, the quantifying operators *m* (many/ much) and *p* (few/little) are used to describe different degrees of intensity around the focal concepts. For example, Figure 3 illustrates the dimension of size, where the positive focal concept is *big* and the negative one is *small.*

FIGURE 3
The semantic dimension of size

m+ (very big)	**+** (big)	**p+** (a little big)	**n+ n-** (neither big nor small)	**p–** (a little small)	**–** (small)	**m–** (very small)
colossal enormous gigantic great huge immense whopping	big large				little small	microscopic midget minuscule minute tiny

As can be seen, a cognitive dimension in which qualities are involved in a meaning opposition can be split into at least two (e.g. non-gradable polarity) and up to seven (e.g. gradable series) cognitively-feasible semantic zones[14]. Therefore, when two or more adjectives from the LDV belong to the same semantic dimension, the adjective which is mapped as the positive focal concept is stored as the basic concept. Other similar cases are found in semantic oppositions such as *alive-dead, male-female, right-wrong, true-false*, etc.

---

13 More particularly, the positive focal concept is selected on the basis of the highest index, and the negative one follows the same criterion but taking into account just those concepts located in the opposite side of the dimension.

14 One of the key features of semantic zones is their “cognitive feasibility”, which does not necessarily imply “lexicalization” (Periñán-Pascual & Arcas-Túnez, 2008). In other words, every semantic zone can be represented by a concept, but it is possible for a particular language to have no lexical realization for that concept. In fact, the difference between series and polarities lies in the cognitive feasibility of the central semantic zone, regardless of the possibility of lexicalization in that zone. For example, in Figure 3, not all semantic zones are lexicalized in English, but they are susceptible to be lexicalized when introducing other natural languages in the knowledge base, i.e. these semantic zones are cognitive feasible. As shown in this figure, the central semantic zone results from the negation of both focal concepts.

Regarding the cognitive clustering, the following tasks were involved:

Task 2.1. Synonyms and quasi-synonyms are gathered under the same concept (e.g. *answer-reply, association-organization, country-nation, kilo-kilogram, lay-put-set, problem-trouble, allow-let-permit, begin-start, do-make, end-finish, beautiful-nice-pretty, big-large, fast-quick-rapid,* etc)[15].

Task 2.2. In the case of verbs, the clustering also occurs with those lexical units which describe the same cognitive scenario. This is the case, for example, of *bring* and *take*, where the difference does not lie in their thematic frames (1) nor in their meaning postulates (2), since both verbs are bound to the same concept (i.e. +TAKE_01), but in their CLSs (3): whereas *bring* is an active accomplishment, *take* is a causative active accomplishment.

(1) (x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal

(2) +(e1: +MOVE_00 (x1)Agent (x2)Theme (x3)Location (x4) Origin (x5)Goal (f1: (e2: +BE_02 (x1)Theme (x4)Location)) Condition (f2: (e3: +BE_02 (x1)Theme (x5)Location))Result (f3: (e4: +HAVE_00 (x1)Theme (x2)Referent))Result)

(3) bring: do (x-Agent, [+TAKE_01 (x-Agent)]) & INGR +TAKE_01 (x-Agent, y-Theme)

take: [<CLS>] CAUSE [do (x-Agent, [+TAKE_01 (x-Agent)]) & INGR +TAKE_01 (y-Theme, z-Goal)]

In the case of verbs such as *buy* and *sell*, whose thematic frame and meaning postulate are presented in (4) and (5) respectively, the difference in their CLSs only entails a reinterpretation of the variables involved, since both verbs are even assigned to the same Aktionsart (i.e. active accomplishment), as shown in (6).

[15] FunGramKB is coarse-grained in comparison with standard lexicography. However, it is fine-grained in comparison with the axioms in other formal ontologies. In SUMO (Suggested Upper Merged Ontology), for instance, the concept RADIATINGLIGHT is related to more than 250 lexical units in English. On the contrary, this SUMO concept is split into tens of FunGramKB concepts.

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

(5) +(e1: +GIVE_00 (x1)Agent (x2)Theme (x3)Origin (x4)Goal (f1: (e2: +PAY_00 (x4)Agent (x5: +MONEY_00)Theme (x4) Origin (x1)Goal))Condition)

(6) buy: do (x-Goal, [+SELL_00 (x-Goal)]) & INGR +SELL_00 (x-Goal, y-Theme)

sell: do (x-Agent, [+SELL_00 (x-Agent)]) & INGR +SELL_00 (x-Agent, y-Goal, z-Theme)

Both *buy* and *sell* describe the same conceptual scenario, i.e. +SELL_00. The semantic difference is found just in the profiling of some given participant in the cognitive scheme: in the case of *buy*, the Goal is the only argument in the first activity, but the Agent defines this first activity in the CLS of *sell*. Other similar cases of cognitive clustering are found with *give-receive*, *remember-remind*, etc.

3.2. The hierarchization phase

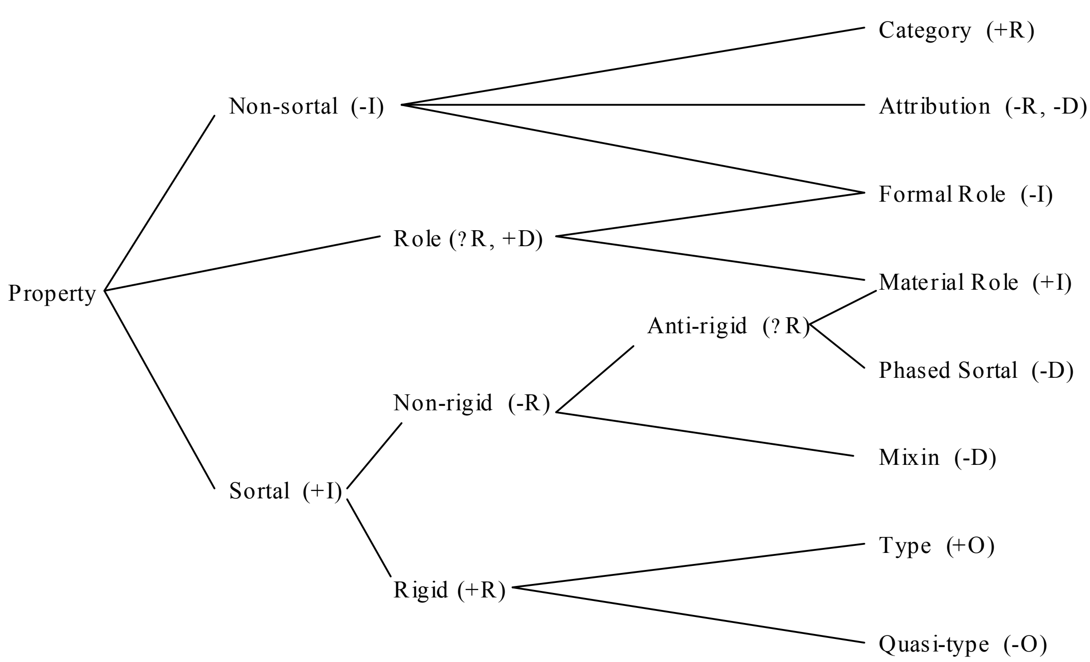
The FunGramKB Ontology comprises three subontologies, whose root metaconcepts are #ENTITY, #EVENT and #QUALITY, allowing the internal organization of full-content nouns, verbs and adjectives respectively. Therefore, the concepts from the previous stage were distributed among the FunGramKB subontologies, and then hierarchically arranged according to the IS-A relation. Along this process, it was necessary to introduce some umbrella concepts, mainly entities, in order to exploit more efficiently the inheritance mechanism in the middle level of the FunGramKB Ontology. A clear example is found in the animal taxonomy, where concepts such as +CRUSTACEAN_00, +MAMMAL_00, +MOLLUSK_00, +INVERTEBRATE_00, +REPTILE_00 and +VERTEBRATE_00 were introduced. Other examples of umbrella concepts which are not derived from the LDV are +CONTAINER_00, +FUEL_00 or +VESSEL_00.

Moreover, since all subordinate concepts must share the meaning postulate of their superordinate concepts (i.e. similarity principle), concepts were provided with their thematic frames and meaning postulates in order to check this ontological commitment (cf. Periñán-Pascual & Arcas-Túnez 2010a). In the case of entities, whose taxonomy is the deepest one, the *OntoClean* methodology (Guarino & Welty, 2000a, 2000b, 2002; Welty & Guarino, 2001) was also applied, since formal meta-properties such as rigidity, identity, unity and dependence can assist ontology engineers to use a more rigorous subsumption relation. These meta-properties can be briefly described as follows:

(i) A property is *rigid* if it is essential to the entity, i.e. it cannot change in any instance of the entity.
(ii) An *identifying* property is unique for the whole instance, distinguishing a specific instance of a certain class from other instances of that class.
(iii) A property carries *unity* if there is a common unifying relation such that all the instances of the property are wholes under that relation.
(iv) A property is *dependent* if each instance of the property implies the existence of another entity.

As shown in Figure 4, the meta-properties rigidity (R), identity (I, O) and dependence (D) are combined to shape the following semantic types (Guarino y Welty 2000a; Welty y Guarino 2001: 63):

FIGURE 4
A typology of properties based on the *OntoClean* meta-properties



Since the meta-properties impose constraints on the subsumption relationship, these formal types also impose some restrictions, such as "*types* can only be subsumed by *categories* or *quasi-types*", and "*phased sortals* must be subsumed by *types*". As in *OntologyWorks* (Guarino & Welty 2002: 65), *FunGramKB* is provided with a checker to validate the consistency of the ontology automatically, once the formal properties of the concepts have been assigned.

3.3. The remodelling phase

In this phase, some basic concepts were demoted to "subconcepts", thus affecting the structure of the ontological hierarchy. To illustrate, let us consider the concept +COVER_00, whose thematic frame and meaning postulate are as follows:

(7) (x1)Agent (x2)Theme (x3)Origin (x4)Goal

(8) +(e1: +PUT_00 (x1)Agent (x2)Theme (x3)Origin (x4)Goal (f2: +ON_00)Position(f1)Instrument (f3: (e2: +HIDE_00 (x1) Theme (x4)Referent))Purpose ^ (f4: (e3: +PROTECT_00 (x1) Theme (x4)Referent))Purpose)

Apparently +COVER_00 could be gathered with the subordinate concepts +BURY_00, +DRESS_00, +FLOOD_00 and +PAINT_00, since all of them share the meaning postulate (8). However, each one of these subordinate concepts presents, indeed, a distinctive feature in the selectional preferences of the argument Theme in their thematic frames (9-12):

(9)	+BURY_00:	(x1: +HUMAN_00 ^ +ANIMAL_00)Agent (x2: +GROUND_00 | +LEAF_00 | +STONE_00) Theme (x3)Origin (x4)Goal
(10)	+DRESS_00:	(x1: +HUMAN_00)Agent (x2: +CLOTHING_00) Theme (x3)Origin (x4: +HUMAN_00 ^ +ANIMAL_00)Goal
(11)	+FLOOD_00:	(x1)Agent (x2: m +WATER_00)Theme (x3)Origin (x4)Goal
(12)	+PAINT_00:	(x1)Agent (x2: +PAINT_00)Theme (x3)Origin (x4) Goal

From the ontological approach, the problem is that the opposition principle is not satisfied at all. In this example, the meaning postulate of these sibling concepts is exactly the same, and the *differentiae* is restricted just to some selectional preference in the thematic frame. As a result, if we replace any of the previous concepts by their superordinate when they appear in other meaning postulates, no semantic loss will actually occur, since events must always be accompanied by their thematic frames when used in any COREL scheme. To illustrate, the predication (13), which is used to define +GRAVE_00, shows this equivalence when +BURY_00 is replaced by +COVER_00:

(13) *(e2: +BURY_00 (x3: +HUMAN_00)Agent (x4: +GROUND_00) Theme (x5)Origin (x6: 1 +HUMAN_00)Goal (f1: +GRAVE_00) Location (f2: (e3: n +BE_01 (x6)Theme (x7: +ALIVE_00) Attribute))Condition) ≡ *(e2: +COVER_00 (x3: +HUMAN_00) Agent (x4: +GROUND_00)Theme (x5)Origin (x6: 1 +HUMAN_00) Goal (f1: +GRAVE_00)Location (f2: (e3: n +BE_01 (x6)Theme (x7: +ALIVE_00)Attribute))Condition)

Initially, two possible solutions were thought in order to remove this type of redundancy:

(a) The subordinate concepts could be merged with their superordinate concept (i.e. +COVER_00), whose thematic frame would then be enriched with those selectional preferences in the thematic frames (9-12), as shown in (14):

(14) (x1: +HUMAN_00 ^ +ANIMAL_00)Agent (x2: (+GROUND_00 | +LEAF_00 | +STONE_00) ^ +CLOTHING_00 ^ m +WATER_00 ^ +PAINT_00)Theme (x3)Origin (x4: +HUMAN_00 ^ +ANIMAL_00)Goal

Moreover, the lexical units linked to the subordinate concepts would be integrated into the superordinate concept. Therefore, *cover*, *spread*, *inter*, *entomb*, *bury*, *lie*, *flood*, *paint*, etc would belong to +COVER_00.

(b) The subordinate concepts, together with their corresponding lexical units, could be merged with their superordinate concept, but now the distinctive selectional preferences would be stored in the lexical entries.

Each one of these two proposals, however, poses serious problems. Concerning solution (a), the problem lies in the fact that different logical operators are not allowed for the various selectional preferences within a single participant. However, although (14) had been a well-formed COREL construct, there would have been many mismatches between the conceptual and the lexical levels. That is, since there would be no explicit correspondence between selectional preferences and lexical units, wrong assertions such as "something can be flooded with soil" or "something can be painted with plastic" could be concluded.

Concerning solution (b), three different problems would arise. Firstly, redundancy would increase in the case that a group of synonyms would share the same selectional preferences in their thematic frames. Secondly, the reasoning process would slow down, since it would be necessary to match the knowledge in the thematic frames located in the Ontology with the knowledge in the lexical templates located in the Lexicon. Thirdly, there would be no clear-cut separation between linguistic knowledge, which is stored in the lexical modules, and non-linguistic knowledge, which is stored in the conceptual modules.

In the end the solution consisted in creating subconcepts linked to superconcepts. Thus, in our example, +COVER_00 plays the role of a superconcept of the subconcepts -BURY, -DRESS, -FLOOD and -PAINT[16]. Subconcepts are not deemed to be real conceptual units but lexically-motivated refinements of the selectional preferences in the thematic frame of already-existing basic or terminal concepts, which serve as superconcepts. Subconcepts share all conceptual properties of their superconcepts, except for some of the selectional preferences in their thematic frames: even the number and role of participants in those thematic frames must be identical[17].

---

[16] The names of subconcepts are preceded by the minus sign.

[17] Appendix 1 presents more examples of subconcepts. As can be seen, subconcepts can also be assigned to terminal concepts (e.g. $SOUND_00).

3.4. The refinement phase

Indeed, this phase will span the whole life cycle of the Ontology and will mainly consist in removing those basic concepts which turn out to be very little productive. In other words, if the definitory potential of a given basic concept is dramatically undermined because it does not take part in the extended meaning postulates[18] of a large number of concepts, then that concept will be merged with its superordinate or demoted to a terminal concept, depending on the presence or lack of subordinate concepts respectively. It should be noted that only when the FunGramKB terminal level has been considerably populated, it is sensible to apply this ontological refinement[19].

4. Conclusions

Within the framework of FunGramKB, this paper discusses COHERENT, the four-phase methodology used for the basic conceptual modelling at the cognitive level. After a brief outline of the linguistic and the conceptual modules of the knowledge base, together with an overview of the three-layered ontology model, the remainder of the paper focuses on the different methodological phases, i.e. conceptualization, hierarchization, remodelling and refinement, which helped knowledge engineers perform a cognitive mapping from the defining vocabulary in *Longman Dictionary of Contemporary English* into a single inventory of about 1,300 basic concepts.

---

[18] Lexical meaning is like an iceberg-only a small amount is visible from the surface, so a word is associated to much more semantic information which is really shown in its meaning postulate (Peters and Kilgarriff, 2000). In FunGramKB, all this underlying conceptual information is revealed through a process called MicroKnowing (Periñán-Pascual and Arcas-Túnez, 2005), Microconceptual-Knowledge Spreading, which can be defined as a multi-level pre-reasoning process for the construction of the extended meaning postulate of a given concept.

[19] The threshold for conceptual productivity can be automatically determined by the information content of basic concepts quantified as negative the log likelihood, -log p(c), where p(c) is the probability of encountering the basic concept c in the FunGramKB extended meaning postulates.

5. Bibliographic References

Allen, James F., 1983: "Maintaining knowledge about temporal intervals", *Communications of the ACM* 26 (11), 832-843.

Allen, James F. & George Ferguson, 1994: "Actions and events in interval temporal logic", *Journal of Logic and Computation* 4 (5), 531-579.

Bizer, Christian *et alii*, 2009: "DBpedia: a crystallization point for the Web of Data", *Journal of Web Semantics: Science, Services and Agents on the World Wide Web* 7, 154-165.

García Carrión, Llanos, 2010: *Representación formal del conocimiento enciclopédico en FunGramKB: el concepto PLACE.* Master thesis, Universidad Nacional de Educación a Distancia.

Goddard, Cliff & Anna Wierzbicka (eds.), 1994: *Semantic and lexical universals. Theory and empirical findings*, Amsterdam: John Benjamins.

—, 2002: *Meaning and universal grammar*, Amsterdam: John Benjamins.

Goddard, Cliff (ed.), 2008: *Cross-linguistic semantics*, Amsterdam: John Benjamins.

Guarino, Nicola & Christopher Welty, 2000a: "A formal ontology of properties" in *Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management*, Berlin-New York: Springer, 97-112.

—, 2000b: "Ontological analysis of taxonomic relationships" in *Proceedings of the 19th International Conference on Conceptual Modeling*, Berlin-New York: Springer, 210-224.

—, 2002: "Evaluating ontological decisions with OntoClean", *Communications of the ACM* 45 (2), 61-65.

Jackendoff, Ray, 1990: *Semantic Structures*, Cambridge (Massachusetts): MIT Press.

Levin, Beth & Malka Rappaport Hovav, 2005: *Argument realization*, Cambridge: Cambridge University Press.

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

Mairal-Usón Ricardo, Carlos Periñán-Pascual & María Beatriz Pérez, in press: "La noción de estructura lógica conceptual" in Ricardo Mairal-Usón, Lilián Guerrero and Carlos González (eds.): *El funcionalismo en la teoría lingüística. La Gramática del Papel y la Referencia. Introducción, avances y aplicaciones*, Madrid: Akal.

Mairal-Usón, Ricardo & Francisco Ruiz de Mendoza, 2009: "Levels of description and explanation in meaning construction" in Chris Butler and Javier Martín (eds.): *Deconstructing constructions*, Amsterdam: John Benjamins, 153-198.

Mairal-Usón, Ricardo, Francisco Ruiz de Mendoza & Carlos Periñán-Pascual, in press: "Constructions within a natural language processing knowledge base" in Hans Boas and Francisco Gonzálvez-García (eds.):

Construction Grammar goes Romance, Amsterdam-Philadelphia: John Benjamins.

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

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

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

—, 2008: "A cognitive approach to qualities for NLP", *Procesamiento del Lenguaje Natural* 41, 137-144.

—, 2010a: "Ontological commitments in FunGramKB", *Procesamiento del Lenguaje Natural* 44, 27-34.

—, 2010b: "The architecture of FunGramKB" in *Proceedings of the 7th International Conference on Language Resources and Evaluation,* European Language Resources Association, 2667-2674.

Periñán-Pascual, Carlos & Ricardo Mairal-Usón, 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 (1), 11-45.

Peters, Wim & Adam Kilgarriff, 2000: "Discovering semantic regularity in lexical resources", *International Journal of Lexicography* 13 (4), 287-312.

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

Pustejovsky, James, 1995: *The Generative Lexicon*, Cambridge (Massachusetts): MIT Press.

Reinhart, Tanya, 2006: *Interface strategies: Optimal and costly computations*, Cambridge (Massachusetts): MIT Press.

Ruiz de Mendoza, Francisco & Ricardo Mairal-Usón, 2008: "Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model", *Folia Linguistica* 42 (2), 355-400.

—, 2011: "Constraints on syntactic alternation: lexical-constructional subsumption in the Lexical-Constructional Model" in Pilar Guerrero (ed.): *Morphosyntactic alternations in English. Functional and cognitive perspectives*, London: Equinox, 62-82.

Van Valin, Robert Jr., 2005: *Exploring the syntax-semantics interface*, Cambridge: Cambridge University Press.

Van Valin, Robert Jr. & Randy LaPolla, 1997: *Syntax: structure, meaning & function*, Cambridge: Cambridge University Press.

Welty, Christopher & Nicola Guarino, 2001: "Supporting ontological analysis of taxonomic relationships", *Data & Knowledge Engineering* 39 (1), 51-74.

6. Appendix 1. A sample of subconcepts

Superordinate	Subconcept	Thematic Frame
+COVER_00	-BURY	(x1: +HUMAN_00 ^ +ANIMAL_00)Agent (x2: +GROUND_00 | +LEAF_00 | +STONE_00) Theme (x3)Origin (x4)Goal
	-DRESS	(x1: +HUMAN_00)Agent (x2: +CLOTHING_00)Theme (x3)Origin (x4: +HUMAN_00 ^ +ANIMAL_00)Goal
	-FLOOD	(x1)Agent (x2: m +WATER_00)Theme (x3) Origin (x4)Goal
	-PAINT	(x1)Agent (x2: +PAINT_00)Theme (x3)Origin (x4)Goal
+CUT_00	-BEHEAD	(x1)Theme (x2: +NECK_00)Referent
	-CARVE	(x1)Theme (x2: +MEAT_00)Referent
	-CERCENAR	(x1)Theme (x2: +BODY_PART_00)Referent
	-DEGOLLAR	(x1)Theme (x2: +THROAT_00)Referent
+DECREASE_00	-ABBREVIATE	(x1)Theme (x2: +WORD_00)Referent
+INGEST_00	-DRINK	(x1: +HUMAN_00 ^ +ANIMAL_00)Agent (x2: +LIQUID_00)Theme (x3: +THROAT_00) Location (x4)Origin (x5: +STOMACH_00) Goal
+SAY_00	-ASK	(x1: +HUMAN_00)Theme (x2: +QUESTION_00)Referent (x3: +HUMAN_00) Goal
$SOUND_00	-BARK	(x1: +DOG_00)Theme (x2: +SOUND_00) Referent
	-BLEAT	(x1: +SHEEP_00)Theme (x2: +SOUND_00) Referent
	-BUZZ	(x1: +BEE_00 ^ +FLY_01)Theme (x2: +SOUND_00)Referent
	-CACKLE	(x1: +CHICKEN_00)Theme (x2: +SOUND_00)Referent
	-CHIRP	(x1: +INSECT_00)Theme (x2: +SOUND_00) Referent
	-GOBBLE	(x1: +TURKEY_00)Theme (x2: +SOUND_00) Referent
	-GRUNT	(x1: +PIG_00)Theme (x2: +SOUND_00) Referent
	-MOO	(x1: +COW_00)Theme (x2: +SOUND_00) Referent
	-NEIGH	(x1: +HORSE_00)Theme (x2: +SOUND_00) Referent
	-QUACK	(x1: +DUCK_00)Theme (x2: +SOUND_00) Referent
	-ROAR	(x1: +BEAR_00 ^ +TIGER_00 ^ +LION_00) Theme (x2: +SOUND_00)Referent
	-TRUMPET	(x1: +ELEPHANT_00)Theme (x2: +SOUND_00)Referent
	-TWITTER	(x1: +BIRD_00)Theme (x2: +SOUND_00) Referent