A knowledge-engineering approach to the cognitive categorization of lexical meaning

Carlos Periñán-Pascual
Universidad Politécnica de Valencia, Spain
jopepas3@upv.es

Abstract

A key challenge in natural language processing is to develop intelligent agents which can retrieve and manage knowledge efficiently as well as simulate human-level reasoning. Undoubtedly, the knowledge base plays a crucial role in such a cognitive architecture. The problem lies in the fact that most approaches to the computational treatment of the meaning of words are restricted to systems of binary lexical relations. The goal of this article is to describe, from the view of linguistics and cognitive science, the theoretical foundation which underlies the construction of the deep semantic representations in FunGramKB, a multipurpose lexico-conceptual knowledge base to be implemented in natural language understanding systems. Thus, the conceptual schemata of thematic frames and meaning postulates may not only provide a full-fledged formalization of lexical semantics in natural language processing but can also facilitate the comprehension of linguistic realizations in artificial intelligence.

Keywords: FunGramKB, meaning postulate, thematic frame, ontology, lexical semantics

Resumen

Uno de los grandes retos del procesamiento del lenguaje natural es el desarrollo de agentes inteligentes que nos permitan no sólo recuperar información y gestionar el conocimiento de forma más eficaz sino también simular el razonamiento humano. En este escenario, la base de conocimiento desempeña un papel crucial dentro de la arquitectura cognitiva. El problema radica en que la mayoría de los enfoques para el tratamiento computacional del significado léxico se limitan a sistemas de relaciones léxicas binarias. El objetivo de este artículo es describir, desde el prisma de la lingüística y la ciencia cognitiva, la fundamentación teórica sobre la que se construyen las representaciones semánticas en FunGramKB, una base de conocimiento léxico-conceptual multipropósito para su implementación en sistemas que requieran la comprensión del lenguaje. De esta manera, los esquemas conceptuales de los marcos temáticos y los postulados de significado no sólo proporcionan una formalización detallada de la semántica léxica en el procesamiento del lenguaje natural sino además facilitan la comprensión de las realizaciones lingüísticas en la inteligencia artificial.

Palabras clave: FunGramKB, postulado de significado, marco temático, ontología, semántica léxica

1. Introduction and overview to FunGramKB

Semantic knowledge is usually required for two main tasks in natural language processing (NLP): parsing (e.g. ambiguity resolution) and partial understanding (e.g. document classification). Performance can be actually improved if the system is provided with a robust knowledge base and a powerful inference component (Vossen, 2003). However, the main problem in the construction of natural language understanding systems is usually found in the lack of a well-developed semantic knowledge base.

With regard to the quality of semantic knowledge, the conceptual content of lexical units can be described by means of semantic features or primitives, or through associations with other lexical units in the lexicon (Velardi *et alii*, 1991). In other words, there exists a clear-cut dichotomy between deep semantics, which is based on conceptual meaning, and surface semantics, which is based on relational meaning. Strictly speaking, the latter cannot provide a real definition of the lexical unit, but it describes its usage in the language via "meaning relations" with other lexical units. WordNet is one of the best-known examples of "relational" lexical database, which provides elaborate lexical networks by means of semantic relations between *synsets* (or clusters of synonymous words). Most current NLP systems adopt a relational approach to represent lexical meanings, since it is easier to state associations among lexical units in the way of meaning relations rather than to describe formally the conceptual content of lexical units. As a result, deep semantics in NLP applications is virtually non-existent, perhaps because most applications exploit WordNet as the source of information. Although surface semantics will certainly be sufficient for systems such as automatic indexing, the construction of a robust knowledge base guarantees that the resource will be reused in most NLP tasks.

In line with the deep semantics approach, FunGramKB (Periñán-Pascual and Arcas-Túnez, 2010) came on the scene as the result of a knowledge-engineering project for natural language understanding. FunGramKB is actually a knowledge base which has been designed to be reused in various NLP tasks (e.g. information retrieval and extraction, machine translation, dialogue-based systems, etc) and with several languages (i.e. English, Spanish, German, French, Italian, Bulgarian and Catalan).

This knowledge base comprises three major knowledge levels, consisting of several independent but interrelated modules:

Lexical level:

- The Lexicon stores morphosyntactic and collocational information about lexical units.
- The Morphicon helps our system to handle cases of inflectional morphology.

Grammatical level:

- The Grammaticon stores the constructional schemata which help Role and Reference Grammar (Van Valin and LaPolla, 1997; Van Valin, 2005) to construct the semantics-to-syntax linking algorithm.

Conceptual level:

- 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, i.e. schemata in which a sequence of stereotypical actions is organised on the basis of temporal continuity.
- The Onomasticon stores information about instances of entities and events, such as Bill Gates or 9/11. This module stores two different types of schemata (i.e. snapshots and stories), since instances can be portrayed synchronically or diachronically.

In the FunGramKB architecture, every lexical or grammatical module is language-dependent; on the contrary, every conceptual module is shared by all languages, where the Ontology becomes the pivotal module for the whole architecture. Moreover, any type of conceptual knowledge, i.e. semantic, procedural or episodic, is represented in FunGramKB through the same formal language, COREL (COnceptual REpresentation Language), so that information sharing takes place more effectively (cf. Periñán-Pascual and Mairal-Usón, 2010).

We aim to demonstrate that the FunGramKB conceptualist approach to language, and more particularly the dual-theories approach to semantic representation, helps to provide a full-fledged formalization of lexical semantics in the NLP framework. The organization of this paper is as follows: section 2 serves to justify the need of deep-semantic representations in natural language understanding, section 3 describes the FunGramKB ontology model in which to anchor lexical meanings, and, finally, section 4 provides an account of the FunGramKB semantic schemata as well as dealing with the linguistic and cognitive theories underlying the construction of those representations.

2. Motivation of the research

2.1. Cognitive architecture and knowledge base

From the initial steps of the FunGramKB project (cf. Periñán-Pascual and Arcas-Túnez, 2005), our efforts have always been aimed at developing a machine-tractable model of conceptualization which could simulate human-level reasoning in a linguistic-aware application. In such a scenario, the natural language understanding system should comprise the knowledge base together with a cognitive architecture, e.g. ACT-R (Anderson, 1993; Anderson and Lebiere, 1998) or Soar (Laird *et alii*, 1986; Newell 1990), among many others. Cognitive architectures serve to determine the underlying infrastructure for the intelligent system, i.e. the cognitive mechanisms which are constant over time and across different applications. In this regard, the knowledge base should not be treated as part of the cognitive architecture, since working memory contents can change over time.

In the last two decades the biggest problem that the artificial intelligence community has been facing is that most practitioners have only been concerned with developing efficient algorithms and formal theories. In other words, most researchers have been interested in the execution process of the cognitive architecture, rather than in the knowledge base itself. However, both the cognitive architecture and the conceptual knowledge base are equally important to cognition. Unfortunately, although cognitive architectures use knowledge in the form of categories, "they often relegate them to opaque symbols, rather than representing their meaning explicitly" (Langley *et al*, 2009: 151). More particularly, the lack of a machine-tractable repository of semantic knowledge becomes the Achilles' heel of many cognitive architectures when implemented in NLP systems. To illustrate, the following section discusses the weaknesses of FrameNet, the most remarkable semantic knowledge base in the current linguistic scene.

2.2. FrameNet

The FrameNet project (Ruppenhofer *et al*, 2006), which is built upon the theory of Frame Semantics (Fillmore, 1982, 1985; Fillmore and Atkins, 1992), is aimed to construct a lexical database where word senses are linked to hand-crafted semantic frames, which become the notational devices for meaning description. In other words, the semantic frame is a schematic arrangement of the frame elements which describe the scenario underlying the meanings of semantically-related words, e.g. the *Theft* frame consists of the core frame elements GOODS, PERPETRATOR, SOURCE and VICTIM, together with the peripheral frame elements MEANS, TIME, MANNER and PLACE. In addition, the frame elements in this collection of semantic frames are used to annotate corpus-extracted sentences manually. In this way, it is possible to retrieve automatically the inventory of syntactic patterns in which lexical units are involved.

The main drawback of the Frame Semantics model lies not only in the syntax-semantics interface but also in the deceptively deep approach to knowledge representation. FrameNet is certainly "a large lexical databank which provides deep semantics" (Fillmore *ettal*, 2001), showing clear advantages over relational lexical databases such as WordNet (cf. Boas, 2005). The controversy arises when the description of meaning takes place in the conceptual realm, where it is restricted to a list of roles (e.g. frame elements) which work as binary semantic relations. In other words, between the poles of a deep approach (e.g. FunGramKB) and a surface approach (e.g. WordNet) to knowledge representation, a "shallow" approach implies that the cognitive content of a lexical unit is described by means of a simple feature-value matrix of conceptual relations (e.g. FrameNet). Consequently, surface and shallow models of natural language understanding are not sufficient for constructing efficient cognitive-based systems, since their expressive power is dramatically restricted (cf. Periñán-Pascual and Arcas-Túnez, 2007b). For instance, if you take into account the verb *forgive*, its *Forgiveness* frame, whose elements are EVALUEE, JUDGE and OFFENSE, cannot fully state the meaning of the verb: "you stop being [1] angry [2] with someone you blamed [3] him or her". That is, neither the semantic frame can represent aspectuality [1] or temporality [3], nor the frame elements can be conceptually qualified [2].

Moreover, although frame elements are deemed to be fine-grained roles, FrameNet ignores the *differentiae* of many verbs by lumping them together under the same semantic frame, resulting in coarse-grained meaning representations. For example, lexical units such as *steal*, *shoplift* and *snatch* are all linked to the *Theft* frame, so it fails to handle the Location and Manner *differentiae* in the meanings of *shoplift* and *snatch* respectively. Indeed, FrameNet researchers possibly opted for this excessive granularity in semantic roles in order to compensate for the deficiencies in this shallow model of lexical meaning.

3. FunGramKB Ontology

A key aspect in knowledge engineering is the design and construction of an ontology model under a series of well-founded guidelines, particularly when you want to reuse it in different applications. Ontology development must be supported by some theory about the elements in the domain, their inherent properties and in which way these elements are related (Gerstl, 1992). Since knowledge engineers face numerous problems in the conceptual modelling of an ontology, it is necessary to work with some underlying "ontological commitments", a term which was first introduced by Quine (1961). These guidelines should help us make decisions on (i) what to be incorporated as a conceptual unit, (ii) where to place it, (iii) how to represent its meaning and (iv) how to organize the structure of the whole ontology (Mahesh, 1996). Consequently, in the following sections we describe the elements, properties and relations in the FunGramKB Ontology.

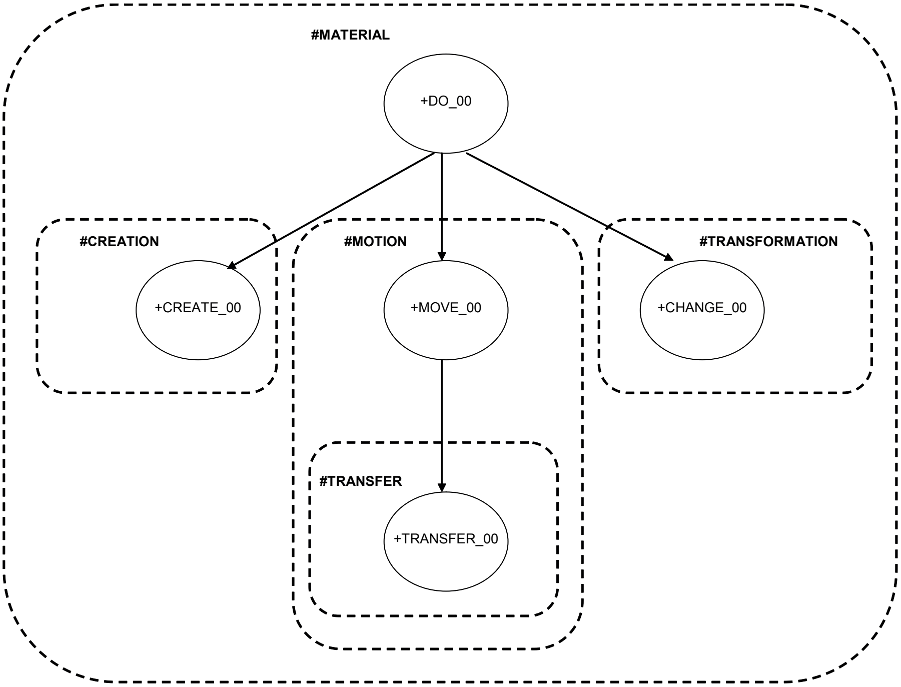
3.1. Conceptual elements

The Core Ontology distinguishes three different conceptual levels, each one of them with concepts of a different type: metaconcepts, basic concepts and terminals. The motivation of constructing such an ontology model responds to the need of a core level of knowledge (i.e. basic concepts) playing a pivotal role between those universal categories which can facilitate ontological interoperatibility (i.e. metaconcepts) and those particular concepts which can grant immediate applicability (i.e. terminals).

The main role of the metaconceptual model is to cover all universal cognitive categories, supporting the integration and exchange of information with other ontologies through a *common parlance* which contributes to standarization and uniformity (Lenci, 2000). Moreover, since metaconcepts reflect cognitive dimensions and not conceptual units, they are not provided with meaning representations. Both metaconcepts and their taxonomic hierarchization arose from the analysis of the main upper-level linguistic ontologies —such as DOLCE, Generalized Upper Model or SUMO, among many others (cf. Periñán-Pascual and Arcas-Túnez, 2007a)— resulting in 42 metaconcepts (e.g. #ABSTRACT, #MOTION, #POSSESSION or #TEMPORAL) distributed in three subontologies: #ENTITY, #EVENT and #QUALITY.

Basic concepts (e.g. +BUILD_00, +COLD_00, +FEEL_00 or +WINDOW_00) 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 selection preferences in thematic frames. The starting point for the identification of the basic concepts in the Core Ontology was the defining vocabulary in the *Longman Dictionary of Contemporary English* (Procter, 1978), though deep revision was required in order to perform the cognitive mapping into an inventory of about 1,300 basic concepts. It is important to note that we move away from a strong approach like that represented by the Natural Semantic Metalanguage approach (cf. Goddard and Wierzbicka, 2002), which identifies a finite and complete inventory of universal semantic primitives that are used to represent meaning. Instead, and following Eagles' recommendations (1998), a weaker approach is adopted in FunGramKB: whereas all semantic primitives turn out to be FunGramKB basic concepts, not all basic concepts are deemed to be universal primitives. Thus, the superordinate of a basic concept is another basic concept (e.g. +SAY_00 > +ANSWER_00) or a metaconcept (e.g. #COMMUNICATION > +SAY_00). Since metaconcepts are not actually conceived as concepts but as cognitive dimensions, those basic concepts which have a metaconceptual superordinate are eventually treated as primitives, where a distinction is made between "metaconceptual primitives" and "semantic primes": whereas metaconceptual primitives are the root basic concepts in every cognitive dimension, semantic primes are those metaconceptual primitives which cannot be further decomposed into other basic concepts. To illustrate, Figure 1 shows that +CREATE_00, +MOVE_00, +TRANSFER_00 and +CHANGE_00 are metaconceptual primitives, and +DO_00 is a semantic prime.

Figure 1. A sample of metaconceptual organization.



The FunGramKB semantic primes, which are in turn metaconceptual primitives, are the only concepts which can serve as universal primitives indeed.

Finally, terminals (e.g. $AMAZE_00, $BARGAIN_00, $SCRAMBLE_00 or $SUBLIMINAL_00) are those concepts which lack definitory potential to take part in the FunGramKB meaning postulates. Terminals are provided with the same type of properties as basic concepts, but the hierarchical structuring of the terminal level is very shallow, and in many cases practically non-existent.

3.2. Conceptual properties

Concepts in the FunGramKB Ontology are not stored as atomic symbols but are provided with semantic properties such as the thematic frame and the meaning postulate. Gruber (1995: 908) noted that "to specify a conceptualization one needs to state axioms that *do* constrain the possible interpretations for the defined terms". Since both types of schemata are given a conceptual status, they are described in more detail in section 3.1.

3.3. Conceptual relations

The fact that knowledge engineers do not understand the term "taxonomy" in the same way makes many ontologies have a chaotic structure. For example, an ontology is often identified with a "lexical taxonomy", which typically includes relations such as IS-A (e.g. animal-dog), KIND-OF (e.g. dog-Alsatian), GROUP-OF (e.g. flock-bird), PART-OF (e.g. car-wheel) or MADE-OF (e.g. lake-water). As a result of modelling the ontology with as many different types of relations as possible, taxonomy structuring tends to be confusing, so it has eventually the opposite effect than expected. Consequently, since ill-designed ontologies make their conceptual model be harder to be reused and integrated, we required a good methodology on which to ground the development of the ontology model. Thus, the OntoClean methodology (Welty and Guarino, 2001; Guarino and Welty, 2002) was applied in the FunGramKB entity taxonomy, where formal meta-properties such as rigidity, identity, unity and dependence assisted ontology engineers to use a more rigorous subsumption (or IS-A) relation.

Subsumption is the only valid taxonomic relation in the whole FunGramKB Ontology. At first sight, it can seem that the exclusive use of the IS-A relation can impoverish the ontology model. Indeed, a consequence of this restriction on the taxonomic relation is found in the modelling of the upper level into three subontologies, where metaconcepts #ENTITY, #EVENT and #QUALITY arrange nouns, verbs and adjectives respectively in cognitive dimensions. However, the fact that concepts linked to lexical units belonging to different grammatical categories are not explicitly connected in the Ontology doesn't prevent FunGramKB to relate those lexical units in the conceptual level through their meaning postulates. In fact, the Ontology establishes a high degree of connectivity among concepts by taking into account conceptual components which are shared by their meaning postulates. In order to incorporate human beings' common sense, the Ontology must identify the relations which can be established among conceptual units, and hence among lexical units. However, displaying conceptual similarities and differences through taxonomic relations themselves turns out to be more chaotic than through meaning postulates linked to conceptual units. As stated above, some ontologies present, for example, PART-OF as a taxonomic relation, in such a way that *blade* or *handle* can be explicitly linked to *knife* through the ontology hierarchy itself. In such a scenario, however, problems arise when inheritance takes place, since the properties of the superordinate are inherited by the subordinate. On the contrary, FunGramKB can retrieve any kind of relation by means of their meaning postulates, as can be seen in (1), as well as maintaining the consistency of the Ontology.

(1) *(e2: +COMPRISE_00 (x1: +KNIFE_00)Theme (x3: 1 +HANDLE_00)Referent)
*(e3: +COMPRISE_00 (x1)Theme (x4: 1 +BLADE_00)Referent)
A knife has one handle and one blade.

4. Schemata in semantic knowledge

4.1. Thematic frames and meaning postulates

Thematic frames and meaning postulates are two complementary types of conceptual scheme which carry the semantic burden of the concepts stored in the FunGramKB Ontology. Both of them fulfill a strong "unicity" criterion, since only one thematic frame and only one meaning postulate can be assigned to every concept.

A thematic frame is a conceptual construct which states the number and type of participants involved in the prototypical cognitive situation portrayed by an event or quality. These participants cannot always be instantiated linguistically, but they are always cognitively necessary, in such a way that it is impossible to understand the concept without taking them into account. To illustrate, we present the thematic frame of (2) $GRILL_00 and (3) +SMOOTH_00:

(2) (x1: +HUMAN_00)Theme (x2: +FOOD_00)Referent
(3) (x1: +SURFACE_00)Theme

Thematic frames can also include those selectional preferences typically involved in the cognitive situation. Indeed, selectional preferences are included when they are sufficiently restricted so as to exert some predictive power on the participant. For example, the thematic frame (2) describes a prototypical cognitive scenario in which "a person (Theme) cooks food (Referent)". It should not be forgotten that, although one or more subcategorization frames can be assigned to a single lexical unit, every concept is provided with just one thematic frame.

An important issue is how these thematic frames should be constructed. In this respect, metaconcepts play a key role. They are not concepts but cognitive dimensions, so they are not provided with either thematic frames or meaning postulates. However, metaconcepts are provided with an inventory of default prototypical participants from which the thematic frames of their subordinate basic/terminal concepts are constructed. Researchers usually agree on the type, role and number of prototypical participants related to any of these cognitive dimensions, but models differ in the specificity of the argument names. This consensus was shown, for example, in Periñán-Pascual and Mairal-Usón (2010: 45), who compared the type of participants involved in Halliday's processes (1985) and the roles involved in Dixon's semantic types (1991) with the prototypical participants assigned to the FunGramKB metaconcepts. In FunGramKB, thematic roles are not specific to a given metaconcept, but the cognitive dimension itself enriches the meaning of thematic roles. In other words, the participants in the thematic frame acquire different interpretations according to the metaconcept under which the given concept is placed. In this way, the inventory of thematic roles is dramatically minimized while preserving their semantic informativeness. In this regard, a key requirement for objectivity is to provide thematic roles with accurate definitions according to the location of thematic frames within the metaconceptual level. For example, although +SEE_00 and +COMPRISE_00 share the same thematic-frame pattern, the conceptual interpretation of the thematic roles is quite different, since these concepts belong to different metaconcepts, i.e. #PERCEPTION and #CONSTITUTION respectively.

(4) +SEE_00 (x1)Theme (x2)Referent
(5) +COMPRISE_00 (x1)Theme (x2)Referent

Whereas the #PERCEPTION dimension involves that Theme refers to the entity that perceives another entity (Referent), the #CONSTITUTION dimension implies that Theme refers to the entity that is made up of other entities (Referent). As can be noted, Theme becomes the key role, because its presence is obligatory in any cognitive situation described within the FunGramKB framework, where the other participants are defined in reference to that role.

On the other hand, a meaning postulate is a set of one or more logically connected predications (e_1, e_2... e_n) carrying the generic features of concepts. Consider (6) and (7) as the representation of the thematic frame and meaning postulate of +FLOAT_00 respectively:

(6) (x1)Theme (x2: +LIQUID_00)Location
(7) +(e1: +LIE_00 (x1)Theme (x2)Location (f1: (e2: n +SINK_00 (x3)Agent (x1) Theme (x2)Location (x4)Origin (x5)Goal))Result)
Something lies on a liquid without sinking.

The thematic frame of +FLOAT_00 depicts a situation in which two participants are typically involved, i.e. something (x1) stays on a liquid (x2). If the semantic burden of this concept, and consequently of words such as *float* (English), *flotar* (Spanish), *galleggiare* (Italian) etc, had been carried just by this thematic frame, then we would not have actually described the cognitive content of those lexical units. If we now consider the predicate *load*, the meaning postulate (9) provides implicatures which can't be derived from the thematic frame (8), such as "loading something involves that the loader places it in/on the loadee" and "the purpose of loading typically involves that the loadee will take the loaded thing to another place".

(8) (x1: +HUMAN_00 ^ +VEHICLE_00)Agent (x2: +CORPUSCULAR_00) Theme (x3)Origin (x4: +HUMAN_00 ^ +ANIMAL_00 ^ +VEHICLE_00)Goal
(9) +(e1: +PUT_00 (x1)Agent (x2)Theme (x3)Origin (x4)Goal (f1: +IN_00 ^ +ON_00)Position (f2: (e2: +TAKE_01 (x4)Agent (x2)Theme (x5)Location (x4) Origin (x6)Goal))Purpose)

As can be seen in the preceding examples, thematic frames are fully integrated into meaning postulates, becoming two sides of the same coin. Since the participants in the thematic frame are cognitively necessary, they must be present in the meaning postulate of the corresponding concept. As a result, every participant in the thematic frame must be referred by co-indexation with a participant in the meaning postulate. For example, the argument x1 in the meaning postulate (9) points back to the argument (x1) in the thematic frame (8), so there is no need to state again the selectional preferences of that participant, i.e. +HUMAN_00 and +VEHICLE_00. Therefore, thematic frames together with meaning postulates make up a single device for the representation of conceptual meaning in the Ontology.

An intriguing issue that divides both linguists and language engineers is the amount and the nature of semantic knowledge which should be stored for a given concept. Concerning the amount of semantic knowledge, the granularity of the metalanguage for meaning description, i.e. how fine-grained or coarse-grained the resulting representation should be, is the main issue to deal with. Indeed, it has been debated heavily, but no consensus has been reached yet, because:

> On the one hand, a concept should encode a considerable amount of information about its instances and exemplars, but on the other, it shouldn't include so much that the concept becomes unwieldy. (Laurence and Margolis, 1999: 29)

The FunGramKB meaning postulates are coarse-grained in comparison with standard lexicography. If NLP knowledge bases stored the same number of meanings that paper-based dictionaries have, it would be very difficult to differentiate formally the various senses of polysemous lexical units, not mentioning the dramatic increase of data to be stored and the combinatory explosion when disambiguating an input text lexically. On the other hand, because of the rich expressivity of COREL, the FunGramKB meaning postulates are fine-grained in comparison with the axioms in other formal ontologies.

Concerning the nature of semantic knowledge, meaning postulates are used to store unsituated prototypical knowledge. Consequently, meaning postulates consist of one or more generic predications, whose role is to describe the regularities making up our common-sense knowledge. These generic predications are inherently "intensional", since a predication describes the mental representation of a feature which determines its applicability to the category of entities to which the linguistic expression refers. Due to this intensional character, we can describe the properties of non-standard entities, e.g. those whose referents disappeared from the real world (e.g. dinosaurs) or belong to non-physical worlds (e.g. unicorns or minotaurs). In fact, intensionality causes predications to be constructed with concepts which make reference to entities in the mental world.

Generic propositions can be interpreted from two different views: the rules-and-regulations approach and the inductive approach (Carlson, 1995). According to the rules-and-regulations approach, every generic predication denotes a (physical, biological, moral…) rule. Therefore, generic predications do not directly describe the properties of entities but the properties of categories of entities. Thus, many rules denoted by generic predications refer to sociocultural conventions, i.e. stereotypes. For example, the statement "foxes are sly" implies a widely extended belief in some cultural communities as a result of the role played in Aesop's Fables and stories from ancient folklore. However, as stated by Papafragou (1996), this is a kind of knowledge which is not completely free from problems. For instance, these features actually state something false about the real world, e.g. there is no objective evidence to justify the cunning intelligence of this kind of animal. Moreover, various speakers can even assign some contradictory stereotypical features to the same concept, resulting in an ambiguous concept. For instance, Lausent (1984) showed that the "silly fox" tends to prevail in many animal stories in the oral traditions of some South-American regions.

On the contrary, according to the inductive approach, there is a semantic relation between generic predications and the properties of entities. In this respect, a predication is true only if there exists a sufficiently large number of relevant entities which satisfy the predicated property. In FunGramKB, we interpret meaning postulates from an inductive approach, because the interlingual nature of our conceptual knowledge representations is against the sociocultural nature of the rules-and-regulations approach. However, the above-described principle of the inductive approach presents two controversial issues, lying on the ambiguity of terms such as "sufficiently large number" and "relevant entities". In FunGramKB, the properties of a predication are applied to "all typical entities", that is, those entities which possess the distinctive properties. Therefore, a predication is true providing that it is true for all typical individuals. More particularly, we support the dual-theories approach when building meaning postulates, because this model turns out to be more efficient for knowledge engineering, as can be seen in the following section.

4.2. Constructing semantic representations

One of the key issues in cognitive systems is knowledge representation. The description of mental categories is usually based on a "feature theory", whose most influential approaches have been the Classical Model (Katz and Postal, 1964; Katz, 1972) and the Prototype Theory (Rosch, 1973, 1975; Rosch and Mervis, 1975). On the one hand, the Classical Model is intended to identify a minimum set of "necessary and sufficient" properties to be used as membership criteria. For example, BACHELOR can be defined by means of the features MAN and NEVER MARRIED. In this way, the concept BACHELOR has a definitional structure whose features are conditions which must be satisfied in order to consider a referent to be a bachelor. The main advantages which make this approach be so attractive are the inferential capacity (i.e. the way you reason with words) and compositionality (i.e. the way words construct the meaning of complex linguistic structures). The problem, however, is rooted in the difficulty to build this list of necessary and sufficient features, since the description of reality is not often so categorical.

On the other hand, the Prototype Theory is founded on those features which are present in most of the exemplars of a category (i.e. central tendency). In other words, concepts are structured in terms of features obtained as the result of the statistical analysis of the properties usually found in the members of a category. According to this theory, concepts are not provided with a real definition but with an open set of properties whose role is to organize "exemplariness". For example, the prototype of BIRD can be defined by the features FLIES, SINGS, LAYS EGGS etc, which are indeed not necessary and sufficient. This model easily conforms to marginal cases, i.e. members which are judged to be atypical (e.g. penguins are birds but they don't fly) or damaged (e.g. one-legged robins are still birds). Unlike the Classical Model, the main advantage of the Prototype Theory is its psychological adequacy. For example, several studies (Rosch, 1973, 1978; Smith *et alii*, 1974) demonstrated that priming in semantic categorization is closely related to the typicality of the exemplar. However, the problem lies in how to handle compositionality.

Finally, influenced by both the Classical Model and the Prototype Theory, Dual Theories of concepts (Osherson and Smith, 1981; Landau, 1982; Armstrong *et al*, 1983; Smith *et ai*, 1988) put forward a hybrid model in cognitive categorization. According to this approach, two main components are involved in conceptual representation:

a) A set of necessary (although not sufficient) features, which have a categorial function, serving to determine class membership.

b) A set of features which have an identification function, serving to determine prototypicality.

In the remainder of this section we deal with some issues derived from the application of the dual-theories model to FunGramKB.

In the FunGramKB Ontology, each predication in a meaning postulate describes a feature, which can be core or exemplary. Core features have a categorial function, whereas exemplary features have an identification function. Formally speaking, core features are represented by strict predications, and exemplary features by defeasible predications. In FunGramKB, each predication taking part in a meaning postulate is preceded by a reasoning operator in order to state if the predication is strict (+) or defeasible (*). Strict predications are law-like rules, which have no exceptions in the real world, as shown in (10):

(10) +(e1: +BE_00 (x1: $ROBIN_00)Theme (x2: +BIRD_00)Referent)
Robins are birds.

On the other hand, defeasible predications can be withdrawn (or defeated) in the light of contrary evidence. For instance, the predication (11) is defeasible, since it is possible to see a three-legged dog and that animal would still be a dog.

(11) *(e1: +HAVE_00 (x1: +DOG_00)Theme (x2: 4 +LEG_00)Referent)
Dogs have four legs.

Therefore, FunGramKB allows monotonic reasoning with strict predications and non-monotonic with defeasible predications.

Meaning postulates can be described by means of features of both types. In spite of their name, core features do not imply that their presence in meaning postulates is obligatory, but that the referent of the concept will certainly have that feature. For instance, the meaning postulate of $FAINT_00 does not have core predications, as shown in (12).

(12) *(e1: +BE_01 (x1: +HUMAN_00)Theme (x2: $FAINT_00)Attribute)
*(e2: +BE_01 (x1)Theme (x3: +WEAK_01)Attribute (f1: (e3: +BE_01 (x1) Theme (x4: +SICK_00 | +HUNGRY_00 | +TIRED_00)Attribute))Reason)
When people feel faint, they do so because they are very ill, tired, or hungry.

It should also be noted that in the case of exemplary features, which are usually far more numerous within meaning postulates, no statistical procedure is used to discover those features, but they are instead generated from lexicographical resources and knowledge engineers' introspection. Dictionaries are reliable repositories of information that several generations of expert speakers have judged to be relevant for lexical meaning. However, a robust knowledge base for NLP should also hold what Jensen (1996) describes as "reasonable knowledge", which is not usually stated in dictionaries because lexicographers rely on the users' own common sense to make definitions be understood. Therefore, dictionaries should be used as the starting point for the construction of the FunGramKB meaning postulates, but the researcher's introspection has also proved to be invaluable for re-constructing the common-sense knowledge that lexicographical resources tend to lack. Accordingly, personal introspection relies on "central tendency" as the main criterion, being the most remarkable determinant predicting typicality in taxonomic categories (Rosh and Mervis, 1975). It should be noted that people can often acquire information about the central tendency without coming across the exemplars directly, as it is the case when the information is acquired from conversations and books (Barsalou, 1991). Although being rather stereotypical, this information is useful to construct the prototypical structure, since the human source of those conversations or books would probably have experienced the exemplars directly and, consequently, could have transmitted the information about the central tendency accurately.

With the purpose of maintaining the homogeneity and consistency among the relevant features which make up meaning postulates, we compiled an inventory of descriptors which serves as a semantic guideline to help knowledge engineers broaden the information in lexical resources. This inventory is based on SIMPLE's model of Extended Qualia Structure (Lenci *et alii*, 2000; SIMPLE Specification Group, 2000), where the values of the Qualia roles proposed by Pustejovsky (1991, 1995) were extended to express fine-grained distinctions between semantic components. As noted by Keil (1979), the relevant features of a category depend hugely on the ontological domain which that category belongs to. Therefore, our descriptors were linked to the Ontology metaconcepts according to their typicality degree, in such a way that the FunGramKB Ontology editor can automatically display those descriptors that become more relevant for the concept whose meaning postulate is being built. By way of example, suppose that we are building the meaning postulate of +BRICK_00, which belongs to the metaconcept #SELF_CONNECTED_OBJECT. Knowledge engineers can use the descriptors presented in Table 1 as a device to guide their introspection process.

Table 1. Descriptors for #SELF_CONNECTED_OBJECT.

a) This entity has some PARTS.
b) The COLOUR or other VISUAL ATTRIBUTES of this entity.
c) The SIZE/LENGTH of this entity.
d) The TOUCH of this entity.
e) The TASTE of this entity.
f) The WEIGHT of this entity.
g) The SHAPE of this entity.
h) The TEMPERATURE of this entity.
i) The SMELL of this entity.
j) The VALUE of this entity.
k) The MANNER in which this entity is OBTAINED/PRODUCED.
l) This entity is made of some MATERIAL.
m) This entity is FOUND IN some places.
n) This entity is used for some PURPOSE.
o) Some ACTIONS related to this entity.
p) This entity is OBTAINED FROM a place.
q) This entity is PRODUCED BY another entity.

Thus, the question corresponding to the feature (a) would be "What typical parts do most bricks have?", where we should like to put the accent on the terms "typical" and "most" because we intend to find out an exemplary feature through this question. In case of a meaningful answer, then we try to find out whether it is a core feature, so we can check it through the question "But are these typical parts necessary?". In case of an affirmative answer, the feature is represented by means of a strict predication; otherwise, the predication becomes defeasible. However, if the first question cannot provide meaningful information, then the feature is irrelevant, so the descriptor is ruled out. In other words, not all descriptors suggested for a given concept can contribute to the construction of predications, so descriptors are used just as a guideline. Continuing with our example, whereas the predications in (13) are able to represent the features (l) and (n), the remaining features in Table 1 are ruled out.

(13) *((e2: +BE_01 (x1)Theme (x3: +CLAY_00)Attribute)(e3: past +BAKE_01 (x4) Theme (x3)Referent))
*(e4: +BUILD_00 (x5)Theme (x6: +WALL_00)Referent (f1: x1)Instrument)

A brick is made of baked clay. It is used for building walls.

We are aware that refining the list of descriptors proposed for every concept would deeply improve the efficiency of the procedure to discover semantic components in meaning postulates. More particularly, we intend to link the descriptors to the FunGramKB basic concepts, since they provide a greater wealth of information than metaconcepts. Thus, the customized inventory of descriptors could better match the conceptual meaning of the terminal subordinates.

5. Conclusions

From the perspectives of both linguistics and cognitive science, we have described the theoretical underpinnings of lexical semantics in FunGramKB, where ontological meaning helps to reveal the common-sense knowledge underlying lexical units. More particularly, meaning postulates together with thematic frames serve to represent unsituated prototypical knowledge, where their predications carry the generic features of the concepts to which the words are linked. Since the construction of lexical meaning is grounded on the dual-theories approach, a hybrid model of cognitive categorization halfway between the Classical Model and the Prototype Theory, we can succeed in developing robust NLP systems provided with deep semantics, where the role of the knowledge base should be as decisive as that of the cognitive architecture. In fact, our efforts to achieve such linguistic-aware systems resulted in ARTEMIS (Periñán-Pascual, in press), i.e. a FunGramKB-based prototype application which is aimed to simulate natural language understanding in the framework of Role and Reference Grammar. Indeed, the research results are so promising that we expect ARTEMIS to bring numerous benefits to many different NLP fields, from information retrieval to machine translation.

Acknowledgement

Financial support for this research has been provided by the DGI, Spanish Ministry of Education and Science, grant FFI2011-29798-C02-01.

References

Anderson, J.R. 1993. *Rules of the Mind*. Hillsdale: Lawrence Erlbaum.

Anderson, J.R. and Lebiere, C. 1998. *The Atomic Components of Thought*. Mahwah: Lawrence Erlbaum.

Armstrong, S., L. Gleitman, and H. Gleitman. 1983. "What some concepts might not be". *Cognition* 13: 263-308.

Barsalou, L.W. 1991. "Deriving categories to achieve goals". In G. Bower (ed). 1991. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 1-64. San Diego: Academic Press.

Boas, H.C. 2005. "From theory to practice: Frame Semantics and the design of FrameNet". In S. Langer and D. Schnorbusch (eds). 2005. *Semantik im Lexikon*, 129-160. Tübingen: Gunter Narr.

Carlson, G. 1995. "Truth-conditions of generic sentences". In G. Carlson and F.J. Pelletier (eds). 1995. *The Generic Book*, 224-237. Chicago: University of Chicago Press.

Dixon, R.M.W. 1991. *A New Approach to English Grammar on Semantic Principles.* Oxford: Clarendon Press.

EAGLES Lexicon Interest Group. 1998. *EAGLES preliminary recommendations on semantic encoding.* Technical report.

Fillmore, C.J. 1982. "Frame Semantics". In Linguistic Society of Korea (ed). 1982. *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*, 111-137. Seoul: Hanshin.

______. 1985. "Frames and the semantics of understanding". *Quaderni di Semantica* 6: 222-254.

Fillmore, C.J. and Atkins, B.T.S. 1992. "Towards a frame-based lexicon: the semantics of RISK and its neighbours". In A. Lehrer and E.F. Kittay (eds). 1992. *Frames, Fields and Contrasts.* Hillsdale: Lawrence Erlbaum.

Fillmore, C.J., Wooters, C., and Baker, C.F. 2001. "Building a large lexical databank which provides deep semantics". In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Hong Kong.

Gerstl, P. 1992. "A model for the interaction of lexical and non-lexical knowledge in the determination of word meaning". In J. Pustejovsky and S. Bergler (eds). 1992. *Lexical Semantics and Knowledge Representation*, 201-218. Berlin: Springer.

Goddard, C. and A. Wierzbicka (eds). 2002. *Meaning and Universal Grammar.* Amsterdam: John Benjamins.

Gruber, T.R. 1995. "Toward principles for the design of ontologies used for knowledge sharing". *International Journal of Human-Computer Studies* 43 (4-5): 907-928.

Guarino, N. and C. Welty. 2002. "Evaluating ontological decisions with OntoClean". *Communications of the ACM* 45 (2): 61-65.

Halliday, M.A.K. 1985. *An Introduction to Functional Grammar.* London: Arnold.

Jensen, M.R. 1996. *Knowledge representation of an encyclopedia article*. Technical report. State University of New York at Buffalo.

Katz, J. 1972. *Semantic Theory*. New York: Harper & Row.

Katz, J. and P.M. Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.

Keil, F.C. 1979. *Semantic and Conceptual Development: An Ontological Perspective*. Cambridge, Mass.: Harvard University Press.

Laird, J. E., Rosenbloom, P. S., and Newell, A. 1986. "Chunking in Soar: the anatomy of a general learning mechanism". *Machine Learning* 1: 11–46.

Landau, B. 1982. "Will the real grandmother please stand up? The psychological reality of dual meaning representations". *Journal of Psycholinguistic Research* 11 (1): 47-62.

Langley, P., Laird, J.E., and Rogers, S. 2009. “Cognitive architectures: research issues and challenges”. *Cognitive Systems Research* 10: 141-160.

Laurence, S. and E. Margolis. 1999. “Concepts and cognitive science”. In E. Margolis and S. Laurence (eds). 1999. *Concepts: Core Readings*, 3-81. Cambridge, Mass.: MIT Press.

Lausent, I. 1984. “El mundo de los animales en Pampas-La Florida”. *Bulletin de l’Institut Français d’Études Andines* 13 (1-2): 81-94.

Lenci, A. 2000. “Building an ontology for the lexicon: semantic types and word meaning”. In Workshop on Ontology-Based Interpretation of Noun Phrases, Kolding.

Lenci, A., N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ogonowski, I. Peters, W. Peters, N. Ruimy, M. Villegas, and A. Zampolli. 2000. “SIMPLE: a general framework for the development of multilingual lexicons”. *International Journal of Lexicography* 13 (4): 249-263.

Mahesh, K. 1996. *Ontology development for machine translation: ideology and methodology*. Technical report. New Mexico State University.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge, Mass: Harvard University Press.

Osherson, D. and E. Smith. 1981. “On the adequacy of Prototype Theory as a theory of concepts”. *Cognition* 9: 35-58.

Papafragou, A. 1996. “On generics”. *UCL Working Papers in Linguistics* 8: 165-98.

Periñán-Pascual, C. In press. “Towards a model of constructional meaning for natural language understanding”. In B. Nolan and E. Diedrichsen (eds). In press. *Linking Constructions into Functional Linguistics: The Role of Constructions in RRG Grammars*. Amsterdam: John Benjamins.

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

______. 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 *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence*, 279-288. Universidad de Salamanca.

______. 2010. “The architecture of FunGramKB”. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, 2667-2674. Valeta: European Language Resources Association.

Periñán-Pascual, C. and R. Mairal-Usón. 2010. "La gramática de COREL: un lenguaje de representación conceptual". *Onomázein* 21: 11-45.

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

Pustejovsky, J. 1991. "The Generative Lexicon". *Computational Linguistics* 17 (4): 409-441.

_____. 1995. *The Generative Lexicon*. Cambridge, Mass.: MIT Press.

Quine, W.O. 1961. *From a Logical Point of View, Nine Logico-Philosophical Essays*. Cambridge, Mass.: Harvard University Press.

Rosch, E. 1973. "On the internal structure of perceptual and semantic categories". In T.E. Moore (ed). 1973. *Cognitive Development and the Acquisition of Language*. New York: Academic Press.

_____. 1975. "Cognitive representations of semantic categories". *Journal of Experimental Psychology: General* 104: 192-232.

_____. 1978. "Principles of categorisation". In E. Rosch and B. Lloyd (eds). 1978. *Cognition and Categorisation*, 27-48. Hillsdale: Erlbaum.

Rosch, E. and C.B. Mervis. 1975. "Family resemblances: studies in the internal structure of categories". *Cognitive Psychology* 7: 573-605.

Ruppenhofer, J., Ellsworth, M., Petruck, M., Johnson, C.R., and Scheffczyk, J. 2006. *FrameNet II: extended theory and practice*. Technical report. University of California at Berkeley.

SIMPLE Specification Group. 2000. *Specification SIMPLE Work Package 2. Linguistic Specifications Deliverable D2.1.* Technical report.

Smith, E., D. Osherson, L. Rips, and M. Keane. 1988. "Combining prototypes: a selective modification model". *Cognitive Science* 12: 485-527.

Smith, E., E. Shoben, and L. Rips. 1974. "Structure and process in semantic memory: a featural model for semantic decisions". *Psychological Review* 81 (3): 214-241.

Van Valin, R.D. Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, R.D. Jr. and LaPolla, R. 1997. *Syntax, Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Velardi, P., M.T. Pazienza, and M. Fasolo. 1991. "How to encode semantic knowledge: a method for meaning representation and computer-aided acquisition". *Computational Linguistics* 17 (2): 153-170.

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