The structure of the lexical component in a lexico-conceptual knowledge base

Isabel Negro Alousque
Universidad Complutense de Madrid
España

Abstract

In the last years a language model, the Lexical Constructional Model (Mairal & Ruiz de Mendoza, 2008, 2009; Ruiz de Mendoza & Mairal, 2008) has been developed to account for all facets of meaning, including pragmatic and discourse phenomena. Recent research in the LCM (Periñán & Arcas, 2004, 2005, 2007a, 2010a, 2010b) has built an online lexico-conceptual base, FunGramKB, for natural language processing (NLP) applications. FunGramKB consists of two levels of information (i.e. a lexical level and a conceptual level) which in turn comprise several modules. The present contribution focuses on the lexicon, one of the two modules in the lexical level. The aim of the paper is twofold: first, to analyse the FunGramKB lexical component; second, to illustrate the semantic representation in the lexicon of the concepts stored in the ontology through a contrastive analysis of the French and English lexical entries for speech act predicates.

Keywords: Lexical Constructional Model; FunGramKB; lexicon; predicate; lexical entry; *Aktionsart*.

Afiliación: Departamento de Inglés, Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid. Madrid, España.
Correo electrónico: inegro@ccee.ucm.es.
Dirección postal: Universidad Complutense. Facultad de Ciencias Económicas y Empresariales. Departamento de Inglés. Campus de Somosaguas. 28223 Pozuelo de Alarcón (Madrid). España.

36   ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

1. Introduction[1]

The Lexical Constructional Model (hereafter LCM[2]; Mairal & Ruiz de Mendoza, 2008, 2009; Ruiz de Mendoza & Mairal, 2008, among others) is a meaning construction model elaborated in the research group Lexicom (www.lexicom.es). As mentioned in Ruiz de Mendoza & Mairal (2008: 355), the LCM approach is based on moderate functional models of language (especially Role and Reference Grammar) and cognitive linguistics (particularly Goldberg's Construction Grammar and Lakoff's Cognitive Semantics). The model provides a thorough semantic description inasmuch as it accounts for all those aspects involved in meaning construction, including traditional implicature, illocutionary force and discourse coherence.

The LCM posits the following modules of semantic description (Ruiz de Mendoza & Mairal 2008; Mairal & Periñán, 2009b):

(i)   The level 1 or argumental module is concerned with the semantic representation of predicates in terms of lexical templates.

(ii)   The implicational or level 2 layer deals with low-level inferences.

(iii)   The level 3 or illocutionary module accounts for illocutionary force.

(iv)   The level 4 or discourse layer addresses discourse aspects, especially cohesion and coherence phenomena.
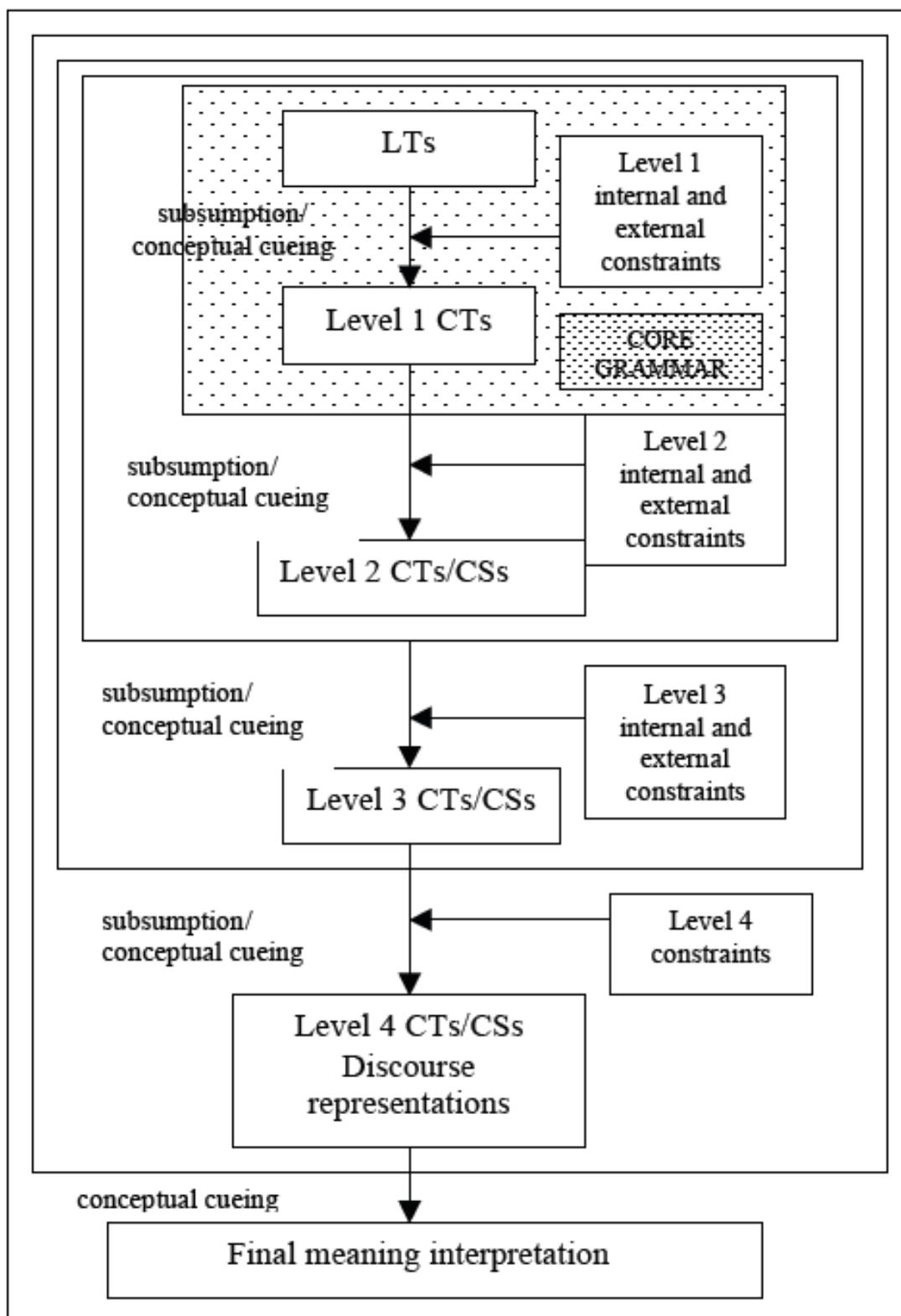
These four layers are interrelated by two cognitive processes: subsumption and cueing. This means that each level is either subsumed into a higher level module or acts as a cue for the activation of relevant conceptual structures that yields an implicit meaning derivation.

The basic organization of the LCM is given in Figure 1.

---

[2]   Abbreviations employed in this article: CLS 'Conceptual Logical Structure'; FunGramKB 'Functional Grammar Knowledge Base'; LCM 'Lexical Constructional Model'; NLP 'natural language processing'; RRG 'Role and Reference Grammar'.

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

37

FIGURE 1
**The overall architecture of the
Lexical Constructional Model**

LT = lexical template; CT = constructional template; CS = Conceptual Structure.

38

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

In the present contribution we focus on the FunGramKB lexicon, which can be considered as one of the key components upon which the knowledge base is built. Along with the Morphicon, the Lexicon, which stores morphosyntactic, pragmatic and collocational information about lexical units, is one of the linguistic modules that make up FunGramKB. It is necessary to highlight that one of the distinguishing features of FunGramKB, which makes it an incomparable model of linguistic representation, is that the lexicon provides a rich semantic description of lexical units including a fine-grained definition, grammatically relevant semantic information and pragmatic information.

Within this framework, our goal is to present the anatomy of the lexical component in FunGramKB, and to illustrate it through the analysis of the representation of speech act verbs in the French and English lexica.

The organization of the paper has the following format. Section 2 gives an overview of FunGramKB, with a special emphasis on the elements of the lexicon. Section 3 describes the organization of the FunGramKB lexicon and the features in FunGramKB lexical entries, as shown in www.fungramkb.com. This is followed by an account of the English and French lexical entries for the predicates belonging to the domain of SPEECH (or subsumed under the metaconcept #COMMUNICATION in FunGramKB). Finally, Section 4 presents a few concluding remarks.

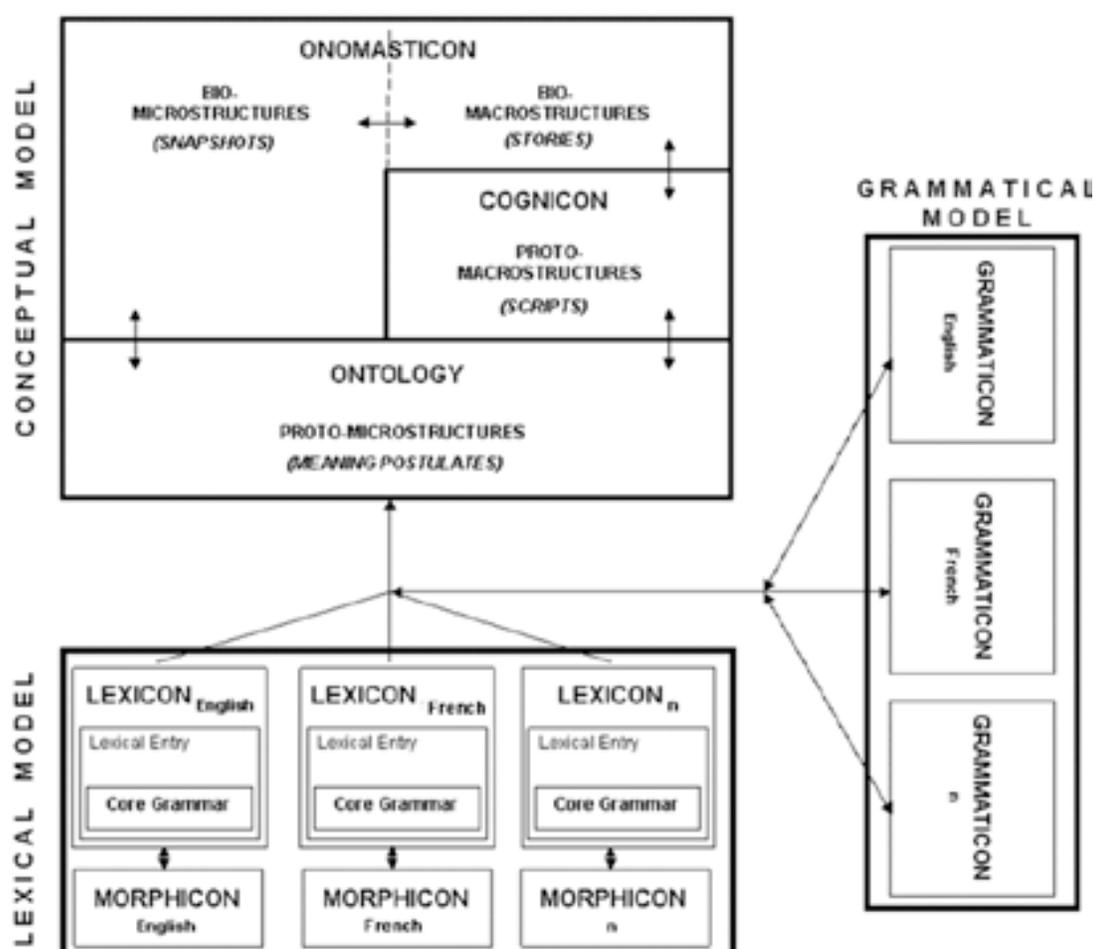## 2. The overall organization of FunGramKB

FunGramKB is a multipurpose (i.e. multifunctional and multilingual) lexico-conceptual knowledge base for NLP systems designed as part of the LCM (Periñán & Arcas, 2004, 2005, 2007a, 2007b, 2010a, 2010b; Mairal & Periñán 2009a, 2009b, 2010; Periñán & Mairal 2009). It is multifunctional and multilingual in the sense that it is meant to be reused in NLP tasks, e.g. document retrieval, information extraction and text categorization, and with many Western languages, including English, Spanish, French, German and Italian.

It is worth mentioning that the FunGramKB French version is an exceptional knowledge base in that it fills a gap in French computational lexicography, where the only language resources so far developed have been dictionaries and corpora.

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

39

Further, it integrates rich morphological, syntactic, semantic and pragmatic information of lexical units.

FunGramKB comprises two interrelated information levels: a lexical level and a conceptual level, which in turn consist of several independent but interrelated modules.

FIGURE 2
FunGramKB modules

CONCEPTUAL MODEL

ONOMASTICON

BIO-MICROSTRUCTURES (SNAPSHOTS)

BIO-MACROSTRUCTURES (STORES)

COGNICON

PROTO-MACROSTRUCTURES (SCRIPTS)

ONTOLOGY

PROTO-MICROSTRUCTURES (MEANING POSTULATES)

GRAMMATICAL MODEL

GRAMMATICON English

GRAMMATICON French

GRAMMATICON n

LEXICAL MODEL

LEXICON English — Lexical Entry — Core Grammar

LEXICON French — Lexical Entry — Core Grammar

LEXICON n — Lexical Entry — Core Grammar

MORPHICON English

MORPHICON French

MORPHICON n

The lexical level is made up of the lexicon and the morphicon, while the conceptual level consists of the ontology, the cognicon and the onomasticon.

Lexical level (i.e. linguistic knowledge)

• The lexicon stores morphosyntactic, pragmatic and collocational information about lexical units.

• The morphicon handles cases of inflectional morphology.

Conceptual level (i.e. non-linguistic knowledge)

- The ontology –the key module in FunGramKB– is presented as a hierarchical catalogue of all the concepts that a person has in mind when talking about everyday situations.
- The cognicon stores procedural knowledge by means of conceptual macrostructures.
- The onomasticon stores information about named entities and events.

A point worth mentioning is that while the lexical level is language-specific, the conceptual level is universal (Perinán & Arcas, 2010a, 2010b). Accordingly, the ontology feeds the different lexica, as pictured in Figure 2.

3. The FunGramKB Lexicon[3]

The FunGramKB lexicon presents two clear advantages:

a) In line with the LCM concern to account for all dimensions of meaning, FunGramKB lexical entries capture conceptual, semantic, syntactic and pragmatic information.

b) The amount of semantic information meets the needs of Artificial Intelligence systems.

Lexical entries in the FunGramKB lexicon are arranged alphabetically and include conceptual information. This is a major difference between FunGramKB and multilingual database such as SIMPLE and EuroWordnet, which describe the meaning of lexical units through associations with other lexical units, thus increasing redundancy through the knowledge base.

The incorporation of conceptual information into the FunGramKB lexical entries is explained by the heavy influence of the ontology on the lexical level to the extent that the lexicon is grounded on the ontology. The conceptual content of a lexical unit is expressed in terms of the concept to which the lexical unit is linked, the thematic frame, the meaning postulate and

---

[3] For a full description of the lexical module we refer to Mairal and Periñán (2009a, 2009b).

the meaning description. FunGramKB distinguishes three conceptual levels (Periñán & Arcas, 2004)[4]:

- Metaconcepts, which are preceded by symbol #, constitute the upper level in the FunGramKB ontology. FunGramKB posits forty-two metaconcepts distributed in three subontologies: #ENTITY for nouns (e.g. #COLLECTION), #EVENT for verbs (e.g. #MOTION, #POSSESSION) and #QUALITY for adjectives and some adverbs (e.g. #ABSTRACT, #TEMPORAL)
- Basic concepts, which are headed by symbol + (e.g. +BOOK_00, +DIRTY_00, +FORGET_00 etc), are used as defining units which enable the construction of meaning postulates (cf. below) for basic concepts and terminals, as well as taking part as selectional preferences in thematic frames (cf. below): e.g. +COLD_00, +MONEY_00, +MOVE_00 etc.
- Terminals, which are preceded by symbol $ (e.g. $JUNE_00, $SENILE_00, $GLEAM_00), are those concepts which have no definitory potential to take part in meaning postulates.

Entries in the lexicon and concepts in the ontology are linked by means of the feature <concept> in such a way that a) lexical entries sharing the same headword are mapped to different concepts and b) lexical entries sharing the same meaning are mapped to the same concept.

On the other hand, as advanced above, in FunGramKB basic and terminal concepts are employed in the definition of lexical units through the thematic frame and the meaning postulate. Every event and quality is assigned one thematic frame, i.e. a conceptual schema which states the number and type of participants involved in the prototypical cognitive scenario described by the event or quality (Periñán & Arcas, 2007a: 267).

The information stated in the TF is integrated into the meaning postulate. A MP comprises a group of one or more logically connected predications (e_1, e_2... e_n), which are conceptual constructs carrying the generic features of concepts (Periñán & Arcas, 2004: 39). The MP of a subordinate concept consists of a *genus* or definiens (i.e. the superordinate concept, a metaconcept or a basic concept in the FunGramKB ontology) and one or more distinctive features (or *differentiae*).

---

[4]   A complete description of the conceptual level is provided in Mairal & Periñán (2009a) and Periñán & Mairal (2009).

The meaning description of a lexical unit provides the natural language equivalent of the meaning postulate.

The following instances exemplify the thematic frame, the meaning postulate and the meaning definition of two basic concepts belonging to the metacognitive dimension #COMMUNICATION> #EVENT, and which some of the lexical units under analysis are linked to.

1. Basic concept +SING_00, to which lexical units like English *sing, yodel* or French *chanter, chantonner, fredonner* are linked.

   (1) TF = (x1: +HUMAN_00 ^ +BIRD_00)Theme (x2)Referent (x3: +HUMAN_00 ^ +ANIMAL_00)Goal

   As we see, TFs also account for those selectional preferences typically involved in the cognitive situation being described (Jiménez & Pérez, 2010): +HUMAN_00, +BIRD_00 and +ANIMAL_00 for the cognitive scenario of 'singing'. Therefore, TF (1) describes this prototypical scenario as involving three participants: (i) entity x1 (Theme), being typically a human or a bird, which in the COREL notation is expressed with the basic concepts +HUMAN_00 and +BIRD_00 connected with the exclusion logical connector "^"; (ii) entity x2 (Referent); (iii) entity x3 (Goal), typically identified with a human or an animal.

   (2) MP = +(e1: +SAY_00 (x1)Theme (x2)Referent (x3)Goal (f1: +MUSIC_00)Manner)

   The MP for (1) presents a semantic representation in which a human being or a bird (x1 = Theme) typically makes sounds with the voice (x2 = Referent) using music (Manner).

   The *genus* of the MP is the superordinate +SAY_00, which is modified by one *differentiae* (f1), which expresses the manner (+MUSIC_00) of the action.

   (3) Meaning description: to make musical sounds with the voice, usually a tune with words.

2. Basic concept +WHISPER_00, to which lexical units like English *whisper* or French *chuchoter, murmurer, susurrer* are linked.

   (4) TF = (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00 ^ +ANIMAL_00)Goal

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

43

The MP of (4) provides a semantic representation in which a human being (x1 = Theme) typically transmits a message (x2 = Referent) to a human or an animal (x3 = Goal) (Location) without making noise (manner), which employing COREL is formalized as follows:

(5) MP = +(e1: +SAY_00 (x1)Theme (x2)Referent (x3)Goal (f1: (e2: +CREATE_00 (x1)Theme (x4: +NOISE_00) Referent))Manner

(6) Meaning description: to speak softly, in a low voice.

We now explain the format of a lexical entry as represented in the FunGramKB lexical module.

The meaning representation of lexical units in the FunGramKB lexicon does not only encode those aspects of the meaning of a word that are grammatically relevant but also semantic and pragmatic properties. It is presented in terms of features or parameters. Table 1 contains the types of features being present in FunGramKB lexical entries for English and French[5].

TABLE 1
Features in FunGramKB lexical entries

	Noun	Adjective	Verb	Adverb
1- Basic				
1.1- Headword	en/fr	en/fr	en/fr	en/fr
1.2- Index	en/fr	en/fr	en/fr	en/fr
1.3- Language	en/fr	en/fr	en/fr	en/fr
2- Morphosyntax				
2.1- Graphical variant	en/fr	en/fr	en/fr	en/fr
2.2- Abbreviation	en/fr	en/fr	en/fr	en/fr
2.3- Phrase constituents: head	en/fr	en/fr	en/fr	en/fr
2.4- Phrase constituents: particle			en/fr	
2.5- Category	en/fr	en/fr	en/fr	en/fr
2.6- Number	en/fr	en/fr		
2.7- Gender	en/fr	en/fr		
2.8- Countability	en/fr			
2.9- Degree		en/fr		en/fr
2.10- Adjectival position		en/fr		
2.11- Verb paradigm and constraints			en/fr	
2.12- Pronominalization			en/fr	

---

5   The "en" and "fr" tags represent English and French languages respectively.

	Noun	Adjective	Verb	Adverb
3- LCM Core Grammar				
3.1- Aktionsart			en/fr	
3.2- Lexical template			en/fr	
3.3- Construction			en/fr	
4- Miscellaneous				
4.1- Dialect	en/fr	en/fr	en/fr	en/fr
4.2- Style	en/fr	en/fr	en/fr	en/fr
4.3- Domain	en/fr	en/fr	en/fr	en/fr
4.4- Example	en/fr	en/fr	en/fr	en/fr
4.5- Translation	en/fr	en/fr	en/fr	en/fr

Our focus will be on the core grammar module, which consists of elements of syntactically relevant semantic information. As shown in Figure 3, the LCM Core Grammar in the lexicon contains those attributes whose values allow the system to build automatically the conceptual logical structures of lexical units after the application of the CLS Constructor Algorithm[6].

Lexical entries include the *Aktionsart* of the predicate and the argument constructions in which the predicate participates. Each predicate is assigned one or more *Aktionsarten* from Vendler's (1967) catalogue of verb classes which is divided into states, activities, achievements, semelfactives, and accomplishments, together with their corresponding causatives.

Examples of each verb class and their formal representation (cf. Van Valin, 2005: 45) are provided in Table 2:

---

[6] CLS are lexical representations resulting from the interaction of TFs and MPs with information stored in lexical entries. The CLS Constructor builds the CLS automatically from the information stored in the LCM Core Grammar together with conceptual knowledge stored in the FunGramKB ontology.

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

45

TABLE 2
Inventory of RRG logical structures

Verb Class	Logical Structure	Example	Instantiation of LS
State	predicate' (x) or (x,y)	see	see' (x,y)
Activity	do' (x, [predicate' (x) or (x,y)]	run	do' (x,[run' (x)])
Achievement	INGR predicate' (x) or (x,y), or INGR do' (x, [predicate' (x) or (x,y)]	pop (burst into tears)	INGR popped' (x)
Semelfactive	SEML predicate' (x) or (x,y) SEML do' (x, [predicate' (x) or (x,y)]	glimpse, cough	SEML see' (x,y)
Accomplishment	BECOME predicate' (x) or (x,y), or BECOME do' (x, [predicate' (x) or (x,y)]	receive	BECOME have' (x,y)
Active accomplishment	do' (x, [predicate$_1$' (x, (y))] & BECOME predicate$_2$' (z,x) or (y)	drink	do' (x,[drink' (x,y)]) & BECOME consumed' (y)
Causative accomplishment	α CAUSES ß where α, ß are LS of any type	kill	[do' (x, ∅)] CAUSE [BECOME [dead' (y)]

State: *L'enfant est blessé.*

Causative state: *Ses remarques m'ennuient.*

Activity: *La voiture a glissé sur l'autoroute.*

Causative activity: *Elle a glissé la note sous la porte.*

Accomplishment: *Les draps ont séché.*

Causative accomplishment: *Le soleil a séché le linge.*

Achievement: *Le pneu a éclaté.*

Causative achievement: *On m'a crevé deux pneus sur ma voiture.*

Semelfactive: *Il toussait sans arrêt.*

Causative semelfactive: *Ne sonne pas.*

Active accomplishment: *Nous avons marché jusqu'à la gare.*

Causative active accomplishment: *J'ai promené le chien jusqu'au parc.*

FunGramKB also incorporates into its core-grammar level of description an inventory of argument structure constructions of the kind proposed by Levin (1993) and Goldberg (1995, 2006). Such incorporation is based on the assumption that the argument structure of a predicate proves insufficient to explain the occurrence of some constituents and that constructions play an active role in determining the type of syntactic configuration in which a predicate is embedded (Mairal & Ruiz de Mendoza, 2009: 157). In this light, constructions contribute arguments to yield the final semantic interpretation of a predicate.

46   ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

FIGURE 3
The LCM Core Grammar in the FunGramKB lexicon

In the next section we discuss the lexicalisation of the metaconcept #COMMUNICATION in English and French.

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

47

4. Lexical representation of the metaconcept #COMMUNICATION in the English and French FG lexica

Speech act verbs belong to the SAY-type. SAY is a type of indefinable suggested by Faber and Mairal (1999) within the lexical domain of speech that serves to define the rest of the verbs in this class. In FunGramKB speech verbs belong to the metaconcept #COMMUNICATION and share the genus of "say".

Figure 4 presents the hierarchical structure of the meta-conceptual dimension #COMMUNICATION.

The metaconcept #COMMUNICATION agglutinates twenty-two basic concepts in the FunGramKB ontology. Basic concepts, which are headed by symbol + , are used 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[7].

There are five terminal concepts (i.e. $CONGRATULATE, $FLATTER, $GOSSIP, $PROPHESY and $SWEAR) located under the basic concept +SAY_00. Terminals, which are preceded by symbol $, are those concepts which have no definitory potential to take part in meaning postulates.

The following meaning definitions provided in the conceptual module show that the semantic particulars of the object permeate the lexical encoding of SAY-type predicates:

congratulate / féliciter: to say someone that you are happy because they have achieved something or because something nice has happened to them.

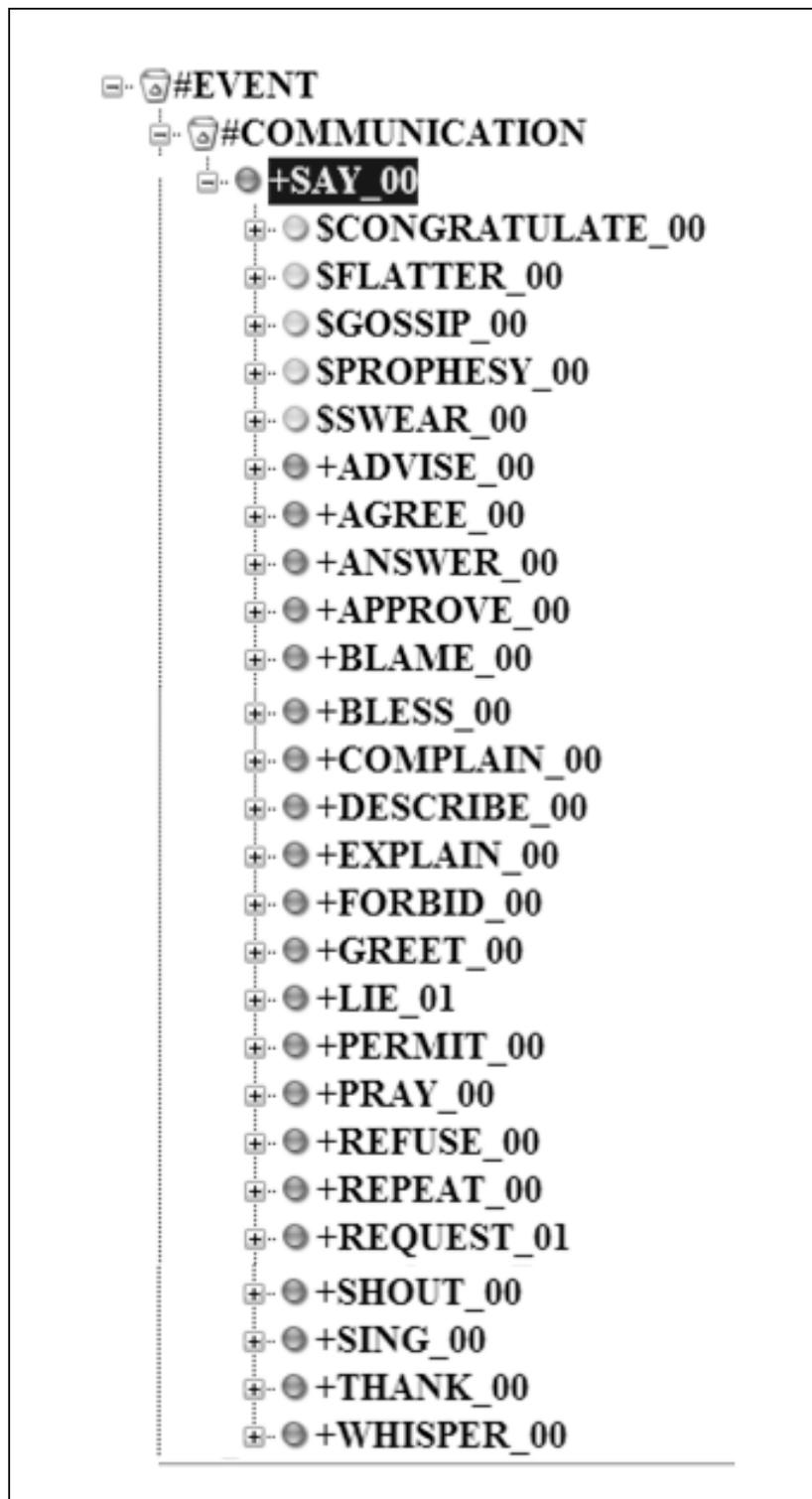*request / demander:* to express the need or desire for something.

accuse / accuser: to say that someone did something wrong or is responsible for something bad happening.

---

[7] In FunGramKB basic and terminal concepts are provided with language-independent conceptual properties such as a thematic frame and a meaning postulate. A thematic frame is a conceptual schema which states the number and type of participants involved in the prototypical cognitive scenario described by the event or quality (Periñán & Arcas, 2007a: 267). A meaning postulate comprises a group of one or more logically connected predications $(e_1, e_2... e_n)$, which are conceptual constructs carrying the generic features of concepts (Periñán & Arcas, 2004: 39).

48   ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

greet / saluer: to say hello to someone.

gossip / potiner: to say about people's behavior and private lives remarks that are unkind or untrue.

FIGURE 4
**The metaconcept #COMMUNICATION in the
FunGramKB ontology**

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

49

The comparative analysis of the lexicalization of the metaconcept #COMMUNICATION reveals significant differences between French and English. This is in accord with the idea that while the lexical level is language-specific, the conceptual level is universal (cf. above). Accordingly, the lexical module codes particular idiosyncrasies such as those discussed below.

A quantitative analysis (cf. Table 3) shows two facts: (i) A similar proportion of predicates are linked to the same concept in English and French; (ii) Terminal concepts are less richly lexicalized in both languages. This fact points to the lexicon-ontology interface.

TABLE 3
Lexicalization of the metaconcept #COMMUNICATION in the English and French FG lexica

Concept	English lexical units	French lexical units
+SAY_00	11	13
$CONGRATULATE_00	2	5
$FLATTER_00	1	2
$GOSSIP_00	1	2
$PROPHESY_00	1	1
$SWEAR_00	1	__
+ADVISE_00	5	4
+AGREE_00	11	6
+ANSWER_00	8	3
+APPROVE_00	1	1
+BLAME_00	12	3
+BLESS_00	1	1
+COMPLAIN_00	17	16
+DESCRIBE_00	5	4
+EXPLAIN_00	8	6
+FORBID_00	7	3

Concept	English lexical units	French lexical units
+GREET_00	3	1
+LIE_00	3	1
+PERMIT_00	3	3
+PRAY_00	1	1
+REFUSE_0	5	4
+REPEAT_00	8	4
+REQUEST_00	11	23
+SHOUT_00	3	15
+SING_00	2	3
+THANK_00	1	1
+WHISPER_00	1	3

The lexicalization of the metaconcept #COMMUNICATION reveals quantitative differences concerning a few concepts. Such differences can be explained by orthographic variations, pragmatic features or semantic parameters relevant to meaning description. Orthographic variations account for the higher proportion of English verbs codified by the basic concept +SAY_00 (*enquire / inquire, verbalize / verbalise*).

Other differences in the lexicalization of the metaconcept +SAY_00 are accountable for on the basis of pragmatic features. Thus, the value <formal> is encoded in a few English verbs subsumed under the basic concepts +ANSWER_00 (*rejoin, retort, riposte*), +BLAME_00 (*reproach, charge, fault, impute, tax*), +EXPLAIN_00 (*elucidate, explicate, expound*), +FORBID_00 (*debar, interdict, prohibit*), +GREET_00 (*hail, salute*), +REPEAT_00 (*iterate, recapitulate, rehearse, reiterate, restate*) and +AGREE_00 (*accede, accord, acquiesce, assent, concur, consent*). In contrast, none of the French lexical items linked to these concepts are diastratically marked.

The encoding of semantic parameters and the semantic specification of the object argument contribute to a richer lexicalization of some concepts in a language. It is the case of the

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

51

concept +REQUEST_00 in French, as the following dictionary definitions show:

> *dicter: demander à quelqu'un en secret et à l'avance de faire quelque chose* ('ask for something secretly beforehand')
>
> *sommer: demander avec autorité de faire quelque chose* ('ask authoritatively for something')
>
> *prier: demander à quelqu'un faire quelque chose avec humilité ou politesse* ('ask for something humbly or politely')
>
> *charger: demander à quelqu'un de remplir une mission/ fonction* ('ask someone to do a task')

The manner parameter is a distinguishing feature in a set of French speech predicates for which there is no translational equivalent in English. Consider the following meaning definitions:

> *chantonner* (+SING_00): *chanter à mi-voix.*
>
> *vanter* ($FLATTER_00): *parler très favorablement de quelqu'un/ quelque chose en exagérant ses mérites/ qualités publiquement.*
>
> *pester* (+COMPLAIN_00): *manifester son mécontentement/ sa colère par des paroles hargneuses et violentes.*
>
> *héler* (+SHOUT_00): *appeler quelqu'un en se servant des mains comme un porte-voix.*

We notice a lexical gap between English and French with regard to the terminal $SWEAR_00. This concept is not lexicalized in French since there is no French lexical item whose meaning is 'use rude and offensive words'. This meaning is expressed by means of the hypernym "say" and an object which is prototypically rude words (*des jurons*).

We will now focus on the most relevant features in the FunGramKB lexicon for the purposes of this study, i.e. pronominalization, *Aktionsart,* syntactic constructions and pragmatic features.

4.1. Pronominalization

Pronominalization serves as a differentiating feature in the domain under analysis. It is true that the bulk of the verbs are never cliticised in either language (, e.g. *gossip, agree, complain,*

pray-prier, request, shout-crier, beg-supplier) and several verbs are optionally cliticised in both languages, such as *ask/demander, express/exprimer, repeat/répéter, congratulate/féliciter, let/laisser.* Consider the following examples:

(1) I can express myself better in writing than verbally.

(2) Le Premier Ministre s'exprimait hier à Strasbourg.

(3) He congratulated himself for doing a good job.

(4) Le patron de la chaîne s'est félicité de ses résultats historiques.

Nonetheless, the analysis of the feature pronominalization in SAY-type verbs yields two differences:

(i) More English verbs are optionally cliticised, while the corresponding French verbs are never cliticized. It is the case of the verbs linked to the concept +DESCRIBE and others (i.e. *reproach, allow*). Let us see a few examples:

(5) The accused gunman described himself as a terrorist.

(6) Le soldat a décrit les mauvais traitements qu'il avait subis.

(7) He characterizes himself as highly moral and tolerant.

(8) Qu'est-ce qui caractérise la France?

(9) Palin depicts herself as one of the Tucson victims.

(10) Cette affaire a dépeint la relation entre les grandes fortunes et les politiques.

(11) How could I allow myself to get manipulated?

(12) Sa foi lui permet de tenir.

(ii) The value <always cliticized>, which is only codified in French, differentiates the lexical items *se plaindre* and *s'enquérir, s'éxclamer* from their English counterparts *complain* and *enquire/inquire* and *exclaim.*

4.2. *Aktionsart*

As regards the verb's *Aktionsart*, the predicates pertaining to the domain of SPEECH denote either an activity or a causative accomplishment. Activities are actions defined by the semantic properties [+dynamic] and [-telic]. The dynamic means that the

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

53

participant in the action does or causes someone to do something. The telic specifies the culmination of the action. Active accomplishments are the telic uses of activities.

Table 4 presents the distribution of the *Aktionsarten* encoded by speech predicates according to the concept the verbs are linked to:

TABLE 4
***Aktionsarten* encoded by speech predicates**

Activity	Causative accomplishment
+SAY_00 $CONGRATULATE_00 $FLATTER_00 $GOSSIP_00 $PROPHESY_00 $SWEAR_00 +AGREE_00 +ANSWER_00 +APPROVE_00 +BLAME_00 +BLESS_00 +COMPLAIN_00 +DESCRIBE_00 +EXPLAIN_00 +GREET_00 +LIE_00 +PRAY_00 +REFUSE_00 +REPEAT_00 +SHOUT_00 +SING_00 +THANK_00 +WHISPER_00	+ADVISE_00 +FORBID_00 +FORBID_00 +REQUEST_00

Both *Aktionsarten* are associated with three participants (or arguments): the Theme, the Referent and the Goal of the action. Syntactically, the goal argument of an activity is realized by a noun phrase, whereas the goal of a causative accomplishment is expressed as an infinitive.

Let us consider the following examples:

• Activities:

(13) He didn't answer all the questions.

(14) The teacher explained the vocabulary in the text.

(15) What are they complaining about?

(16) Je lui ai reproché de ne pas travailler assez.

(17) La petite criait à pleins poumons.

(18) Je ne sais pas pourquoi il a menti.

The goal is an optional argument of a set of predicates (e.g. *say/dire, shout/crier, approve/approuver, answer/répondre, repeat/répéter*). We illustrate this point through examples:

(19) The boy shouted in pain.

(20) I don't approve the project.

(21) Il répète toujours les mêmes histoires.

The predicates belonging to the conceptual dimensions +ADVISE_00, +FORBID_00, +PERMIT_00 and +REQUEST_00 designate causative accomplishments. The Goal argument is an obligatory argument, as illustrated in the following examples:

(22) He forbade his son to smoke in the house.

(23) She didn't allow me to go.

(24) Je l'ai prié de bien y réféchir.

(25) Je vous conseille d'aller voir un spécialiste.

The incorporation of the predicate's *Aktionsart* into the lexical entries is particularly relevant for French since it accounts for semantic distinctions such as those exemplified below that are not explained in traditional grammatical descriptions (e.g. Grevisse, 2007, 2009):

(26) Je lui ai dit comment j'avais réussi. (Activity).

(27) Je lui ai dit de venir tout de suite. (Causative accomplishment).

(28) Nous n'avons pas demandé le prix. (Activity).

(29) Ils nous ont demandé de les aider. (Causative accomplishment).

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

55

In traditional grammatical accounts of French *dire* and *demander* would simply be ditransitive verbs whose direct object may be realized as a noun phrase (28), a nominal clause (26) or an infinitive (27) (29), the description of a verb's syntactic behaviour being based on the notion of transitivity[8] and the syntactic function of the arguments (direct object, indirect object, etc.). In contrast, the lexical entry for *dire* in FunGramKB states the two *Aktionsarten* encoded by the predicate[9] and the semantic function of the arguments, i.e. Theme, Referent and Goal. A Theme, prototypically human, transmits a message, i.e. set of propositions (Referent) to another entity (Goal).

4.3. Syntactic constructions

The comparative analysis of English and French SAY-type predicates yields important differences concerning the argument constructions in which lexical predicates take part.

A higher proportion of English speech predicates take part in argument constructions. As a matter of fact, several constructions are only sensitive to English verbs. It is the case of the dative alternation[10], the *together-* reciprocal alternation, the reaction object alternation, the *as*-alternation and the caused motion construction. The following instances exemplify them:

(30) I recommended the camera to all my friends. / I recommended all my friends the camera.

(31) I never spoke with him. / We never spoke together.

(32) They groaned their envy.

(33) She whispered thanks and left.

(34) I strongly recommend Wilt as a good book.

(35) He shouted her into the room.

(36) The principal called me into his office.

(37) My friends talked me into going on a diet.

---

8 French distinguishes between transitive (e.g. *manger*), ditransitive (e.g. *envoyer*) and intransitive (e.g. *obéir*) verbs.

9 Interestingly enough, both *Aktionsarten* have the same number of variables.

10 Syntactic alternations are regarded as formal variants in the expression of arguments that verbs may participate in (Levin, 1993: 2).

The dative alternation is not feasible in French since the indirect object is syntactically introduced by the preposition *à*, e.g. *J'ai recommandé l'appareil à tous mes amis*; *Il a expliqué le poème aux élèves.*

Similarly, the *together*-reciprocal alternation, as illustrated by (31), is impossible in French since the item *ensemble* 'together' is used with verbs denoting physical action or motion (*Nous sommes allés ensemble / *Nous avons parlé ensemble*).

Several French and English verbs take part in other constructions such as the reciprocal alternation and the cognate object construction.

- Reciprocal alternation (intransitive)

  S/NP1 + v + OP-*from, into, to,with* (NP2) ➔ S/(NP & NP2) + v

The predicates *speak-parler* and *agree-être d'accord,* which have a prepositional object, can occur without it. This being the case, the subject must be a collective noun phrase. Subjects are typically animate and volitional (Levin, 1993: 59).

Consider the following examples:

(38) The Republicans didn't agree with the Democrats about the health reform.
(39) The Republicans and the Democrats didn't agree about the health reform.
(40) Jean a parlé à Marie des vacances à la montagne.
(41) Jean et Marie ont parlé des vacances à la montagne.

- Cognate object construction

  S/NP1 + v ➔ S/NP1 + v + O/NP2

The intransitive verbs *pray* and *sing-chanter* conveying a non-verbal expression take as their object a noun that is zero-related to the verb and so called *cognate object.*

(42) [Have you ever prayed a pray like Solomon prayed that day?
(43) Sing us a song then. Just to pass the time.
(44) Il nous a chanté une belle chanson.

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

57

4.4. Pragmatic features

As regards pragmatic information, the feature <dialect> is only pertinent to the meaning description of four English lexical units employed in the American variety (e.g. *verbalise, characterize, kvetch* and *holler*).

The feature <style> is codified by a group of English and French lexical units. On the one hand, twenty-nine English predicates share the value <formal>: *utter, verbalise, praise, flatter, counsel, concur, assent, accede, accord, acquiesce, consent, rejoin, retort, riposte, reproach, fault, tax, impute, explicate, expound, iterate, recapitulate, reiterate, elucidate, debar, permit, entreat, solicit, request.*

In contrast, in the French language this value is only relevant to the meaning description of one lexical unit (i.e. *complimenter*).

On the other hand, a considerable number of lexical units linked to the concept +COMPLAIN_00 share the value <informal>. The set of predicates includes English *grizzle, holler, moan, bitch, beef, belly-ache, kvetch* and *grouse,* and French *bougonner, clabauder, geindre, maronner, ronchonner, rouscailler* and *rouspéter*. There is also a slang predicate, *gueuler,* linked to the concept +SHOUT_00.

5. Conclusion

This paper has offered an overview of FunGramKB, a knowledge base designed as a part of the LCM which integrates rich semantic and syntactic information and will eventually serve for the development of NLP applications. The discussion of the anatomy of the FunGramKB lexicon through the lexical entries for English and French speech act predicates shows how fully-fledged and fine-grained the semantic representation of lexical units in FunGramKB is since lexical entries specify the morphosyntactic, pragmatic and collocational information of predicates. We sustain that the description of the various kinds of data that form part of the FunGramKB lexical entries shows that the lexical module is a major achievement of the LCM. On the one hand, the lexical entries have been enriched with relevant conceptual, semantic, syntactic and pragmatic

information. On the other hand, lexical representation has been provided with a formal metalanguage.

6. Bibliographic references

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

GOLDBERG, Adele, 1995: *Constructions: A Construction Grammar Approach to Argument Structure*, Chicago: Chicago University Press.

—, 2006: *Constructions at Work: The Nature of Generalization in Language*, Oxford: Oxford University Press.

GREVISSE, Maurice, 2007: *Le bon usage: Grammaire française*, Paris: Duculot.

—, 2009: *Le PetitGrevisse: Grammaire française*, Paris: Duculot.

JIMÉNEZ, Rocío & Beatriz PÉREZ, 2010: "An account of selection restrictions within a conceptual framework: Its relevance for Role and Reference Grammar" in *Proceedings of the 2009 International Conference on Role and Reference Grammar* [available from http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/ ProceedingsofRRG2009_02.pdf], 131-146.

LEVIN, Beth, 1993: *English Verb Classes and Alternations*, Chicago: University of Chicago Press.

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

—, 2009b: "Role and Reference Grammar and ontological engineering" in *Volumen Homenaje a Enrique Alcaraz*, Alicante: Universidad de Alicante,

—, 2010: "Teoría lingüística y representación del conocimiento: una discusión preliminar" in Dolores GARCÍA PADRÓN & María del Carmen FUMERO PÉREZ (eds.): *Tendencias en lingüística general y aplicada*, Berlin: Peter Lang, 155-168.

MAIRAL, Ricardo & Francisco José RUIZ DE MENDOZA, 2008: "New challenges for lexical representation within the Lexical-Constructional Model", *Revista Canaria de Estudios Ingleses* 57, 137-158.

—, 2009: "Levels of description and explanation in meaning construction" in Christopher BUTLER & Javier MARTÍN ARISTA (eds.): *Deconstructing Constructions*, Amsterdam/Philadelphia: John Benjamins, 153-198.

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

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

ONOMÁZEIN 24 (2011/2): 35-59
Isabel Negro Alousque:
The structure of the lexical component in a lexico-conceptual knowledge base

59

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

—, 2007b: "Deep semantics in an NLP knowledge base" in Daniel Borrajo, Luis Castillo & Juan Manuel Corchado (eds.): *12ᵗʰ Conference of the Spanish Association for Artificial Intelligence*, Salamanca: Universidad de Salamanca, 279-288.

—, 2010a: "The Architecture of FunGramKB" in *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Malta, 2667-2674.

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

Periñán, Carlos & Ricardo Mairal, 2009: "Bringing Role and Reference Grammar to Natural Language Understanding", *Procesamiento del Lenguaje Natural* 43, 265-273.

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

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

Vendler, Zeno, 1967: *Linguistics in Philosophy*, Ithaca, NY: Cornwell University Press.