

ONTOLOGICAL SEMANTICS IN THE LEXICAL CONSTRUCTIONAL MODEL¹

MARIA BEATRIZ PÉREZ CABELLO DE ALBA
UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (UNED)

Abstract. In the last two years, the Lexical Constructional Model (hereafter LCM) has gradually evolved from a lexicalist towards a conceptualist model. In this paper, we discuss some of the methodological consequences of this conceptual shift. In so doing, we concentrate on the following two issues: (i) the notion of *conceptual logical structures* (CLS) as an alternative form to lexical templates and (ii) the nature of the linguistic-conceptual interface. Some other more relevant notions are discussed peripherally: i.e. the anatomy of what has been termed a *constructicon*, which paves the way towards a computational treatment of constructions (cf. Bod 2009, Mairal, Ruiz de Mendoza and Perrián 2011). Finally, we discuss some of the most relevant applications that have emerged.

Keywords: *Lexical Constructional Model, Conceptual Logical Structures, Linguistic-Conceptual Interface, Selection Restrictions.*

1. Introduction

Since 2006 a group of researchers working in different paradigms – functional (Role and Reference Grammar) and cognitive (Construction Grammar and Metaphor and Metonymy Theory) joined forces with the aim of developing a comprehensive model of meaning construction. This research activity has resulted in the Lexical Constructional Model (LCM) which, unlike some other theories, provide the analytical tools to account for those aspects of meaning construction that go beyond grammar, i.e. implicational, illocutionary and even discourse features (cf. Ruiz de Mendoza and Mairal 2008, 2010, Mairal and Ruiz de Mendoza 2009).² As the outcome of this model is a fully-fledged semantic representation, this model could be a good candidate for the development of natural language processing applications, especially in the domain of crosslinguistic information retrieval, or what has been called the semantic web.

In close connection with this, a further step was taken and this linguistic apparatus was thought to be part of FunGramKB, (Perrián and Arcas 2004, 2005, 2007a,b, 2010a,b, Mairal and Perrián 2009, 2010, Perrián and Mairal 2009, 2010a,b). Figure 1 shows the new architecture of the knowledge base, where the LCM is an integral part. This knowledge base includes two levels of information: a linguistic level and a conceptual level. Each one of these two levels consists of several independent but interrelated modules. Hence, the lexicon is connected up to the ontology and in fact it retrieves information from it, and the same applies to the grammicon. Technical details aside, there is an important defining feature that cannot be underestimated: the prominent conceptual grounding of the LCM, which is no longer a lexicalist-oriented but a conceptualist model. This shift has brought about an important

¹ Financial support for this research has been provided by the DGI, Spanish Ministry of Education and Science, grant FFI2008-05035-C02-01/FILO and ACTUALing.

² For a complete account of the LCM, we refer the reader to the webpage: www.lexicom.es and the publications therein. Butler (2009) provides a critical assessment of the model.

REVISTA ELECTRÓNICA DE LINGÜÍSTICA APLICADA (ISSN 1885-9089)
2011, Número 10, páginas 187-202

Recibido: 31/05/2011

Aceptación comunicada: 22/07/2011

number of methodological corollaries that have deeply marked the internal structure of both the lexicon and the grammaticon.³

Within this framework, the organization of this paper goes as follows: Section 2 provides an overview of the most relevant methodological corollaries this conceptual shift has introduced. From this conceptual panorama we move on to analyse the internal structure of the lexicon component. In this regard, Section 3 focuses on the format of a conceptual logical structure, an alternative system of lexical representation that is more powerful than standard logical structures as propounded in Role and Reference Grammar (RRG) (cf. Van Valin 2005; Mairal, Perrián and Pérez, forthcoming). Section 4 deals with a pervasive issue for lexical and lexicographic theories, i.e. the notion of selection restrictions. Finally, Section 5 shows some of the most relevant applications of this knowledge base in the industry.

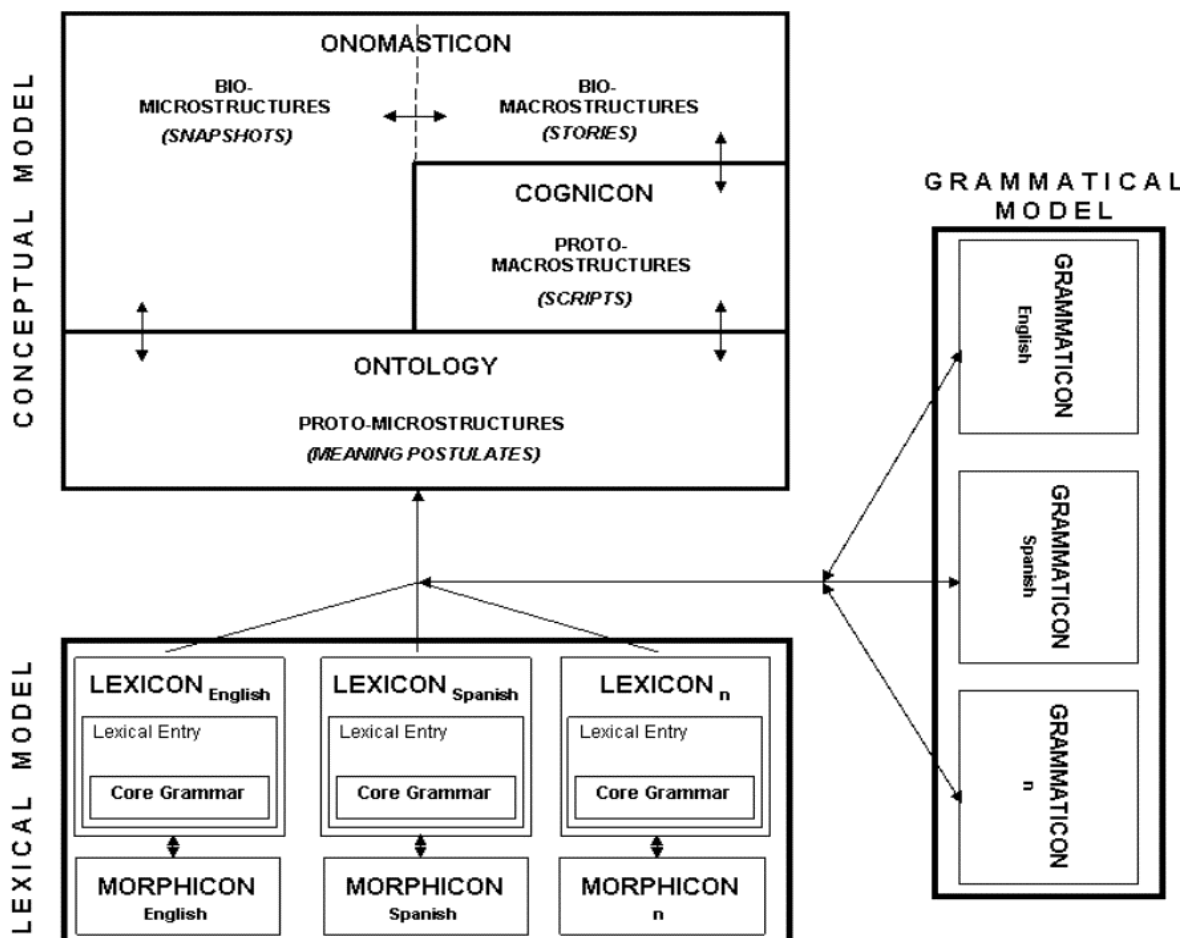


Figure 1: FunGramKB modules (at www.fungramkb.com).

2. Conceptually based methodological underpinnings

As discussed above, the LCM is now seen as part of a lexical conceptual multilingual knowledge base, whereby the lexicon and the grammatical component are connected up to a

³ This linkage is expressed in terms of what has been termed *the linguistic-conceptual interface*, a theoretical aspect that spells out the exact way the ontology really affects the internal structure of a lexical entry, or the way the ontology has a deep impact in the semantic interpretation of grammatical constructions. This set of interfaces along the path ontology-lexicon-ontology-grammaticon-lexicon is essential for measuring out the analytical capacity of the knowledge base in terms of text analysis. We refer the reader to Mairal, Ruiz de Mendoza and Perrián (2011) for a preliminary discussion of this interface.

cognitive level. This cognitive level consists of an ontology, an onomasticon and a cognicon, each corresponding to the three types of knowledge distinguished in cognitive psychology, namely, semantic, episodic and procedural knowledge (cf. Mairal and Perrián 2010, Perrián and Arcas 2007b, 2010a). What is really noteworthy at this stage is that the linguistic level is connected up to the cognitive level, and, what is more, the cognitive level constitutes the pivot of the whole machinery: for example, the ontology feeds the lexicon, shapes lexical entries and marks the explanatory scope of the lexicon. Consequently, as will be shown in Sections 3 and 4 the power of the ontology is such that the explanatory scope of lexical entries (what has been termed *conceptual logical structures*) is far more reaching than standard lexical representations since even encyclopaedic knowledge can be accessed through the knowledge base and the reasoning tools that are part of it. Prior to explaining the real consequences of an ontological approach, let us highlight some methodological corollaries which are essential to have a clear picture of the real aim of this paper.

2.1. A strong linguistic commitment

Far from following statistical and stochastic approaches like most knowledge bases in natural language processing, FunGramKB takes a strong linguistic commitment and contextualizes the whole machinery within the context of a robust meaning construction model like the LCM. Recall that the LCM, unlike some other linguistic models, provides a comprehensive account of meaning construction, including those aspects that go beyond core grammar, i.e. implicational, illocutionary and even discourse (cf. Ruiz de Mendoza and Mairal 2008; Mairal and Ruiz de Mendoza 2009). As noted in the LCM literature, four levels of descriptions are posited:

- (i) Level 1 produces core grammar characterizations.
- (ii) Level 2 accounts for heavily conventionalized situation-based low-level meaning implications.
- (iii) Level 3 is concerned with conventionalized illocutionary meaning (situation-based high-level implications).
- (iv) Level 4 deals with very schematic discourse structures.⁴

It is claimed that meaning construction is regulated by two cognitive operations: *subsumption* and *conceptual cueing*. Hence, a representation at an $n_{\text{level-1}}$ either *cues* for the activation at an $n_{\text{level-2}}$ or is constructionally subsumed into it. In an attempt at further constraining the system a bit more, these two cognitive operations are in turn regulated by a number of internal and external constraints so that meaning construction is not only *described* but also *explained*. Note that italics are used to emphasize the fact that, rather than theories that describe data, what is needed is a model that offers explanatory principles that are constrained, motivated, functionally and cognitively grounded, and, in essence, provide a real picture of the rich process involved in meaning construction. Then it comes as no surprise that the LCM can play a very active part in knowledge representation applications whose central core is precisely meaning description.⁵

⁴ So far, most of the work on the LCM has focused on Level-1, the lexical and argument structure level (i.e. Baicchi 2010; Peña 2009; Ruiz de Mendoza and Luzondo, fc.), while work at the rest of levels is the core of a research project that is currently being undertaken. There are a few preliminary results that are worth looking at since these constitute a real challenge for natural language processing given that what is aimed for is to make explicit whatever is implicit (i.e. Del Campo 2011; Ruiz de Mendoza and González 2011).

⁵ What is really essential in this section is to draw the reader's attention to the fact that comprehensive linguistic models should form part of natural language processing applications, which, for whatever the reason, has not been the case. FunGramKB is an exception.

2.2. division of labour between the lexicon and the ontology.

As shown in Figure 1, the cognitive and the linguistic levels are clearly differentiated. This division has strong theoretical implications. It is claimed that the ontology, as one of the modules in the cognitive level, has a universal status, that is, whatever is encoded is language-independent: there is just one single ontology that is common to different languages. This contrasts with the language-specific nature of the linguistic level: there is one different lexicon for each of the languages that form part of the knowledge base: Spanish, German, French, English, Bulgarian and Italian. As shown in Figure 2, each conceptual unit is associated to different words in each language, a feature which is encoded in the different lexica.

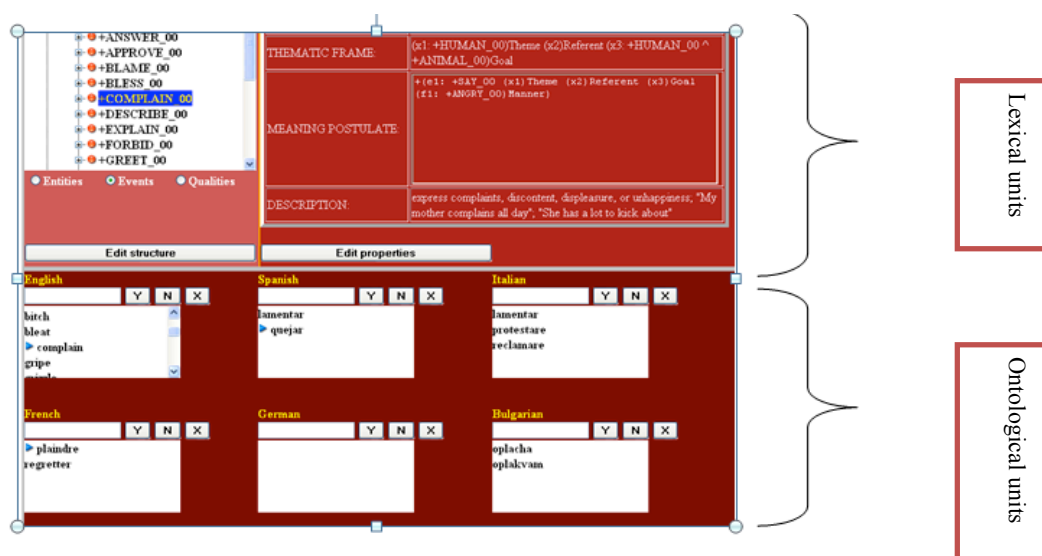


Figure 2: Ontological properties and different lexica.

In close connection with this statement, it follows the question of how to encode cultural distinctiveness: how should one proceed if there is not a straightforward relation between a concept and the different lexicalization patterns in each language? What about those languages that have a particular word to express a relevant conceptual distinction, while this distinction is not pertinent in the rest of the languages involved? Can we go on so far as to affirm that just the same ontology will be valid for the European, Oriental and Arabic linguistic communities? These are some of the central issues that are dealt with in Mairal and Perrián (fc).

The same can be said of the grammaticon and the language-specific nature of the constructions. One of the methodological corollaries is that cross-lingual differences in syntactic constructions do not necessarily involve conceptual differences (cf. Jackendoff 1990). For example, consider a high-taking complement predicate like *remember*, where the different syntactic constructions do not necessarily entail a change in the denotative meaning:

- (1)
 - a. I remember the person who was sitting next to me.
 - b. I remember to check my e-mails every morning.
 - c. I remember that Peter's theory was not precise enough.
 - d. I remember giving him the key.
 - e. I remember her reading a book under a tree.

2.3. Conceptual units are provided with two types of properties

FunGramKB is a three-layered conceptual model whereby each conceptual unit is provided with two types of properties: a thematic frame and a meaning postulate.⁶ These two properties provide definitions for each conceptual unit using a machine readable language, COREL (*Conceptual Representation Language*).⁷ This is in fact a distinctive feature if compared to the relational approaches adopted in some other models, i.e. EuroWordNet. What is really noteworthy at this stage is that each basic and terminal concept is provided with a meaning definition in the format of a thematic frame and a meaning postulate. Suffice the description of the conceptual unit +WEAVE_00 as an illustration of these two notions:

Concept	+WEAVE_00
Superordinate	+CREATE_00
Thematic Frame	(x1: +HUMAN_00)Theme (x2: +CLOTH_00 ^ +BASKET_00)Referent
Meaning Postulate	+(e1: +CREATE_00 (x1)Theme (x2)Referent (f1: +HAND_00 +MACHINE_00)Instrument (f2: +THREAD_00 ^ +STRING_00)Means (f3: +HUMAN_00)Beneficiary)

Table 1: The conceptual properties of +WEAVE_00.

In our knowledge base we find that +WEAVE_00 is a hyponym of the superordinate conceptual unit, +CREATE_00, from which the former inherits all its properties.⁸ +WEAVE_00 has two features: a thematic frame, “a conceptual construct which states the number and type of participants involved in the prototypical cognitive situation portrayed by the event” (Mairal and Perrián 2010), which indicates that this concept belongs in the metaconceptual dimension #CREATION, which designates two participants: an entity (Theme) creates another entity (Referent). Note that a thematic frame can also include selectional preferences that further define each of the participants involved in the cognitive situation, i.e. there is a human that typically weaves an object that is made of cloth or a basket.⁹ Moreover, each conceptual unit is provided with a meaning postulate, “a set of one or more logically connected predications ($e_1, e_2... e_n$), i.e. cognitive constructs carrying the generic features of concepts” (Mairal and Perrián 2010). If we look at the meaning postulate in Table I, we gather that an entity, typically a human, creates something, typically an object made of cloth or a basket, using an instrument, typically a hand or a machine, by means of using thread or strings, and all this is done at the benefit of someone.

Details aside, we would like to draw your attention to the fact that by building lexical representations in terms of conceptual units we will be able to expand the explanatory power of our representations since the knowledge base will provide us with information that trespasses the frontiers of linguistic knowledge. In a sense, an ontological scenario offers a nice setting to account for accessing world knowledge or encyclopaedic knowledge. This is in fact the core of the following section.

⁶ These three types of conceptual units are: (i) metaconcepts; (ii) basic concepts and (iii) terminal concepts (cf. Perrián and Arcas 2010a).

⁷ see Perrián and Mairal (2010) for a complete description of the syntax and semantics of this metalanguage

⁸ Note that non-monotonic multiple inheritance is one of the ontological commitments of FunGramKB (cf. Perrián and Arcas 2010b).

⁹ We will return to the notion of selectional preferences below since an ontological approach like the one adopted here is really influential in delineating the contrasts between a selectional preference and a collocation.

3. Conceptual Logical Structures

As a natural consequence of this ontological orientation, FunGramKB has formulated the notion of *Conceptual Logical Structures (CLS)* as a system for lexical representation. As discussed in Mairal, Periñán and Pérez (forthcoming), CLSs arise as an alternative to standard systems of semantic decomposition approaches, i.e. standard *logical structures* in RRG (table 2) (cf. Van Valin (2005:45):), the lexical templates as part of the Theory of the Generative Lexicon (TGL) (figure 3) (cf. Pustejovsky 1995:82):¹⁰

VERB CLASS	LOGICAL STRUCTURE	INSTANTIATION OF LS
State	predicate' (x) or (x,y)	see' (x,y)
Activity	do' (x, [predicate' (x) or (x,y)])	do' (x,[run' (x)])
Achievement	INGR predicate' (x) or (x,y), or INGR do' (x, [predicate' (x) or (x,y)])	INGR popped' (x)
Semelfactive	SEML predicate' (x) or (x,y) SEML do' (x, [predicate' (x) or (x,y)])	SEML see' (x,y)
Accomplishment	BECOME predicate' (x) or (x,y), or BECOME do' (x, [predicate' (x) or (x,y)])	BECOME have' (x,y)
Active accomplishment	do' (x, [predicate ₁ ' (x, (y))] & BECOME predicate ₂ ' (z,x) or (y)	do' (x,[drink' (x,y)]) & BECOME consumed' (y)
Causative accomplishment	α CAUSES β where α, β are LS of any type	[do' (x, ∅)] CAUSE [BECOME [dead' (y)]]

Table 2: Logical structures in RRG (adapted from Van Valin 2005:45)

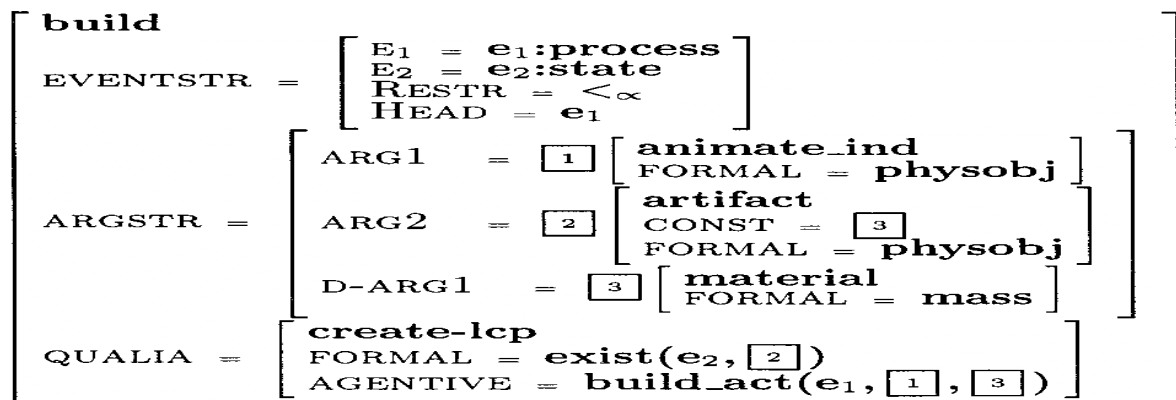


Figure 3: Example taken from Pustejovsky (1995: 82).

Technical details aside, the two representations coincide in a number of important features. First, both include an event structure as part of a predicate's lexical entry and therefore predicates are classified in terms of their *aktionsart* or its internal aspectual properties. Second, there is an argument structure that specifies the number of arguments involved in the predication. Furthermore, each lexical entry is provided with a prime or in the case of the TGL there is a mark (*lcp*) towards the type of lexical class. Both are said to have an ontological grounding. Additionally, lexical entries in the TGL include a more fine grained semantic description of each of the arguments following the format of quale (cf. Pustejovsky

¹⁰ The same can be said of the *event structure templates* in Rappaport and Levin (1998) or the constructional schemata in Goldberg (1995).

1995). However, these two systems meet a number of challenges, some of which are even recognized by their proponents:

a) Primes should have further semantic decomposition as shown within the RRG literature: “Because there is as yet no adequate decompositional representation for the primitive state and activity predicates which are the building blocks of the system (...)” (Van Valin y LaPolla 1997:114)

In connection with this, Van Valin and Wilkins (1993) and Van Valin and LaPolla (1997) developed lexical templates for the predicate *remember* and for *speech act verbs*, where each predicate is enriched with a number of internal variables that account for those parameters (both semantic and pragmatic) that go beyond syntax.¹¹ The TGL, however, is more ambitious in this respect and it includes those semantic parameters that are formalizable in terms of qualia structure.¹²

b) Although constants or primes are claimed to have an ontological status, nothing is said about the format of the ontology, which is thought to serve as the theoretical basement:

“we assume that the set of constants is open-ended. Each constant has an ontological categorization (...) drawn from a fixed set of types (i.e. state, thing, place, manner, etc.). [...] we assume that the ontological type of a constant determines its basic association with a particular event structure template; (...)” (Rappaport and Levin 1998: 108)

The same can be said of the mark “*lcp*” in the TGL. This feature is associated to a number of ontological units but to the best of our knowledge we have not seen the ontology that supports this decision-making process.

c) The resulting representations only include those aspects that are grammatically relevant, which signifies that a predicate’s meaning is confined to those aspects that are syntactically visible. However, lexical meaning goes beyond syntax and, in fact, it includes semantic, pragmatic and world knowledge that seems to be fundamental for, say, the development of natural language applications, i.e. tools for crosslinguistic information retrieval.

Hence, in the light of these observations FunGramKB has formulated the notion of CLS, a semantic syntax-motivated formalism that is more robust than semantic decompositional approaches such as those described above. Semantic representation now takes the form of a CLS, which is defined as “*a lexically-oriented schema containing concepts; that is, a CLS is not made of lexical material*”. This proposal provides compelling answers to some of the thorny issues as posited in most lexical representation theories, i.e. the universal nature of primes and constants. Given that a CLS is made of concepts, and concepts are shown to be part of an ontology, and an ontology is supposed to capture the world knowledge we all have

¹¹ In a similar fashion, the following works within the LCM have provided analytical tools to provide fully-fledged semantic decompositions of both state and activity predicates (cf. Mairal and Faber 2002, 2007, Mairal and Ruiz de Mendoza 2008, 2009).

¹² Recall that the qualia structure, according to the TGL, includes the following four features of a predicate’s meaning: constitutive (Q_C): the relation between an object and its constituent parts (material, weight, parts and component elements); formal (Q_F): that which distinguishes it within a larger domain (i.e. orientation, magnitude, shape, dimensionality, color, position); telic (Q_T): its purpose and function; agentive (Q_A): factors involved in its origin or ‘bringing it about’ (i.e. creator, artifact, natural kind, causal chain.). We refer to Pustejovsky 1995: chapter 5, especially section 5.4, and chapter 6.

(cognitive knowledge), then it follows that concepts have a universal grounding, and by extension a CLS has crosslinguistic validity.

Recent research has developed the following inventory of CLSs, which has a very similar format to standard logical structure with the exception that primes are now substituted by concepts, signaled by the notational device <C>.

VERBAL CLASS	CONCEPTUAL LOGICAL STRUCTURE
State	<C> (x) or (x,y)
Activity	do (x, [<C> (x) or (x,y)])
Achievement	INGR <C> (x) or (x,y), or INGR do (x, [<C> (x) o (x,y)])
Semelfactive	SEML <C> (x) or (x,y) SEML do (x, [<C> (x) or (x,y)])
Accomplishment	BECOME <C> (x) or (x,y), or BECOME do (x, [<C> (x) o (x,y)])
Active accomplishment	do (x, [<C> (x, (y))]) & BECOME <C> (z,x) or (y)
Causative	α CAUSE β , where α and β are conceptual logical structures of any type.

Figure 4: Inventory of CLSs.

A predicate's lexical entry is no longer formed by primes or constants but real conceptual units, which have their own conceptual properties, i.e. a thematic frame and a meaning postulate. The interaction between the linguistic and the ontological level is actually accounted for in terms of a linguistic-conceptual linkage, which we summarize below (see Mairal, Ruiz de Mendoza and Perrián 2011, for a detailed explanation). In the lexicon as part of the core grammar, each predicate is provided with a number of different properties that range from its basic aktionsart (A), number of variables (Ai), macrorole assignment (Aii), selection of semantic roles (Aiii), and idiosyncratic features like prepositions and collocations (Aiv). Moreover, syntactic knowledge is also specified via the type of syntactic constructions. Figures 5 and 6 show the interface of the core grammar lexicon where all these aforementioned features are specified:

LCM CORE GRAMMAR:

AktionsArt:

- State
- Activity
- Accomplishment
- Achievement
- Semelfactive

You determine the canonical lexical class(es) of the verb.

Variables: x, y

Idiosyncratic features:

[MR 1], [U = <--- no value selected --->]

Thematic frame mapping:

X = Theme, Y = Goal, Z = [no function]

A REMINDER OF FUNGRAMKB PARTICIPANTS:
 AGENT: Entity that makes another entity move.
 THEME: Entity that changes its place or position.
 LOCATION: Location in which an entity moves.
 ORIGIN: Location from which an entity moves.
 GOAL: Location to which an entity moves.

Lexical Template:

Prepositions:

X = aboard, Y = aboard, Z = aboard

N to towards

Collocations:

Figure 5: Part of the LCM core grammar.

CONSTRUCTIONS:

AFFECTING TRANSITIVITY:
 alternations involving a change in the verb's transitivity

- Causative/inchoative alternation
- Conative alternation
- Induced action alternation
- Middle construction
- Preposition drop alternation

PHRASE SHIFT:
 alternations involving the shift of some phrase found with the verb but without a change in transitivity

- Apart-reciprocal alternation (intransitive)
- Apart-reciprocal alternation (transitive)
- As alternation

PHRASE ADDITION/REMOVAL:
 alternations involving a change in the number of phrases found with the verb but without a change in transitivity, resulting in oblique subject alternations

- Abstract cause subject alternation
- Container subject alternation
- Instrument subject alternation
- Location subject alternation
- Locatum subject alternation

MISCELLANEOUS:
 other constructions

- Caused-motion construction (intransitive)
- Caused-motion construction (transitive)
- Cognate object construction

Figure 6: The inventory of syntactic constructions in the lexicon.

Once these features are specified we arrive at a CLS via an automatic process, i.e. a CLS constructor. This is an algorithmic tool that takes conceptual and lexical information as input and automatically delivers the basic CLS of a predicate as output. Figure 6 shows the result of the two CLSs for the predicate *run* in English, an activity and an active accomplishment CLS:

CLS Constructor

ACTIVITY:
 do (x-Theme, [+RUN_00 (x-Theme, y-Goal)])

ACTIVE ACCOMPLISHMENT:
 do (x-Theme, [+RUN_00 (x-Theme)]) & INGR +RUN_00 (x-Theme, y-Goal)

Figure 6: The CLS Constructor.

A relevant issue that cannot be underestimated is the fact that a CLS connects up to the ontology, or else, to the world knowledge we have about a predicate. In other words, through

a CLS we find a nice way to link linguistic knowledge with world knowledge. Continuing with our example of the predicate *run*, this lexical unit is linked to the basic concept +RUN_00, which, in turn, is provided with a thematic frame and a meaning postulate in the ontology, as shown in figure 7:

Conceptual Information:	
LEXICAL UNIT:	run
CONCEPT:	+RUN_00
THEMATIC FRAME:	(x1)Agent (x2: +HUMAN_00 ^ +ANIMAL_00)Theme (x3: +GROUND_00)Location (x4)Origin (x5)Goal
MEANING POSTULATE:	+(e1: +WALK_00 (x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal (f1: +FAST_00)Speed)
DESCRIPTION:	move fast by using one's feet, with one foot off the ground at any given time

Figure 7: Conceptual properties for +RUN_00.

From the meaning postulate we can retrieve knowledge that goes beyond syntax and gather that *run* is a verb of movement whereby a human or an animal moves on the ground from a location to a destination. Additionally, this movement is performed fast. All these parameters are not visible to syntax but are part of our world knowledge.

Finally, a CLS is seen as an interlingua, a language-independent representation, which performs a twofold function: (i) it provides the format for the semantic representation that is the initial step in the semantic-to-syntax linking algorithm; (ii) it serves as input for COREL, which is the machine readable metalanguage that is used for the reasoning engine. Figure 8 illustrates the relationship between syntactic and conceptual representation:

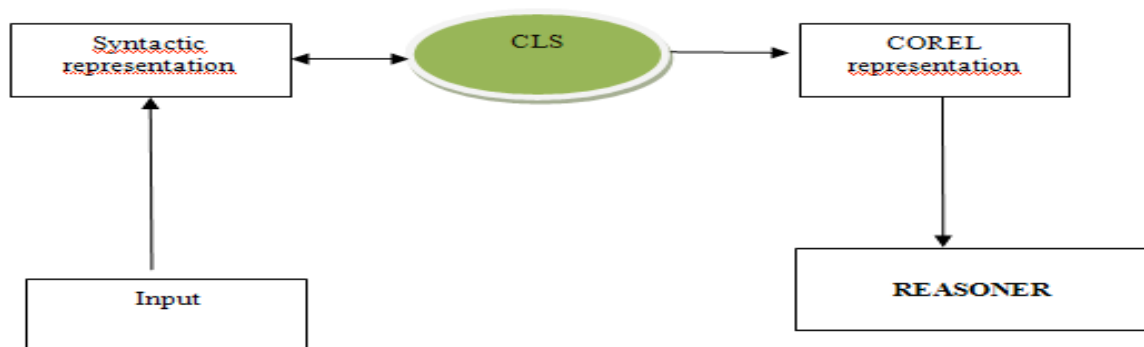


Figure 8: Linking syntactic and conceptual representation in FunGramKB.

4. Selectional preferences and collocations

One of the most pervasive issues in lexical representation is to account for the nature of selection restrictions, i.e. the semantic/pragmatic constraints that define the referential nature of the different arguments in a predicate's lexical entry. This has been a drawback for most theories, both formal and functional, and consequently these have always invoked too general principles that are stipulated on an *ad hoc* basis. A case in point is RRG that, for example, accounts for the fact that the first argument in the logical structure of verbs of perception, cognition, propositional attitude, emotion and internal experience must be a sentient, animate entity (Van Valin and LaPolla 1997: 156). As shown elsewhere (Jiménez and Pérez 2011), an ontological /conceptual approach offers a standardized procedure to account for the nature of selection restrictions.

First, as a natural consequence of the a clear demarcation between the linguistic and the conceptual level, the broad, and often unclear notion of “selection restrictions” is now substituted by “selectional preferences” and “collocations”, each working at the conceptual and the linguistic level respectively. Hence, selectional preferences are part of the ontology and reflect our knowledge about the world, i.e. they are language-independent labels. For example, if we return to the concept +WEAVE_00 the two arguments in the thematic frame are marked with the selectional preferences +HUMAN_00, +CLOTH_00 ^ +BASKET_00, which are, in turn, conceptual units that are part of the knowledge base, and more particularly, of the ontology.

- (2) +WEAVE_00
 Thematic Frame:
 (x1: +HUMAN_00)Theme (x2: +CLOTH_00 ^ +BASKET_00)Referent

Selectional preferences are then seen as conceptual constraints that are prototypically related to cognitive situations. They are not word-oriented but conceptually-oriented. It might be the case that selectional preferences are so tightly linked to specific conceptual units that new conceptual units are created. Here comes the notion of subconcepts. A subconcept is a refinement of the selectional preferences in the thematic frame of an already existing basic or terminal concept. This means that in our endeavour to design a knowledge base that is based on a fine-grained semantic approach we are sensitive to those cases in which the conceptual restriction or specification takes place exclusively in one or all of the participants of the thematic frame of a basic or terminal concept, without varying its meaning postulate. Here is a sample of four subconcepts that belong in the terminal concept \$SOUND_00 (cf. Periñán and Mairal (fact)):

-BARK	(x1: +DOG_00)Theme (x2: +SOUND_00)Referent
-BLEAT	(x1: +SHEEP_00)Theme (x2: +SOUND_00)Referent
-BUZZ	(x1: +BEE_00 ^ +FLY_01)Theme (x2: +SOUND_00)Referent
-CACKLE	(x1: +CHICKEN_00)Theme (x2: +SOUND_00)Referent

Table 3: A sample of subconcepts from \$SOUND_00.

Thus far, this is what goes on at the cognitive level. However, how does FunGramKB account for “collocations”? Collocations, unlike selectional preferences, are word-oriented so they are stored in their appropriate lexica, depending on the language the word is associated with. Collocations are language specific and cannot be retrieved from our knowledge of the world, as was the case for selectional preferences. Consequently, each lexica in FunGramKB has a slot to encode these language-specific features. For example, if we look up the word “perder” (“lose”) in the Spanish lexicon, on the top of the page we find the conceptual information of the conceptual unit to which “perder” is associated to (i.e. +LOSE_00). The thematic frame of this conceptual unit captures what we all know about the cognitive situation invoked by this conceptual unit: there are two arguments, the first one can be human or animal, and the second is a corpuscular object.

The screenshot shows a search interface for the Spanish predicate "perder". On the left, there is a search bar with "perder" and a list of suggestions including "a menudo", "a salvo", "a veces", "abajo", "abarcar", "abastecer", "abatido", "abatimiento", "abdomen", and "abecé". Below the suggestions is a keyboard layout with letters A-Z. A status bar indicates "303 / 2625 lexical units (2877 senses)". The "Part of Speech" section shows radio buttons for N, V, A, and a selected "no filter" option. The main area displays "Conceptual Information:" for "perder" (+LOSE_00). The information is structured as follows:

LEXICAL UNIT:	perder
CONCEPT:	+LOSE_00
THEMATIC FRAME:	(x1: +HUMAN_00^+ANIMAL_00)Theme (x2: +CORPUSCULAR_00)Referent
MEANING POSTULATE:	+(e1: +LACK_00 (x1)Theme (x2)Referent (f1: (e2: past +PUT_00 (x1)Agent (x2)Theme (x3)Origin (x4)Goal))Reason (f2: (e3: n +REMEMBER_00 (x1)Agent (x1)Theme (x4)Referent))Reason)
DESCRIPTION:	place (something) where one cannot find it again; "I misplaced my eyeglasses"

Figure 9: Conceptual properties of the Spanish predicate “perder”.

However, the Spanish predicate *perder* collocates with a finite number of words, that is, its collocations. These are language-specific and are not part of our world knowledge. Therefore, these should go to the linguistic level, and more particularly, to the Spanish lexicon, as shown in the following figure:

The screenshot shows a "Collocations:" interface for the Spanish predicate "perder". It features three input fields labeled X, Y, and Z, each with a "Y" button and a "N" button. The Y field is currently active and displays a list of collocations: "camino", "dirección", "norte", and "rumbo".

Figure 10: Collocations of the Spanish predicate “perder”.

Provided that selectional preferences are language specific, it comes as no surprise that the predicates *atesorar* and *hoard* that belong in to the same metaconceptual dimension in the ontology, #POSSESSION > +STORE_00, have different collocational patterns as shown in the corresponding Spanish and English lexica:¹³

The screenshot shows a "Collocations:" interface for the Spanish predicate "atesorar". It features three input fields labeled X, Y, and Z, each with "Y" and "N" buttons. The Y field is active and displays a list of collocations: "éxito", "información", "recuerdo", "secreto", and "victoria".

Figure 11: Collocations of *atesorar* (hoard) in FunGramKB’s Spanish lexicon.

The screenshot shows a "Collocations:" interface for the English predicate "hoard". It features three input fields labeled X, Y, and Z, each with "Y" and "N" buttons. The Y field is active and displays a list of collocations: "information", "time", "flaw", "misery", and "nostalgia".

Figure 12: Collocations of *hoard* in FunGramKB’s English lexicon. +HUMAN_00)Them2: +CLOTH_00.

¹³ These collocates have been obtained from REDES (for Spanish) and the *Collins Corpus Concordance and Collocation Sampler* (for English).

5. Applications in the industry

One of the applications that is currently being developed is the expansion of the general purpose ontology towards specialized ontologies or satellite ontologies. Satellite ontologies deal with knowledge representation of specialized domains, for example legal or environmental languages. The following figure shows the interaction between the core ontology and the satellite ontologies (cf. Mairal, Periñán and Samaniego 2011):

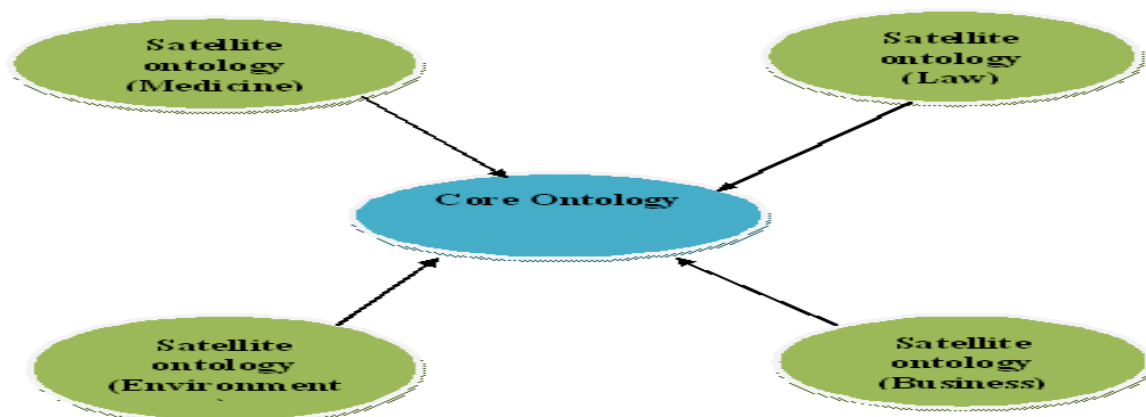


Figure 12: Core Ontology and satellite ontologies.

Now, a FunGramKB user has access to a rich inventory of lexical information that has been extracted from both the ontology and the lexicon. Here is an example with the verb “run”:

Lexica

This is a sample of the data stored in the FunGramKB lexica.

Insert a verb:

English
Ⓢ

1 2 3	
Headword	run
Concept	+RUN_00
Thematic Frame	(x1)Agent (x2: +HUMAN_00 ^ +ANIMAL_00) Theme (x3: +GROUND_00) Location (x4)Origin (x5)Goal
Meaning Postulate	+(e1: +WALK_00 (x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal (f1: +FAST_00)Speed)
POS	v
Paradigm	Irregular
Tense Constraints	no
Reflexivity	Never
Reciprocity	Never
Aktionsart	Activity, Active accomplishment
Lexical template: variables	X-Theme, Y-Goal
Lexical template: restrictions	MR1
Prepositions	Y = to, towards
Constructions	Induced action alternation
Dialect	Standard
Usage	Common
Topic	FACTOTUM
Examples	1. When the boat was under way she looked back and there were the children running in the dunes [BNC/FPF 1542] 2. Frightened by the stranger he ran to the kitchen [BNC/AMB 49]
1 2 3	

Figure 13: Linguistic and conceptual information for the predicate run.

The primary aim is to use all this data in the development of natural language processing applications for crosslinguistic information retrieval, machine translation (Periñán and Mairal 2010) or even for lexicography, in the sense that conceptual search queries will be readable and interpretable by the knowledge base.

5. Conclusions

This paper discusses the methodological impact the ontological approach adopted for the LCM has had in lexical representation. We claim that the notion of CLS is shown to have more explanatory and expressive power than standard decompositional representations like RRG's logical structures. From a CLS we have access to world knowledge as encoded in the ontology and therefore these representations go beyond those aspects that are syntactically relevant and provide a nice format to combine both linguistic and non-linguistic knowledge. Moreover, an ontological approach also offers solutions to the pervasive nature of selection restrictions in the sense that a twofold division is followed between selectional preferences and collocations. The former are part of the ontology and encode world knowledge, while the latter are language-specific and part of the lexicon. Finally, we have given a brief account of some of the current applications in the development of natural language processing and specialized language representation.

Acknowledgements

We would like to thank Dr. Ricardo Mairal for his invaluable help and continued support.

References

- Baicchi, A. 2010. Metaphoric motivation in grammatical structure. The case of the caused-motion construction from the perspective of the Lexical-Constructional Model. In G. Radden, K.U. Panther and P. Koch (eds.), *Motivation in Lexicon, Grammar, and Discourse*. Amsterdam/Philadelphia: John Benjamins.
- Bod, R. 2009. Constructions at work or at rest? *Cognitive Linguistics* 20/1: 129-134.
- Butler, C.S. 2009. The Lexical Constructional Model: Genesis, strengths and challenges. In C.S. Butler and J. Martín (eds.), *Deconstructing Constructions*. 117-152. Amsterdam/Philadelphia: John Benjamins.
- Del Campo, N. 2011. Cognitive modelling in illocutionary meaning. *Review of Cognitive Linguistics* 9: 392-412.
- Goldberg, A. 1995. *A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Jackendoff, R. 1990. *Semantic Structures*. Cambridge, Mass: MIT Press.
- Jiménez Briones, R. and M.B. Pérez Cabello de Alba (2011) An account of selection restrictions in Role and Reference Grammar. *RCEI* 62: 99-122.
- Mairal Usón, R. and P. Faber. 2002. Functional Grammar and Lexical Templates. In R. Mairal and M.J. Pérez Quintero (eds.), *New perspectives on predicate argument structure in Functional Grammar*. 41-98. Berlin and New York: Mouton de Gruyter.
- Mairal, R. and Faber, P. 2007. Lexical templates within a functional cognitive theory of meaning. *Annual Review of Cognitive Linguistics* 5: 137-172.

- Mairal, R. and F.J. Ruiz de Mendoza. 2008. New challenges for lexical representation within the Lexical-Constructional Model (LCM). *Revista Canaria de Estudios Ingleses* 57: 137-158.
- Mairal-Usón, R., F.J. Ruiz de Mendoza and C. Perrián-Pascual. Constructions within a Natural Language Processing Knowledge Base. In H. Boas and F. González-García (eds.), *Construction Grammar goes Romance*. Amsterdam/Philadelphia: John Benjamins. (Forthcoming)
- Mairal, R. and C. Perrián. 2009. The anatomy of the lexicon component within the framework of a conceptual knowledge base. *Revista Española de Lingüística Aplicada* 22: 217-244.
- Mairal, R. and C. Perrián 2010. Role and Reference Grammar and Ontological Engineering. In J.L. Cifuentes, A. Gómez, A. Lillo, J. Mateo and F. Yus (eds.) *Los caminos de la lengua. Estudios en homenaje a Enrique Alcaraz Varó*. 649-665. Alicante: Universidad de Alicante.
- Mairal R., C. Perrián and M.B. Pérez. La noción de estructura lógica conceptual. In R. Mairal, L. Guerrero and C. González. *El funcionalismo en la teoría lingüística. La Gramática del Papel y la Referencia. Introducción, avances y aplicaciones*. Akal: Madrid. (Forthcoming)
- Mairal, R. and Perrián, C. Universality and culturality in a Natural Language Processing Knowledge Base. (forthcoming)
- Mairal, R., C. Perrián and E. Samaniego 2011. Using ontologies for terminological knowledge representation: a preliminary discussion. In N. Talaván, E. Martín Monje y F. Palazón (eds.), *Technological innovation in the teaching and processing of LSPs: Proceedings of TISLID'10*. UNED: Madrid.
- Mairal, R. and F.J. Ruiz de Mendoza, F. 2009. Levels of description and explanation in meaning construction. In C. Butler and J. Martín (eds.) *Deconstructing Constructions*. 153-198. Amsterdam: John Benjamins.
- Peña, M.S. 2009. Constraints on subsumption in the caused-motion construction. *Language Sciences* 31: 740-765.
- Perrián, C. and F. Arcas. 2004. Meaning postulates in a lexico-conceptual knowledge base. *Proceedings of the 15th International Workshop on Databases and Expert Systems Applications*, IEEE: 38-42.
- Perrián, C. and F. Arcas. 2005. Microconceptual-Knowledge Spreading in FunGramKB. *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*. 239-244.
- Perrián, C. and Arcas, F. 2007a. Cognitive modules of an NLP knowledge base for language understanding. *Procesamiento del Lenguaje Natural* 39: 197-204.
- Perrián, C. and F. Arcas. 2007b. Deep semantics in an NLP knowledge base. *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence*. 279-288.
- Perrián, C. and F. Arcas. 2010a. The architecture of FunGramKB. *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. 2667-2674.
- Perrián, C. and F. Arcas. 2010b. Ontological commitments in FunGramKB. *Procesamiento del Lenguaje Natural* 44: 27-34.
- Perrián, C. and R. Mairal. 2009. Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural* 43: 265-273.
- Perrián, C. and R. Mairal. 2010. Enhancing UniArab with FunGramKB. *Procesamiento del Lenguaje Natural* 44: 19-26.
- Perrián, C. and Mairal, R. 2010. La Gramática de COREL: un lenguaje de representación conceptual. *Onomazein*. 21/1: 11-45.

- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, Mass.: MIT Press.
- Rappaport Hovav, M. and B. Levin. 1998. Building verb meanings. In M. Butt and W. Geuder (eds.), *The Projection of Arguments: Lexical and Compositional Factors*. 97-134. Stanford, CA: CSLI Publications
- Ruiz de Mendoza, F.J. and A. Luzondo. Lexical-constructional subsumption in resultative constructions in English. In M. Brdar, *et al.* (eds.), *Cognitive Linguistics. Between Universality and Variation*. Cambridge Scholars Publishing. (Forthcoming).
- Ruiz de Mendoza, F.J. and F. González. 2011. Constructional Integration in the Lexical Constructional Model. *BAS (British and American Studies)* 17: 75-95.
- Ruiz de Mendoza, F.J. and R. Mairal. 2008. Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia Linguistica* 42/2: 355-400.
- Ruiz de Mendoza, F.J. and R. Mairal 2010. Constraints on syntactic alternation: lexical-constructional subsumption in the Lexical-Constructional Model. In P. Guerrero (ed.), *Morphosyntactic Alternations in English. Functional and Cognitive Perspectives*. 62-82. London, UK/ Oakville, CT: Equinox.
- Van Valin, R. D. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- Van Valin, R.D. and D. Wilkins. 1993. Predicting syntax from semantics. In R.D Van Valin Jr. (ed.), *Advances in Role and Reference Grammar*. 499–534. Amsterdam/ Philadelphia: John Benjamins.
- Van Valin, R.D. and R. LaPolla. 1997. *Syntax, Structure, Meaning and Function*. Cambridge: Cambridge University Press.