Decomposing semantic decomposition: Towards a semantic metalanguage in RRG[1]

Ricardo Mairal Usón
UNED, Madrid, Spain

Pamela Faber
University of Granada, Spain

1 Introduction

Although logical structures in Role and Reference Grammar (RRG) aspire to cross-linguistic validity, they lack an enhanced semantic component capable of encoding the set of semantic and pragmatic parameters that underlie the meaning of each predicate. Such a component would require the specification of criteria upon which to base the selection of semantic primitives. The existence of a set of undefinables has the advantage of permitting the systematic description of predicate meaning within a unified framework. However, this is presently not the case in RRG because no standardized procedure for this type of semantic codification has so far been specified. In this article we propose the use of a core set of semantic primitives as the basis for a semantic metalanguage or controlled vocabulary for the conceptual description of predicates. This would enrich logical structures by making them more systematic.

2 Setting the scene

Although the inventory of logical structures has been proven to have cross-linguistic validity (cf. Van Valin, in press), the status of the set of primitives that form an integral part of these representations is somewhat less certain. To the best of our knowledge, the RRG Framework has never specified a set of semantic primitives or even a metalanguage for conceptual representation. For this reason, it is our opinion that the present inventory of logical structures is not systematic in its treatment of primitives. Example (1) shows that certain activity predicates defined by logical structures suffer from circularity since their definiens coincides with the definiendum:

(1) Activity predicates

sing	**do'** (x, [**sing'** (x)])
walk	**do'** (x, [**walk'** (x)])
drink	**do'** (x, [**drink'** (x)])

In other cases the central part of the definition of a predicate is the past participle of the same term being defined. This occurs in state predicates that encode the end result of the accomplishment /activity event:

(2) State predicates

melt:	BECOME **melted'** (x)
shatter:	INGR **shattered´** (x)
break:	**do'** (x,ϕ)] CAUSE [BECOME **broken'** (y)]

---

[1] Financial support for this research has come from the following sources: (1) the research project funded by the Comunidad Autónoma de Madrid., grant no. 06/HSE/0132/2004; (2) research project BFF2003-04720, funded by the Spanish Ministry of Education.

The examples in (2) contrast with those in (3) which point to a more fine-grained semantic decompositional system:

(3) Full semantic decompositional system:

learn	BECOME **know'** (x, y)
receive	BECOME **have'** (x, y)
kill	[**do'** (x, ∅)] CAUSE [BECOME [**dead'** (y)]
show:	[**do'** (x, ∅)] CAUSE [BECOME [**see'** (y,z)]
cook:	[**do'** (x,ϕ)] CAUSE [BECOME **baked'** (y)]

The preceding examples in (3) are evidently based on the presupposition that certain predicates are more basic than others. From the semantic side of the fence, we might well ask *sing, walk, drink, melt* and *shatter* can truly be regarded as primitives, and if their meaning is not open to further decomposition. These predicates appear to differ substantially from others such as *have* and *know,* which seem to be better candidates for universal entities.

Nonetheless, there have been serious attempts to provide richer semantic representations in terms of lexical templates, as can be seen in the examples below (cf. Van Valin and Wilkins, 1993; Van Valin and LaPolla, 1997, Mairal, 2003, 2004):

(4) Speech act verbs and the entry for *promise* (Van Valin and La Polla, 1997:117)

a. **do'** (x, [**express**(α).**to**(β).**in.language**.(γ)' (x,y)])
b. **do'** (x, [**express**(α).**to(β).in.language**.(γ)' (x, y)]) CAUSE [BECOME **obligated'** (x, w)
α = w
β = y

(5) The predicate *remember* (Van Valin and Wilkins, 1993: 511)

BECOME think.again (x) about something.be.in.mind.from.before (y)

(6) The lexical class of *contact-by-impact* verbs (Mairal, 2003)

[[**do´** (w, [**use.tool.**(α).**in.**(β).**manner.for.**(δ)´ (w, x)] CAUSE [**do´** (x, [**move.toward´** (x, y) & INGR **be.in.contact.with´** (y, x)], α = x.

The problem in (4), (5) and (6) resides in the fact that *express, obligated, think, move* are regarded as primitives. There seems to be no reason or explicit criteria for such a decision. It is as though *express,* for example, had been plucked out of the air without the use of any heuristic procedure for designating a set of primitives or, for that matter, an inventory of semantic fields, not to mention a description of their internal organization.

In this sense it is doubtful that **obligated'**, which is posited as the primitive predicate for the lexical entry for *promise* in (4), can even be remotely called a primitive. Furthermore, in (5) and (6), although *think* and *move* are putative universals, nothing is said of how these predicates have been arrived at or where they have come from. Although couched in more elaborate semantic decompositions, Mairal's (2003) lexical templates are still not systematic enough in their use of activity and state primitives. Primitives such as *manner*, *tool*, and *use* appear in these representations, but no explanation is given of how they have been obtained.

In this article we propose a framework for the specification of a set of undefinables together with a set of combinatory rules. This preliminary proposal is based on previous research in semantic analysis as well as extensive lexical and cross-linguistic analyses carried out in the framework of Wierzbicka's Natural Semantic Metalanguage (Wierzbicka 1980, 1987, 1995, 1996), Mel'cuk's Text-Meaning Theory (Mel'cuk 1988, 1989, 1996), and the Functional-Lexematic Model (Martín Mingorance 1990, 1995; Faber and Mairal 1999).

3 Towards a semantic metalanguage

The establishment of a semantic metalanguage is a complex task because it presumably entails some type of semantic decomposition, and would make explicit, at least to a certain extent, the relation between language and thought[2].

Semantic decomposition of any kind is problematic because there is no simple solution for the atomization of meaning. Although semantic features or attributes are not presently in fashion, they still pop up in different guises within a wide variety of approaches. There is much discussion as to the possible nature of such units, but whether they are encoded in natural language or conceptual metalanguage[3], the final inventory must be systematic, finite, and with some sort of internal organization. It is simply not feasible to create an endless *ad hoc* list of semantic primitives to be used every time the need arises. Furthermore, any such inventory must be created on the basis of methodological principles that justify all choices.

Even though semantic decomposition may at first appear to be an impossible enterprise, the intuition persists that smaller meaning units must exist at some level to encode conceptual content[4]. Our proposal for a semantic metalanguage consists of two basic modules, which are subject to cross-linguistic validity and psychological adequacy:

a. The semantics of the metalanguage.
b. The grammar (or syntax) of the metalanguage.

The following theses constitute the basis of our framework. The first two (T1 and T2) refer to the semantics of the metalanguage, while the remaining three (T3, T4 and T5) refer to the grammar as well as the typological and psychological adequacy of the theory:

- [T1]: <u>Organization of the lexicon</u>: Semantic primitives are extracted from a lexicon organized into lexical classes and lexical hierarchies.
- [T2]: <u>Procedure for the specification of primitives</u>: Factorization determines where the chain of the decompositional system actually ends.
- [T3]: <u>Syntax of the primitives</u>: The semantic metalanguage proposed consists of a list of primitives and a set of operators that are used as a basis for the encoding

---

[2] As Levinson (1997:13) points out, the problem lies in the relation between the medium in which we think and the medium in which we talk. He raises the question of whether or not there is a distinction between semantics and underlying conceptual representations. Even though there are many reasons why the relation may not be direct or isomorphic, the fact remains that some type of relationship must exist. It is a question of finding it.

[3] Jackendoff's Conceptual Structures are a case in point.

[4] Cognitive models are a good example of this line of research, i.e. Lakoff (1987). However, things are changing and there is a tendency to identify what they call primary metaphors with the status of primitives or near primitives. The same can be said of the notion of construction which is at first accorded a universal status whereas in subsequent publications it is claimed that constructions are language-specific. See Goddard (1998) for a critique of those arguments which reject the existence of primitives.

of meaning. Operators and semantic primitives are combined to encode the conceptual meaning of lexical units whose structured meaning definitions signal the major distinctions within lexical domain organization.

- [T4]: Typological adequacy: Arriving at a set of undefinables that can serve as the basis for a semantic metalanguage means the adoption of a system of lexical decomposition which can function cross-linguistically[5]. In this regard, we show that each domain has a set of functions that act on the superordinate term to give more specific hyponyms and codify the most relevant subdomains. However, the degree of lexicalization in each domain evidently varies from language to language. Hence both the set of lexical functions and the primitives are regarded as universal though the specific combination of both is language specific. Despite the fact that metaphoric extensions for lexical units may vary from language to language, the first and primary meaning generally coincides. If there were no shared conceptual meanings that underlie texts in different languages, interlinguistic translation would be impossible.[6]
- [T5]: The conceptual basis of the metalanguage: Ideally, the metalanguage established through semantic decomposition would be adequate for the codification of conceptual representations. Not surprisingly, conceptual representation is an important cognitive function that lies at the center of language.

All of the above assertions seem to point to the fact that semantic decomposition in some form is possible and ultimately desirable. We acknowledge that this is uncomfortable terrain for those who seek clear-cut solutions to linguistic representation because meaning is undeniably messy. Nevertheless, lexical items happen to be the linguistic designation of concepts, and linguists who are inevitably forced to enter the fuzzy realm of word meaning have little choice but to deal with issues such as categorization, conceptual representation, and a semantic metalanguage for conceptual description.

3.1. The semantics of the metalanguage

Linguistic theories that endeavour to account for syntactic structures and leave meaning for later inevitably encounter stumbling blocks because language is not so much about grammatical constructions as about meaning. Thus, in a manner of speaking, the lexicon, our conceptual storage house, is "where the action is". As is true for most storage places, its value and efficiency depend not only on the contents, but above all on how they are organized.

Since there is considerable neurological evidence in favour of a lexicon organized in semantic categories (Damasio and Damasio 1992; H. Damasio et al 1996), it seems reasonable to assume that the semantic component of a linguistic theory aspiring to psychological adequacy would be organized, at least at some level, according to areas of meaning. One might even go so far as to say that the syntax of lexical units would depend on their meaning instead of vice versa since meaning is prior to syntax. As is well known, each lexical unit is linked to an underlying meaning

[5] Important initiatives in this respect are Wierzbicka's Natural Semantic Metalanguage (NSM) (cf. Goddard and Wierzbicka, 2002), Mel'cuk's Meaning Text Theory (e.g. Wanner 1996), Jackendoff's (1983, 1991) conceptual semantics, or even ontological semantics (cf. Nirenburg and Raskin 2004).

[6] Note that the primitives, which are very similar to the NSM, and the operators, which resemble Mel'cuk's Lexical Functions, have been tested against a number of typologically different languages (cf. Goddard and Wierzbicka, 2002; Mel'cuk, 1989).

representation or concept, which is in turn linked to other concepts by means of different types of conceptual relations. In theory, such knowledge representations constitute the *tertium comparationis* that makes translation possible.

The question is how to find the type of onomasiological organization most in consonance with that of our mental lexicon as well as the conceptual invariants upon which such a structure would presumably be based.

3.1.1. The Functional Lexematic Model (FLM)

According to Levy (2003), it is only through study of the usage of terms in a public language that we can have an independent way of fixing the contents of people's concepts. This is in consonance with Dummett's (1991) Priority Thesis, which takes language first and concepts second. Since dictionaries are the codification of conceptual content in public language, they are valid texts for the extraction of conceptual information regarding meaning parameters, arguments, semantic roles, etc. It goes without saying that this type of information must afterwards be validated by means of corpus analysis. Lexical hierarchies, which reflect conceptual hierarchies, could thus be traced and constructed through the analysis of dictionary definitions[7].

The FLM (Martín Mingorance 1984, 1990, 1995), slightly modified from Faber and Mairal (1999), formulates the following inventory of lexical classes extracted by the extensive factorization of dictionary definitions. The resulting classes are the following:

Lexical domain	**Nuclear term**
EXISTENCE	be/happen
CHANGE	become
POSSESSION	have
SPEECH	say
EMOTION	feel
ACTION	do, make
COGNITION	know, think
MOVEMENT	move (go/come)
PHYSICAL PERCEPTION	see / hear / taste / smell / touch
MANIPULATION	use

In our opinion, the superordinate terms for each basic conceptual category can be regarded as possible candidates for the inventory of more basic terms or primitives that comprise one component of a semantic metalanguage.[8] Accordingly, they are the basis for the formulation of the meaning of more specific lexical items. These generic terms would be the starter terms for lexical hierarchies that would provide the core structure for the construction of a conceptual network with a rich set of relations. Each category would be organized in terms of meaning parameters establishing lexical dimensions that constitute the internal organization of the lexical domain.

Our proposal is that these meaning parameters can be encoded partly in terms of semantic primitives, such as those posited by Wierzbicka (1996), and partly in terms of

[7] This type of analysis is far from new. Amsler (1980) did precisely that in order to derive hyponymic information about nouns.

[8] This is in consonance with Mel'cuk's (1996:76) general principle of *Lexical Inheritance:* "All lexicographic data shared by a family of semantically related lexical units should be stored just once – under one LU (lexical unit) of the corresponding cable or under the generic LU of the corresponding semantic field, from where these data can be 'inherited' in each particular case."

lexical functions, similar to those formulated by Mel'cuk (1996). The generic terms or near primitives specified within the FLM (Faber and Mairal 1999) correspond to a great extent to those proposed by Wierzbicka (see appendix 1).

In this article we analyze the lexical domain of COGNITION in which *know* and *think* are evident candidates for universals (as also specified in the NSM). As shall be seen, these primitives can be articulated by means of a finite set of lexical functions to codify the basic conceptual meaning of lexical units.

3.2. The grammar of the metalanguage

One of the criticisms that has been levelled against Wierzbicka's NSM approach is that nothing is said (or at least nothing is said explicitly) about the formal nature ascribed to their representations, i.e. whether they are formulated in tree or dependency structures etc (cf. Van Valin and Wilkins, 1993: 504). Our proposal is compatible with the RRG lexical representation formalism, i.e. logical structures or the more elaborate version of lexical templates (cf. 3.2.2. and 4).

Regarding the tools that make up the grammar of the metalanguage, it is first necessary to find out how the set of indefinables combine so that we can define the whole set of predicates that converge within a lexical class. In connection with this, we propose a series of operators derived from those in RRG (e.g. CAUSE, BECOME, INGR) complemented by others, which either coincide with or are similar to Mel'cuk's Lexical Functions (e.g. MAGN, INSTR).

It has been observed that standard lexical functions within Mel'cuk's *Explanatory and Combinatorial Lexicology* (ECL) framework do not fully accommodate paradigmatic relationships (Grimes 1990; Fontenelle 1997) since such relationships refer to encyclopaedic knowledge rather than represent lexical relations[9]. L'Homme (in press) points out that although they have not been designed for that purpose, lexical functions are useful to reveal groups of semantically-related terms, and explains how this can be done. We have gone much farther in this respect that L'Homme and have made major changes to the original model by adapting functions so that they can account for the lexical domain-specific relationships to be represented.

3.2.1. The schema

Each lexical domain thus has a set of functions/operators that act on the superordinate term to generate more specific hyponyms and codify the most relevant lexical domains and subdomains. However, unlike Mel'cuk's lexical functions, our inventory of operators is used to organize the lexicon vertically instead of horizontally[10]. The list of lexical functions (or operators) used in the meaning representation of the verbs in this

[9] Within this inventory there is a certain lack of equilibrium since syntagmatic lexical functions are richer and considerably more numerous than paradigmatic ones. In the ECL framework, paradigmatic lexical functions, such as GENER and SPEC (CF. GRIMES, 1990) for generic-specific relationships, and MULT and SING for meronymic relationships are not sufficient in themselves to represent the internal structure of lexical domains. For this reason, as shall be seen, we have adapted syntagmatic functions and made them paradigmatic..

[10] According to Mel'cuk *et al* (1995: 126-127), A lexical function (LF) is written as: **f**(x) = y, where f represents the function, x, the argument, and y, the value expressed by the function when applied to a given argument. The meaning associated with an LF is abstract and general and can produce a relatively high number of values. For example, **Magn** is a function that expresses intensification. It can be applied to different LUs and produces a high set of values (e.g. **Magn** (smoker) = heavy; **Magn**(bachelor) = confirmed, etc.) (Mel'cuk *et al*, 1995: 126-127). Most of these functions refer to the combinatorial potential of lexical items. The ECL framework has a set of approximately sixty standard lexical functions divided into paradigmatic and syntagmatic functions (Mel'cuk 1998; Mel'cuk *et al.* 1995; Wanner 1996).

article can be found in Appendix 2 and constitute an expansion of the set of operators formulated in Van Valin and LaPolla (1997) for lexical representations.

A further issue concerns the development of a formalized system for lexical representation. In connection with this, we can either preserve the inventory of logical structures and try to integrate the new metalanguage, or alternatively, establish a new lexical representation system. We think that the best option is the former given the typological adequacy of lexical structures as well as the fact that they have been shown to work fairly well within the linking algorithm.

In consonance with Van Valin and LaPolla (1997) regarding speech act verbs and later in Mairal (2003, 2004) and Mairal and Faber (2002) regarding the internal structure of a lexical template, we believe that logical structures can be enriched by adding a new semantic component as shown in the following schema:

(7) [semantic representation] + logical structure = predicate

As seen in (7) the format of an RRG lexical entry would consist of two basic modules:

- the set of semantic parameters that differentiate one predicate from others within the same domain;
- a description of the event structure and the set of grammatically salient properties.

The first module is encoded by means of lexical functions that are essentially paradigmatic, while the second follows the orthodox practice of RRG. As for the notational device, we will be using two types of variables: internal and external variables. Internal variables are marked with numerical subscripts, while external variables are represented by Roman characters[11]. Let us consider the lexical entry for the following predicate:

(8) **fathom**: [MAGNOBSTR & $\text{CULM}_{12[\text{INTENT}]}$] **know'** (x, y)

The entry in (8) has two parts: (i) the semantic component in brackets; (ii) the representation of the logical structure. In this case, this predicate is represented by a state logical structure which takes know' as a primitive and has two arguments. Furthermore, this logical structure is in turn modified by a lexical function (or operator) MAGNOBSTR, which refers to the difficulty involved in carrying out an action, and in the case of *fathom*, there is great difficulty. As can be deduced from (8), lexical inheritance allows the packaging of enriched lexical information into one unified format given that hyponyms inherit the properties of their superordinate terms.

Independently of the complex semantic parameters that are characteristic of a predicate, one of the methodological prerequisites that cannot be violated is that all the units involved in the semantic representation must be drawn from the inventory of primitives and functions; hence, universally valid analytical units must be used. This is evident in the following lexical entry for *clarify*:

(9) **clarify**: [CAUS_{123}INSTR(BONCAUS(see))$_{123}$ & $\text{CULM}_{12[\text{INTENT}]}$]
do' (x, Ø) CAUSE [BECOME **know'** (x, y)]
x = 1; y = 2; z = 3

[11] Recall that Van Valin and LaPolla (1997: 117) use Roman and Greek letters to account for what they call external and internal variables.

In the same way as in the previous example, the lexical entry has two components; a causative accomplishment logical structure with three arguments and a semantic component which provides the distinguishing semantic specifications characteristic of this predicate: $\text{CAUS}_{123}\text{INSTR}(\text{BONCAUS}(\text{see}))_{123}$ [understand]. In *clarify,* an agent (arg1) causes (CAUS) a mental percept (arg2) to be understood better by a receiver-beneficiary (arg3).The means (INSTR) by which this is achieved is by causing (CAUS) somebody (arg3) to see(VISION) it (arg2) better (BON). As shall be observed, all the units in the lexical entry are part of the universal inventory of primitives (e.g. *see*) or operators.

We might ask ourselves why these representations are better than logical structures or lexical templates. If we compare this new formalism with logical structures, it is evident that the new representation is based on a list of semantic primitives and contains a richer semantic decompositional system. Let us compare the representation for *regre*t in terms of logical structure and the new formalism:

(10)

a. **regret'** (x, y)

b. $[\text{SYMPT}_{12}$ (sadness) INVOLV_{12}(want)$\text{DEGRAD}_{(\text{do})2} \leftarrow / _{(\text{become})2}\leftarrow]$ **feel'** (x, y)

This predicate is a commentative predicate which means that it asserts an emotional reaction and takes the complement as background (cf. Noonan, 1985). In such a case, the LS proposed in (10a) ignores this semantic interpretation and uses **regret** as a primitive, a decision which is highly questionable. In contrast (10b) includes a semantic representation which describes the emotional reaction that the effector experiences. The effector feels/experiences an emotion, a physical symptom (SYMPT) of sadness about *y*, an event. Furthermore, there is a subactivity INVOLV_{12} implied by the predicate such that the effector wants the second argument (an event) not to have happened (*become*)[12] in the past. The question of how to specify the semantics of the complement and the consequences for linking is an issue dealt with in section 5. Moreover, this representation is typologically adequate since it only uses elements that are extracted from the metalanguage. In this regard, **feel** seems to be a better candidate than **regret** as a primitive.

If we compare this new formalism to lexical templates, it is evident that this formalism has the advantage of being simpler and more consistent in the use of a semantic metalanguage.

4. The metalanguage in action: COGNITION

One of the reasons why it is so difficult to posit a valid set of semantic primitives is that it cannot be done without first having an overall view of the semantic field or lexical domain in question. The specification of primitives cannot be done only by focusing on token lexical units. Without a previous organization of conceptual information, one runs the risk of not perceiving the forest for the trees.

For this reason we are going to examine a series of lexical units and their configuration within the context of an entire lexical domain. Not surprisingly, the domain of COGNITION is one of the richest and most complex in the lexicon. Thinking is something that human entities do. It is part of our biological and psychological make-up. Despite linguistic and cultural differences, it would be difficult to imagine a language without a word to refer to this process. Not surprisingly, both *know* and *think* are included in Wierzbicka's inventory of undefinables.

[12] *Become* is another primitive. It is also part of the NSM inventory.

COGNITION consists of various dimensions that encode the different ways that we conceptualize a *thinking/knowing* experience or event. In this respect it can be regarded as a multidimensional concept, which can have different types and degrees of lexicalization, depending on the language.

Human cognition basically involves two components. The first is operational and consists of a mental process or activity (*think*). The second component refers to knowledge or the desired result of that process (*know*). *Know* and *think* or the underlying concepts that they represent, lie enmeshed in a conceptual network. The more basic a concept, the more connections it has with others.

On a syntagmatic level, such connectivity is evident in the wide variety of different complementation patterns that both verbs possess as well. This is an indication of their generality, or how high they are located in the semantic hierarchy of verbs of COGNITION. Although the syntax of a predicate is not fine-grained enough to base a theory of categorization on, it does provide significant clues regarding category assignment and whether a term is more general or more specific. On a paradigmatic level, the degree of generality of a lexical unit is reflected in the number of hyponyms that it generates.

However, the hyponyms of a more general lexical unit are not random sets, but form dimensions and subdimensions that are structured in terms of significant cognitive parameters. The application of operators to the generic undefinable of the lexical field generates the rest of the lexemes.

4.1. The inceptive dimension

In COGNITION, lexical units are defined in terms of *know*, which is the superordinate or generic lexeme for both *learn* and *understand. Learn* is conceptualized as a continuing process that should lead to a positive result, *understand*. *Understand* is thus seen as the result of learning or the culmination of *know.*

Learn, for example, can be defined as inceptive *know.* The superordinate term is *know,* which is considered a primitive, and all of the other terms are defined either directly or indirectly in terms of the superordinate:

(11)

Know *learn*	to come to **know**	
	find out	to **learn** sth intentionally
	discover	to **learn** sth unintentionally.
	memorize	to **learn** sth so that you can say it
	teach	to cause sb to **learn** sth by saying it to them or causing them to see it.

The various hyponyms of *know* are generated by applying functions to them. As previously mentioned, some of these functions are of our own invention and others have been adapted from Mel'cuk's lexical functions for our purposes to generate paradigmatic lexical structure[13].

[13] We are aware that strictly speaking, lexical functions are applied to the combinatorial possibilities of a lexical unit, and that this is a somewhat unorthodox variation of how they are used in ECL framework and Meaning-Text-Theory.

The first function used in this dimension is INCEP (the beginning of the lexical unit). If we apply it to *know,* we obtain *learn,* which signifies the inceptive state of *know. Learn* can thus be semantically represented in the following way:

(12) INCEP$_{12}$[know] = *learn*

Given that this predicate only encodes the beginning or inchoation of the accomplishment structure without the intervention of other parameters, the format of this lexical entry will only contain an accomplishment logical structure without any further semantic specification:

(13) BECOME **know'** (x, y)

In the same way as in the ECL framework, the two arguments of *learn* are represented by subscripts ($_{12}$). These are the basic arguments for all of the hyponyms of *learn,* which inherit the semantic information as well as a portion of the semantic information of the superordinate term. The first argument is a cognizer/experiencer, while the second is a percept/mental entity.

As examples of the English lexicalization of this subdimension, we have also represented four hyponyms of *learn* (i.e. *find out, discover, memorize,* and *teach*). *Find out* and *discover* lexicalize the presence or absence of intentionality on the part of the cognizer. They are obtained by applying other lexical functions and/or primitives to the representation of *learn* (i.e. INCEP$_{12}$[know]). The lexical function used to indicate intentionality is INTENT. This function is combined with another to indicate the presence (PLUS) or absence (MINUS) of intentionality.

(14) INCEP$_{12}$[know] = *learn*
- PLUSINTENT [learn] = *find out*
- MINUSINTENT [learn] = *discover*

As advanced above, these semantic representations are part of the larger lexical representation system such that the resulting formalism would have the following format:

(15) Representation for *learn*: BECOME **know'** (x, y)
(16) Representation for *find out*: [PLUSINTENT$_{12}$] BECOME **know'** (x, y)
(17) Representation for *discover*: [MINUSINTENT$_{12}$] BECOME **know'** (x, y)

Memorize is a goal-directed act. The purpose of this type of learning is to be able to say what one has learned. The purpose is represented by PURP, a function that we have created. The speech act itself is encoded by using *say,* the generic undefinable for the domain of speech act verbs, which is used as a semantic primitive. The subscripts ($_{12}$) refer to the human entity that effects the action (arg1) and the text that is learnt in this way (arg2).

(18) INCEP$_{12}$[know] = *learn*
- PURP(say)$_{12}$ [learn]= *memorize*

(19) [PURP(say)$_{12}$] BECOME **know'** (x, y)

The causative subdimension of these verbs is predictably generated by applying the operator CAUS. It can be argued, of course, that teaching does not always cause recipients to learn. In fact, it is possible to say:

(20) She taught the class Chinese for two hours, but no one understood a word that she said.

However, it should be pointed out that there is more than one *teach* (i.e. *teach* [SPEECH] and *teach* [COGNITION]). In (20) *teach* would be a speech verb since *teach* [COGNITION] necessarily entails mental activity or incipient modification of cognitive structures in the recipients of the action. What usually happens in this type of causative subdimension is that domains of abstract activity such as COGNITION begin to shade into ACTION because causation implies some sort of action on the part of the cognizer.

(21) INCEP_{12}[know] = *learn*
- CAUS_{3} [learn] = *teach*

(22) [CAUSINSTR_{123}(say)/CAUS(see)$_{123}$] **do'** (x, Ø) CAUSE [BECOME **know'** (y, z)

This lexical entry contains a causative accomplishment logical structure such that an effector (x) causes (y) to come to know something. Furthermore, this causal chain is modified by a semantic module which specifies that the effector carries out this causative action by means of using an instrument such that (x) says something to (y) or alternatively (x) causes (y) to see something.

4.2. Dimensions of COGNITION

The domain of COGNITION has various conceptual dimensions. In this paper due to restrictions of time and space, we can only provide salient examples from five of them. As shown, the hyponyms of *know* have been created by applying a set of operators to the more general term. As seen in section 4.1, the first of these dimensions envisions cognition as a goal-directed act in which the cognizer/experiencer comes to know something (*learn*) as well as one that leads to a final result (i.e. a state of knowledge (*understand*)). The second dimension conceptualizes cognition as the positive or negative achievement of understanding. The third dimension focuses on the temporal nature of the mental percept, in other words, whether it belongs to the past, present, or future. The fourth dimension regards the truth value of the proposition, which is in the mind of the cognizer. The last dimension is that in which cognition is linked to visual perception. For reasons of space, dimensions 3 and 4 will not be analyzed here although some passing comments will be made in section 5.

4.2.1. The achievement dimension

The achievement dimension of COGNITION is divided into two parts, one in which knowledge is achieved (*realize, grasp, fathom, etc.*) and a more negative one in which it is not (*misunderstand*). *Understand* is conceived as the highest point (or objective) (CULM) of *know.*

The predicates in this subdimension are generated through the application of functions such as INSTR (instrument) and OBSTR (implied difficulty). This is only natural because achieving understanding is arduous. Thus, it is done by means of strategies (INSTR) (i.e. *seeing in the mind, giving meaning*). LOC_{in} refers to a place which in this case is the mind conceptualized as an (abstract) body part.

This links *mind* to the NSM primitives, one of which is *body* and the meronymic relation PART_OF. Evidently, the only way to systematically generate nouns is for them to be linked to a concept within an ontology. Explicit reference to such an ontology is also seen in the category FEELING_TYPE, which is where the various feelings (*worry*, *surprise*, *doubt*) generated by lack of understanding are drawn from. *Feel* is also an NSM primitive.

know Arg$_1$ = Cognizer Arg$_2$ = Mental percept		
Causative subdimension: Arg1 = Cognizer Arg2 = Mental percept Arg3 = Recipient		
INCEP$_{12}$[know]		**learn** to come to know sth
	PLUSINTENT [learn] MINUSINTENT [learn] PURP(say)$_{12}$ [learn]	**find (out)** to learn sth intentionally. **discover** to learn sth unintentionally. **memorize** to learn sth so that you can say it
	CAUS$_3$ [learn] CAUSINSTR$_{123}$ (say)/CAUS(see)$_{123}$ [learn]	**teach** to cause sb to learn sth by saying it to them or causing them to see it.
CULM$_{12[\text{INTENT}]}$ [know]		**understand** to know the meaning of sth.
	INSTR (see)$_{12}$LOC$_{in}$ (BODY_PART: mind) [understand] INCEP$_{12}$[realize] INSTR (CAUS (have))$_{12\ ((\text{other})\ \text{INTENT})\ 3}$ [understand] OBSTR [understand] MAGNOBSTR [understand]	**realize** to understand sth by seeing it in the mind **dawn on** to realize sth gradually. **interpret** to understand sth by giving it a certain meaning. **grasp** to understand sth with difficulty **fathom** to understand sth with great difficulty
	CAUS$_{123}$ [understand] CAUS$_{123}$ INSTR (BONCAUS (see))$_{12\ 3}$ [understand] CAUS$_{123}$INSTR (say)$_{123}$ [understand]	**enlighten** to cause sb to understand sth **clarify** to cause sth to be understood (by sb) by causing it to be seen better. **explain** to cause sth to be understood (by sb) by saying things about it.
	DEGRAD$_{12}$ [understand]	**misunderstand** to understand sth wrongly.
	CAUSDEGRAD$_{123}$ [understand] INVOLVCAUS CONT(think)$_{123}$ [confuse] INVOLVCAUSSYMPT$_{12}$ (FEELING_TYPE: surprise, doubt) [confuse] INVOLVCAUSSYMPT$_{12}$ (FEELING_TYPE: worry) [confuse]	**confuse** to cause sb not to understand sth. **puzzle** to confuse sb, causing them to think for a long time. **confound** to confuse sb causing them to feel surprise and doubt. **perplex** to confuse sb, causing them to feel worry.

The following combinations of functions and primitives are an integral part of the format of the lexical entries in the domain. Let us discuss each of these definitions and their corresponding formalized representations from an RRG perspective:

(23)

- Understand: [Culm$_{12[\text{intent}]}$] **know’** (x, y)
  x = 1 and y = 2

Example (23) has a state logical structure modified by a semantic component which is coded in brackets. The semantic part, CULM$_{12[\text{INTENT}]}$ [know], has the following interpretation: the cognizer (arg1) comes to know the meaning of the mental percept (arg2). The concept of meaning is represented by the application of INTENT to arg2, which signifies its full intentionality.

(24)

- Realize: [Instr (see)$_{12}$Loc$_{\text{in}}$ (body_part: mind) & Culm$_{12[\text{intent}]}$] **know’** (x, y)
  x = 1 and y = 2

In the same way as in (23), example (24) has a state logical structure which is inherited from the superordinate term and a semantic description of the idiosyncratic properties of this predicate. This semantic part, INSTR (see)$_{12}$LOC$_{\text{in}}$ (BODY_PART: mind) [understand], is interpreted as follows: the cognizer (arg1) comes to know or understand a mental percept (arg2). This is done by *seeing* (semantic primitive) it in his/her *mind* conceptualized as a location (LOC$_{\text{in}}$). The mind is represented as an abstract body_part, which means it is in a partitive relationship to *body*, one of Wierzbicka’s semantic primitives. Note that this predicate inherits the properties of its immediate superordinate, a feature that we have marked with the symbol &.

(25)

- Interpret: [Instr (Caus (have))$_{12\ (\ (\text{other})\ \text{intent})\ \ 3}$ & Culm$_{12[\text{intent}]}$] **know’** (x, y)

Example (25) is a logical state structure plus a semantic description, INSTR (CAUS (have))$_{12\ (\ (\text{other})\ \text{INTENT})\ \ 3}$ [understand]. In *interpret,* the cognizer (arg1) understands a mental percept (arg2) causing it (CAUS) to *have* (semantic primitive) an*other* (semantic primitive) meaning. In the same way as in *understand,* the concept of meaning is represented by the application of INTENT to arg2, which signifies its full intentionality.

(26)

- Grasp: [Obstr & Culm$_{12[\text{intent}]}$] **know’** (x, y)

Both *grasp* (26) and *fathom* (8) inherit the properties of the superordinate *know* and the semantic part is codified by the operator OBSTR [understand]. The operator OBSTR refers to when there is difficulty involved in carrying out an action. In the case of understanding, this generates *grasp* and when there is great difficulty (MAGNOBSTR), *fathom.*

(27)

- Explain: [CAUS$_{123}$INSTR (say)$_{123}$ & CULM$_{12[\text{INTENT}]}$] **do’** (x, Ø) CAUSE [BECOME **know’** (x, y)]
  x = 1; y = 2; z = 3

This lexical entry codifies a causative accomplishment logical structure modified by a semantic representation of the following type: CAUS$_{123}$INSTR (say)$_{123}$ [understand]. This representation indicates that an agent (arg1) causes (CAUS) a mental percept (arg2) to be understood better by a receiver-beneficiary (arg3). The means (INSTR) by which this is achieved is by *say*ing (semantic primitive) things about it (arg2) to the receiver beneficiary (arg3).

(28)

- Misunderstand: [DEGRAD$_{12}$] **know'** (x, y)

Example (28) consists of a state logical structure modified by the lexical operator DEGRAD$_{12}$ [understand]. In *misunderstand,* a cognizer (arg1) does not achieve correct understanding of a mental percept (arg2). The function DEGRAD is thus applied to the basic meaning of the dimension to signify when an action is not performed or carried out well.

(29)

- Confuse: [CAUSDEGRAD$_{123]}$] **do'** (x, Ø) CAUSE [BECOME NOT **know'** (x, y)]
  x = 1; y = 2; z = 3

Example (29) is a causative accomplishment logical structure modified by a semantic chain which contains the following operators: CAUSDEGRAD$_{123}$ [understand]. In this lexical entry, *confuse,* an agent (arg1) causes (CAUS) a mental percept (arg2) not to be understood (DEGRAD) by a receiver-beneficiary (arg3).

(30)

- Puzzle: [INVOLVCAUSCONT(think)$_{123}$] **do'** (x, Ø) CAUSE [BECOME NOT **know'** (x, y)]
  x = 1; y = 2; z = 3

Example (30) designates a causative accomplishment logical structure and a semantic module of the following type: INVOLVCAUSCONT(think)$_{123}$ [confuse]. In *puzzle,* an agent (arg1) causes (CAUS) a mental percept (arg2) not to be understood (DEGRAD) by a receiver-beneficiary (arg3). This activity involves (INVOLV) causing (CAUS) the receiver to think (semantic primitive) for a long time (CONT).

(31)

- Confound and perplex

INVOLVCAUSESYMPT$_{123}$ (FEELING_TYPE: surprise, doubt) **do'** (x, Ø) CAUSE [BECOME NOT **know'** (x, y)] x = 1; y = 2; z = 3

In (31) there is a causative accomplishment logical structure modified by INVOLVCAUSESYMPT$_{123}$ (FEELING_TYPE: surprise, doubt) [understand]. In *confound* and *perplex,* an agent (arg1) causes (CAUS) a mental percept (arg2) not to be understood (DEGRAD) by a receiver-beneficiary (arg3). Both activities involve (INVOLV) the experiencing of feelings (SYMPT) on the part of the receiver (arg3). The difference between the two predicates is that in *confound* the FEELING_TYPE is surprise and doubt, whereas in *perplex*, the FEELING_TYPE is worry.

4.2.2. The temporal dimension

The temporal dimension of COGNITION encodes the fact that one can think about somebody or something from the past (*remember, recall*), in the present (*consider1, study*) or in the future (*consider2, plan*). These temporal subdimensions are represented by the function LOC. The subscript ($_{in}$) refers to position and the superscript (^{temp}) to the fact that LOC in this case refers to temporal location: The resulting function, $\text{LOC}_{in}^{\text{TEMP}}$ is followed by arrows that indicate if that temporal location is the past ($^{\leftarrow}$), present ($^{\leftrightarrow}$), or future ($^{\rightarrow}$).

think Arg_1 = Thinker Arg_2 = Mental percept		
$\text{LOC}_{in}^{\text{TEMP}}{}_{12}$ [think]		to think about sb/sth in time (past/present/future)
$\text{LOC}_{in}^{\text{TEMP}\leftarrow}{}_{12}$ [think]		to think about sb/sth in the past
$\text{LOC}_{in}^{\text{TEMP}\leftarrow}{}_{12}$ INSTR $(\text{CAUS}(\text{move}))_{12}$ LOC_{ad1} (BODY_PART: mind) [think]		**remember** to think about sb/sth in the past by causing it to move again to your mind.
	PLUSINTENT_{12} [remember]	**recall** to remember sb/sth making an effort to do so.
	$\text{INVOLV}(\text{say})_{12}\text{SYMPT}_1$ (FEELING_TYPE: pleasure) [remember]	**reminisce** to remember sb/sth, saying things about them with pleasure.
	DEGRAD_{12} [remember]	**forget** to not remember.
	CAUS_{12} [remember]	**remind** to cause sb to remember.
$\text{LOC}_{in}^{\text{TEMP}\leftrightarrow}{}_{12}$ [think]		to think about sth in the present
	$\text{LOC}_{in}^{\text{TEMP}\leftrightarrow}{}_{12}$ CONT [think]	**consider₁** to think carefully about sth
	PURP $\text{MAGN}(\text{know})_{12}$[consider1]	**study** to consider sth in order to know more things about it.
$\text{LOC}_{in}^{\text{TEMP}\rightarrow}{}_{12}$ [think]		to think about sth in the future
	CONTPURP $(\text{POSS}(\text{do}))_{12}$ $\text{LOC}_{in}^{\text{TEMP}\rightarrow}$ [think]	**consider₂** to think carefully about sth so that you might do in the future.
	$\text{PLUSINTENT}_{12}\text{PURP}(\text{do})_{12}$ $\text{LOC}_{in}^{\text{TEMP}\rightarrow}$ [think]	**plan** to think intentionally about sth in order to do it in the future.

The predicates in this subdomain can be represented as follows:

(32)

- Remember: [$\text{Loc}_{in}^{\text{temp}\leftarrow}{}_{12}$ Instr $(\text{Caus}(\text{move}))_{12}$ loc_{ad1} (body_part: mind)] **think’** (x, y)

Example (32) designates a state logical structure with *think* as a primitive and modified by two arguments. Besides, this logical structure is complemented and enriched by a semantic representation that has the following interpretation: a cognizer (arg1) thinks about a mental percept (arg2) located in the past ($\text{LOC}_{in}^{\text{TEMP}\leftarrow}$). The means (INSTR) used to do this is by causing (CAUS) the mental percept (arg2) to *move* (semantic primitive) to a location (LOC_{ad}) which is the mind (category BODY_PART) of the cognizer (arg1).

(33)

- Reminisce: [Involv(say)$_{12}$Sympt$_1$ (feeling_type: pleasure)] **think'** (x, y)

The predicate in (32) inherits the properties of *remember* and thus encodes a state logical structure, which is in turn modified by a semantic representation with the following format: INVOLV(say)$_{12}$SYMPT$_1$ (FEELING_TYPE: pleasure) [remember]. *Reminisce* has the same basic meaning as *remember,* but involves (INVOLV) a cognizer (arg1) saying (*say*) things about it (arg2). This subactivity produces physical symptoms or feelings (SYMPT) in the cognizer/experiencer, which come from FEELING_TYPE: pleasure.

(34)

- Consider: [LOC$_{in}$$^{TEMP\leftrightarrow}$$_{12}$ CONT] **think'** (x, y)

Example (34) includes a state logical structure and a semantic representation LOC$_{in}$$^{TEMP\leftrightarrow}$$_{12}$ CONT [think], which is interpreted as follows: a cognizer (arg1) thinks about a mental percept (arg2) located in the present *(*LOC$_{in}$$^{TEMP\leftrightarrow}$) for a long time (CONT).

(35)

- Study: [PurpMagn(know)$_{12}$ & Loc$_{in}$$^{temp\leftrightarrow}$$_{12}$ Cont] **think'** (x, y)

As in the previous examples, the state logical structure in (35) is enriched by a semantic representation PURPMAGN(know)$_{12}$ [consider1]. In *study,* a cognizer (arg1) thinks about a mental percept (arg2) located in the present. PURP codifies the purpose of the action, which is for the cognizer to *know* (semantic primitive) more (MAGN) about the mental percept (arg2).

(36)

- Consider: [ContPurp(Poss(do))$_{12}$ Loc$_{in}$$^{temp\rightarrow}$] **think'** (x, y)

The semantic representation in (36), CONTPURP(POSS(do))$_{12}$ LOC$_{in}$$^{TEMP\rightarrow}$ [think] has the following interpretation: a cognizer (arg1) thinks about a mental percept (arg2) located in the future (LOC$_{in}$$^{TEMP\rightarrow}$) for a long time (CONT). The purpose of the action (PURP) is to possibly (POSS) *do* (semantic primitive) it in the future (LOC$_{in}$$^{TEMP\rightarrow}$).

(37)

- Plan: [PlusIntent$_{12}$Purp(do)$_{12}$ Loc$_{in}$$^{temp\rightarrow}$] **think'** (x, y)

Apart from the state logical structure, the predicate in (37) includes a semantic description in the following terms: PLUSINTENT$_{12}$PURP(do)$_{12}$ LOC$_{in}$$^{TEMP\rightarrow}$ [think]. In *plan* the cognizer (arg1) thinks about a mental percept (arg2) with intentionality (PLUSINTENT). The purpose (PURP) of his action is to *do* (semantic primitive) it in the future (LOC$_{in}$$^{TEMP\rightarrow}$).

5. Predicting morphosyntactic patterns from semantic representations

Another very important issue here is whether morphosyntactic patterns can be obtained from the information in these lexical representations. In previous research we formulated lexical mapping rules to show how morphosyntactic structure could be derived from the information in the template. The question now is if this is also true for this new system of lexical representation.

In the representations so far, the semantics of the second argument (y) has not been specified. However, in the case that concerns us here, the semantic nature of the complement can be further narrowed down by using the operators and primitives of the metalanguage. Section 5.1 describes exactly how this can be done (cf. section 5.1.).

Section 5.2 explains how the set of functions can be related to the morphosyntactic patterns of each of the complements through the notion of lexical rules or semantic redundancy rules as first postulated in Van Valin and Wilkins (1993). These proposals signify a partial reformulation of the Semantic Hierarchy.

5.1. The semantics of the complement phrase

Certainly, most of the representations in the examples so far either encode the manner, the means, or the instrument used to carry out the action, in other words, additional semantic parameters that differentiate one predicate from others within the same lexical domain. However, nothing is said of the semantic nature of the complement, especially in those cases that involve complex constructions. If these representations are to be syntactically informative we will need to include, when necessary, a semantic description of the second argument, based based on the semantic metalanguage we have been using.

A relevant example can be found in the valorative subdomain of COGNITION. The valorative subdomain focuses on the truth value of the proposition entertained in the mind of the cognizer. The proposition is represented by the function FACT, which means something (in this case, an idea) that is realized. The cognizer evaluates the proposition in terms of its truth value. In this sense, we use the semantic primitive *true* (from the NSM inventory) and combine it with Mel'cuk's function ANTI when it is necessary to express the contrary.

The subdimensions are structured in terms of the certainty of the cognizer. Total certainty is to think that something is true (*believe*) or that it will happen (*expect*). The second dimension contains those verbs that encode the probability (PROB) or the possibility (POSS) that something is true or that it will happen. The last dimension contains verbs such as *doubt* that signify thinking that something is possibly (POSS) not true (ANTI true). The following are the most relevant predicates within the subdimension:

(38)

- Believe: [FACT$_{12}$(true)$_2$] **think'** (x, y)

Example (38) contains a state logical structure composed of a primitive *think* and two arguments. In addition, this structure is modified by a semantic component which specifies the semantic nature of the second argument, in other words, the proposition that the mental precept is *true* (semantic primitive).

(39)

- Persuade: [[Fact$_{12}$(true)$_2$ / FACT$_{12}$ (become)] Caus$_{123}$ Instr (say)$_{123}$] [**do'** (x, Ø)] cause [become **think'** (y, z)]

The lexical entry for *persuade* (39) includes a causative accomplishment logical structure which is in turn modified by a semantic component. Firstly, this component specifies the semantic nature of the second argument which is: (i) the expression of the effector's attitude or judgement regardint the truth of the proposition (this reading is in fact inherited from its superordinate *believe*); (ii) a mental disposition regarding a possible action. Besides, there is an instrumental (INSTR) parameter that indicates that the effector carries out this action by *say*ing (semantic primitive) something to y. As shown below, the two interpretations are very revealing of the type of morphosyntactic structure and syntactic clause linkage of the predicate and the complement.

(40)
- Expect: [FACT$_{12}$ (true)$_2$ / FACT $_{12}$ (become)$_2$] **think'** (x, y)

The lexical entry in (40) designates a state logical structure which is subject to two possible interpretations; (i) the effector's evaluation that the proposition will be true in the future; (ii) the effector's mental disposition that some action will occur in the future. Again, these two interpretations have important consequences for the type of morphosyntactic structure of the complements.

(41)
- Suspect: [Fact $_{12}$ (poss true)$_2$] **think'** (x, y)

In *suspect* (41) the cognizer (arg1) thinks about a mental percept (arg2), which is a proposition (FACT). This proposition is evaluated as possibly *true* (POSS true).

(42)
- Doubt: [Fact $_{12}$ (poss anti true)$_2$] **think'** (x, y)

In *doubt* (42) the cognizer (arg1) thinks about a mental percept (arg2), which is a proposition (FACT). This proposition is evaluated as possibly untrue (POSS ANTI true).

This is just a glimpse of how we can encode the semantics of the second argument by drawing on the units of the metalanguage. Now, let us see how we can codify these functions so that we can derive morphosyntactic patterns.

5.2. Lexical rules and morphosyntactic structure

As a first approximation, we have used the notion of lexical rule as first formulated in Van Valin and Wilkins (1993). Although we maintain that these rules provide a sound framework to predict morphosyntactic structure from semantic representation, some important modifications should be introduced which concern both the nature and the format of the lexical rule. Let us first look at the type of rules proposed in Van Valin and Wilkins (1993) so that we can see the differences with respect to the new proposal:

(43) The format of SRR:

something.be.in.mind…..

a) intention(s): **something.x.intends.be.in.mind**
b) knowledge: **something.x.knows.be.in.mind**
c) belief(s): **something.x.believes.be.in.mind**
d) perception: **something.x.perceived.be.in.mind**

One of the problems with these rules is that they are based on a number of semantic labels extracted from the Interclausal Semantic Hierarchy. We believe that this Semantic Hierarchy can be challenged for the following reasons:

- Semantic categories such as perception, cognition, etc. include a vast inventory of verbs, many of which do not conform to the syntactic patterns encoded in the hierarchy.

- The mapping between the syntactic and the semantic scales can be refined by giving an exact account of the lexical primitive or operator encoded within a lexical class.

For these reasons, the Interclausal Semantic Hierarchy presents semantic labels such as cognition, perception, jussive, etc. that are too vague and unspecific since not all the verbs that belong in these classes are subject to the types of clause linkages suggested in the Hierarchy. This is why it is much safer to bind lexical rules to specific operators (or lexical functions) or even primitives that are an integral part of the semantic representation of the predicate. In this way, we avoid using overly abstract and sometimes confusing labels.

Accordingly, the functions/operators and primitives should serve as input for the formulation of an inventory of semantic redundancy rules which account for the type of semantic and syntactic bond between the predicate and the complements. As for the format of the rules, we use a mechanism based on the syntax of the metalanguage:

(44) Operator [operandum] = value

Based on the research done so far, we can tentatively propose the following list of rules although we are aware that it can and must be completed by adding new functions after investigating the internal architecture of different lexical classes[14]:

(45)	FACT (true) [y] =	Clausal Subordination
	FACT (become) [y] =	Core cosubordination
	PURP do [y] =	Core cosubordination
	$\text{INVOLV}_{12\ (\text{ACTION})}$ [y] =	Core coordination
	$\text{INVOLV}_{12\ (\text{see})\ 12}$ [y] =	Core cosubordination
	INTENT_{2} [y] =	Clausal subordination

The first function can be broken down into three depending on the degree of truth (see below). This analysis leads to a new conception of the Semantic Hierarchy as presented in Van Valin and LaPolla (1997:479). Although this hierarchy is basically correct, we can reformulate part of it (the first seven categories which are the only ones that refer to lexical classes) by substituting the semantic labels for lexical functions. Here is the resulting scale[15]:

(46) Semantic and syntactic hierarchy based on lexical functions

CAUSE	Nuclear cosubdordination

[14] The inventory of semantic redundancy rules is far from being exhaustive. Here is a sample of those that we have found in our analysis of two lexical domains. We are sure that this inventory will be enriched when other lexical classes are analyzed.

[15] There is a clear correspondence between these functions and the semantic labels used in Van Valin and LaPolla (1997: 479): CAUSE corresponds to Causative; INCEP, CONT, and CESS stand for the three phases of aspectual verbs; FACT [become] corresponds to Psych-action verbs; PURP [do] corresponds to purposive verbs; INVOLV12 (ACTION) corresponds to Jussive verbs; INVOLV 12 [SEE] corresponds to direct perception; FACT and its various degrees of truth correspond to propositional attitude verbs; INTENT corresponds to cognition verbs.

INCEP	Core cosubordination
CONT	Core cosubordination
CESS	Core cosubordination
FACT [become]	Core cosubordination
PURP [do]	Core cosubordination
INVOLV12 (ACTION)	Core coordination
INVOLV 12 (SEE)	Core coordination
FACT (true)	Clausal subordination
FACT](PROB true)$_2$	Clausal subordination
FACT (POSS true)$_2$	Clausal subordination
FACT (anti true)$_2$	Clausal subordination
INTENT	Clausal subordination

This proposal is meant to be an instantiation of those broad semantic labels of the type, 'psych-action', direct perception' etc. We are not saying that these labels are incorrect, but they would certainly benefit from a clearer formulation and definition of their referential scope. That is why they have been replaced by lexical functions which have more precise reference.

Let us see how the whole proposal works for *learn,* which focalizes the inceptive phase of know:

(47) BECOME **know'** (x, y)

We have argued that this structure is sufficient as it stands since there are not any further semantic parameters intervening in the representation of this predicate. However, if we want to predict morphosyntactic structure from lexical entries, then it is necessary to enrich this semantic representation by specifying the semantic potential of the y argument:

(48) [involv12 fact$_{12}$ (become) / intent$_2$] BECOME **know'** (x, y)

This representation consists of a semantic module coded in brackets and an accomplishment logical structure. The semantic representation provides a semantic description of the nature of the second argument (subscript 2) which is subject to two possible interpretations: (i) there is a mental disposition on the part of the first argument to carry out an action; (ii) the second argument is an expression of knowledge. According to our semantic redundancy rules, the first reading results in a core co-subordination clause linkage while the second results in a clausal subordination. Adapting the proposal in Van Valin and Wilkins (1993) we can use the following diagram to illustrate the different semantic interpretations of the y argument and their corresponding morphosyntactic values according to the Semantic Hierarchy Scale:

(49) Decomposed semantic representations

Decomposed Semantic Representations

Semantic Redundancy Rules [Semantic subclasses filling variable slots]

$$
\begin{array}{cccl}
\alpha & \beta & \gamma & \text{Interpretations} \\
\downarrow & \downarrow & \downarrow & \\
\text{a} & \text{(b)} & \text{(c)} & \text{Interclausal Semantic Relations (IRH)} \\
\downarrow & \downarrow & \downarrow & \\
\text{j-n1} & \text{j-n2} & \text{j-n3} & \text{Syntactic Clause Linkage Categories (IRH)} \\
\downarrow & \downarrow & \downarrow & \\
\text{mscf1} & \text{mscf2} & \text{mscf3} & \text{Language-specific Morphosyntactic Complement Types}
\end{array}
$$

(50) *learn*

learn: [INVOLV12 FACT_{12} (BECOME) / INTENT_2] BECOME **know'** (x, y)

↓

Semantic redundancy rules for y

FACT (become) [y]	INTENT $_2$ [y]
↓	↓
FACT [become]	INTENT
↓	↓
Core cosubordination	Core/Clausal Subordination
↓	↓
Subjectless *to* + infinitive	*That*-Clause

Examples of this algorithm are the following instances[16]:

(51) Core co-subordination

In addition, over 11,000 have the opportunity to **learn** to play an orchestral instrument 12.499 c:\bnc\g\gx\gxj.dcv 16
Charles Darwin (1872) pointed out that: children **learning** to write often twist about their tongues as their fingers move, in a ridiculous fashion. 33.524 c:\bnc\f\fe\fed 87
According to Miller & Dollard (1941), "The follower (the student) **learns** to model his behaviour on that of another (the teacher) through responding to cues of sameness,

[16] These examples have been taken from the British National Corpus.

for which he is re 7.924 c:\bnc\h\h0\h0y 21
When a child is **learning** to read, he will point with his finger to the word on which he is concentrating. 30.050 c:\bnc\a\ay\ayk 78

(52) Clausal subordination

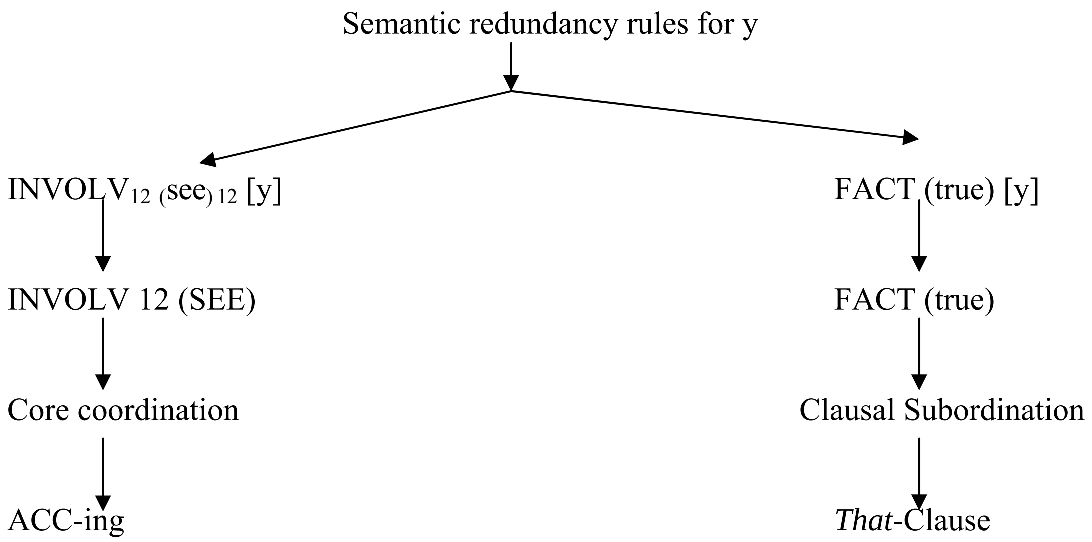
I think we **learned** then that there are times when it is best to play with some safety in 34.864c:\bnc\h\ha\hae 50
and we **learn** from the Word of God that He “made peace through the blood of his cross. 12.914 c:\bnc\b\b2\b29 48
It’s encouraging to **learn** that tourists come to farms to enjoy what is already there; they are interested in walking, visiting open farms and buying produce, rather than in specific entertainments. 40.405 c:\bnc\a\ac\acr 66
But Mount Kenya is over 17,000ft (5100m) and we’d also **learnt** that many walkers and climbers fail to reach the top simply because of altitude sickness.

The fifth subdomain of COGNITION is linked to visual perception because the cognizer sees the percept in his mind. A case in point is the predicate *imagine* with the following representation:

(53) *imagine*: [FACT $_{12}$ INVOLV$_{12}$(see)$_{12}$ LOC$_{in}$ (abstract body_part: mind)] **think’** (x, y)

The lexical entry in (53) consists of a state predicate defined by the primitive *think* and two arguments such that the first argument thinks about a mental percept (FACT). The semantic component specifies an activity that involves *see*ing (semantic primitive) it in his mind. This predicate entertains two possible interpretations of the second argument, i.e. as a mental proposition and as a direct percept. This can be seen in the following diagram:

(54) Semantic redundancy rules for y



6. The conceptual basis of the metalanguage

Since meaning is never clear-cut, many theoretical frameworks tend to tiptoe around it as though it were a sleeping crocodile. This is a paradoxical situation because every linguistic theory has some sort of dictionary either implicitly or explicitly at its base. In fact, one could even go so far as to say that the success of a theory depends on how well this underlying dictionary has been designed and elaborated[17].

According to Givón (1995:395), the conceptual lexicon is a repository of relatively time-stable, culturally-shared, well-coded knowledge about our external-physical, social-cultural and internal-mental universe. It is a network of interconnecting codes, in which the activation of one node activates other related nodes in what neurologists refer to as a spreading activation pattern. Such concepts are types of conventionalized experience, which can be divided into the categories of entities, events, properties, and relations. This assertion supports those who are in favor of establishing an ontology of concepts upon which to anchor language structure (Nirenburg and Raskin 2004).

Meaning definitions within the lexicon point to the position a concept may have within a network. This is in line with Pustejovsky (1996:6) when he writes:

> […], the meanings of words should somehow reflect the deeper conceptual structures in the cognitive system, and the domain it operates in. This is tantamount to stating that the semantics of natural language should be the image of nonlinguistic conceptual organizing principles, whatever their structure.

Such a conceptual configuration would be organized onomasiologically (in meaning areas) rather than semasiologically (in alphabetical order). The definitions in an ideal dictionary would thus reflect conceptual categories, codified in the genus or conceptual label of the definition, and differentiating features would be codified in the adverbial modification of the genus. Definitions would be not only coherent on a microstructural level, but also on a macrostructural one.

It is well-known that categories have a basically hierarchical organization, given that hierarchies are central to cognition. Jackendoff (1997:16) writes:

> From the point of view of psychology and neuroscience, of course, redundancy is expected. Moreover, so are multiple sources of infinite variability, each with hierarchical structure. One can understand an unlimited number of hierarchically organized visual scenes and conjure up an unlimited number of visual images; one can plan and carry out an action in an unlimited number of hierarchically organized ways; one can appreciate an unlimited number of hierarchically organized tunes.

Within such hierarchies concepts are related both vertically and horizontally by different types of conceptual relations. When this type of organization is applied to the structure of a semantic or conceptual domain, the resulting structure is an ontology (in the artificial intelligence sense rather than the philosophical sense) defined by Gruber (1993) as an explicit specification of a conceptualization. One crucial property of a conceptual system is that no concept can be described without an account of its relationships to various others (Lamb 1998: 147).

In this sense it is impossible to understand the domain of COGNITION without understanding its overall structure as well as that of other related domains. This can be seen in the fact that in English and other languages abstract concepts such as knowing and thinking are conceptualized in a more concrete form to facilitate understanding. For

---

[17] Nevertheless, when meanings are evoked within linguistic frameworks, they usually appear in the form of a phrase that has been hastily cut and pasted from the nearest available dictionary. Although no one can deny that a dictionary is a valuable lexical resource, dictionaries vary in quality, and at the present time there are none so perfect as to qualify as a source for the automatic extraction of conceptual meanings.

example, in order to better understand abstract thought, we turn ideas into objects that can be possessed (*grasp* [POSSESSION]) and looked at (*consider* [VISUAL PERCEPTION]).

The lexical units that we have used as examples also have evident secondary connections to the domains of SPEECH, FEELING, VISUAL PERCEPTION, MOVEMENT, and POSSESSION. For example, *reminisce* and *persuade* have connections to the domain of SPEECH; *anticipate, confound,* and *perplex* to the domain of FEELING: *imagine* and *plot* to VISUAL PERCEPTION; *remember* to MOVEMENT; and *interpret* and *grasp* to POSSESSION.

The multidimensional nature of this area of the lexicon can be seen in the multiple ways that the activity of thinking is conceptualized. How we think about thinking and knowing also has repercussions in how we are able to encode it linguistically, in other words, in the complementation patterns of predicates or its syntactic potential.

6. Conclusions

One of the greatest challenges in linguistics today is to find a kind of representation that will adequately transmit the interface between syntax and semantics. Although syntax is undoubtedly important, it hardly tells the whole story (or even the main part of the story). No linguistic framework can aspire to any sort of adequacy unless it takes a position on the conceptual meaning of lexical units and endeavors to show how this type of meaning is related to their syntactic potential.

In this regard we have taken a set of predicates within the lexical domain of COGNITION, one of the most complex in the lexicon, because language conceptualizes the process of thinking in different ways. This is evidenced by the various dimensions and subdimensions that these predicates fall into. Each lexical unit contains the meaning of a superordinate term. The generic term or terms in each domain are regarded as undefinables. Their meaning definitions reflect their relationships with each other as well as those with predicates of other domains. A semantic metalanguage is used for this codification, which is based on Wierzbicka's Natural Semantic Metalanguage, Mel'cuk's lexical functions, and the lexical organization of Martín Mingorance's Functional Lexematic Model.

The semantic primitives of Wierzbicka, though extensively researched have the disadvantage of creating natural language definitions that are extremely unwieldy. As a result, codification of this type is not viable if one's goal is conciseness. Mel'cuk's lexical functions, on the other hand, are extremely concise and elegant. If applied in conjunction with the NSM, the resulting codification benefits from the insights of both models.

Nevertheless, it must be pointed out that neither framework has ever tried to deal with paradigmatic lexical structure. In contrast, the FLM is based on lexical structure obtained by the factorization of dictionary entries as well as the convergence of both semantic and syntactic information.

As shown in this article, this type of semantic codification is compatible with RRG lexical structures and lexical templates. It uses a finite inventory of functions and primitives to encode lexical distinctions that despite having no significant impact on syntax, are fundamental to meaning.

7. References

Amsler, R.A. 1980. The structure of the Merriam-Webster pocket dictionary. *Technical Report* TR – 164. Austin: University of Texas.

Apresjan, J.D. 1993. Systemic lexicography as a basis of dictionary-making. *Journal of the Dictionary Society of North America* 14: 79-87.

Damasio, A. and H. Damasio 1992. Brain and language. *Scientific American* 267 (3): 63-71.

Damasio, H., T. Grabowski, D. Granel, R. Hichwa, A. Damasio 1996. A neural basis for lexical retrieval. *Nature* 380: 499-505.

Dancette, J. and M.C. L'Homme (2002). "The Gate to Knowledge in a Multilingual Specialized Dictionary: Using Lexical Functions for Taxonomic and Partitive Relations", In *EURALEX 2002 Proceedings*, August 13-17 2002, Copenhagen (Denmark), pp. 597-606.

Dummett, M. 1991. *Frege and other philosophers*. Oxford: Oxford University Press.

Faber, P. and R. Mairal Usón 1999. *Constructing a lexicon of English verbs.* Berlin: Mouton de Gruyter.

Fontenelle, T. 1997. Turning a bilingual dictionary into a lexical semantic database. Tübingen: Max Niemeyer.

Givon, T. 1995. *Functionalism and grammar.* Amsterdam: John Benjamins.

Goddard, C. 1998. *Semantic analysis: A practical introduction.* Oxford: Oxford University Press.

Goddard, C. and A. Wierzbicka 2002. *Meaning and Universal Grammar: Theory and Empirical Findings.* Amsterdam: John Benjamins.

Grimes, J. 1990. Inverse Lexical Functions, in: J. Steele (Ed.). *Meaning-Text Theory: Linguistics, Lexicography and Implications*, Ottawa: Ottawa University Press, pp.350-364.

Gruber, T. R. 1993: "A translation approach to portable ontology specifications". *Knowledge Acquisition* 5, 199-220.

Jackendoff, R. 1983. *Semantics and cognition.* Cambridge, MA: MIT Press.

Jackendoff, R. 1990. *Semantic structures.* Cambridge, MA: MIT Press.

Jackendoff, R. 1997: *The Architecture of the Language Faculty.* Cambridge MA: MIT Press.

Lakoff, G. 1987. *Women, fire, and dangerous things.* Chicago: Chicago University Press.

Lamb, S. M. 1998: *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam/Philadelphia: John Benjamins.

Levinson, S. 1997. From outer to inner space: Linguistic categories and non-linguistic thinking. In J. Nuyts and E. Pederson (eds.): 1997. *Language and conceptualization.* Cambridge: Cambridge University Press, 13-45.

Levy, D. K. 2003. Concepts, language, and privacy: An argument ''vaguely Viennese in provenance''. *Language and Cognitive Processes*, 2003, 18 (5/6), 693–723.

L'Homme, M.C. (2005, forthcoming). "Using Explanatory and Combinatorial Lexicology to Describe Terms", In Wanner. L. (ed.). *Selected Lexical and Grammatical Topics in the Meaning-Text Theory. In Honour of Igor Mel'cuk,* Amsterdam/Philadelphia: John Benjamins.

Mairal Usón, R. 2003. "Why the notion of lexical template". *Anglogermánica online* 2002-1.

Mairal Usón, R. 2004. "La reglas léxicas en la Gramática del Papel y la Referencia". In Wotjak, G. y Cuartero Otal, J.: *Algunos problemas específicos de la descripción sintáctico-semántica*. Berlín, Frank & Timme Verlag., pp. 175-196.

Mairal Usón, R. and M. J. Pérez Quintero (eds) 2002. *New perspectives on argument structure in Functional Grammar.* Berlin: Mouton de Gruyter.

Mairal, R. and P. Faber, 2002. Functional Grammar and lexical templates. In R. Mairal Usón and M. J. Perez Quintero (eds.): 39-94.

Martín Mingorance, L. 1984. “Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valency dictionary”. In Hartmann, Reinhart ed. *LEX’eter ‘83 Proceedings. Papers from the International Conference on Lexicography at Exeter*. Tübingen: Max Niemeyer, 226-236.

Martín Mingorance, L. 1990. “Functional Grammar and Lexematics”. In Tomaszczyk, Jerzy and Barbara Lewandowska-Tomaszczyk eds. 1990, 227-253.

Martín Mingorance, L. 1995. “Lexical logic and structural semantics: methodological underpinnings in the structuring of a lexical database for natural language processing”. In Hoinkes, Ulrich ed. 1995. *Panorama der Lexikalischen Semantik*. Tübingen: Gunter Narr, 461-474.

Mel’cuk, I. 1981. Meaning-Text Models: A recent trend in Soviet linguistics. *Annual Review of Anthropology* 10: 27-62.

Mel’cuk, I. 1988. Semantic description or lexical units in an explanatory combinatorial dictionary: Basic principles and heuristic criteria. *International Journal of Lexicography* 1 (3): 165-188.

Mel’cuk, I. 1989. Semantic primitives from the viewpoint of the Meaning-Text Linguistic Theory. *Quaderni di Semantica* 10 (1): 65-102.

Mel’cuk, I., N. Arbathewsky-Jumarie, L. Elnitsky, L. Iordanskaja, and A. Lessard 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I.* Montreal: University of Montreal Press.

Mel’cuk I., A. Clas et A. Polguère. 1995. *Introduction à la lexicologieexplicative et combinatoire*, Louvain-la-Neuve (Belgique), Duculot / Aupelf - UREF.

Mel’cuk and L. Wanner 1996. Lexical functions and lexical inheritance for Emotion lexemes in German. In: L. Wanner (ed.): *Recent trends in Meaning-Text Theory.* Amsterdam: John Benjamins, 209-227.

Nirenburg, S. and V. Raskin 2004. *Ontological semantics.* Cambridge, MA: MIT Press.

Pustejovsky, J. 1996. *The Generative Lexicon.* Cambridge MA: MIT Press.

Van Valin, R.D. 1993. *Advances in Role and Reference Grammar.* Amsterdam: John Benjamins.

Van Valin, R.D. 2003. On the relationship between syntactic theory and models of language processing. Available at: http://linguistics.buffalo.edu/research/rrg/vanvalin_papers/SynThLgProc.pdf

Van Valin, R.D. and D. Wilkins, 1993. Prediction syntax from semantics. In. R.D. Van Valin (ed.): 499-534.

Van Valin and R. LaPolla, 1997. *Syntax: Structure, meaning and function.* Cambridge: Cambridge University Press.

Wierzbicka, A. 1980. *Lingua mentalis.* Sydney/New York: Academic Press.

Wierzbicka A. 1987. *English speech act verbs: A semantic dictionary.* Sydney/New York: Academic Press.

Wierzbicka, A. 1996. *Semantics: Primes and universals.* Oxford: Oxford University Press

Appendix 1:

Grammatical category	Wierzbicka’s Semantic Primitives
nouns	I, YOU, SOMEONE, PEOPLE, SOMETHING/THING, BODY
determiners	THIS, THE SAME, OTHER
quantifiers	ONE, TWO, SOME, ALL, MANY/MUCH
evaluators	GOOD, BAD

descriptors	BIG, SMALL, (LONG)
intensifier	VERY
mental predicates	THINK, KNOW, WANT, FEEL, SEE, HEAR
speech	SAY, WORD, TRUE
actions, events and movement	DO, HAPPEN, MOVE
existence and possession	THERE IS, HAVE
life and death	LIVE, DIE
time	WHEN/TIME, NOW, BEFORE, AFTER, A LONG TIME, A SHORT TIME, FOR SOME TIME, MOMENT
space	WHERE/PLACE, HERE, ABOVE, BELOW; FAR, NEAR; SIDE, INSIDE; TOUCHING
“logical” concepts	NOT, MAYBE, CAN, BECAUSE, IF
augmentor:	MORE
taxonomy, partonomy	KIND OF, PART OF;
similarity	LIKE

Appendix 2: Functions used in this article

Lexical Function	**Definition**
ECL Lexical Functions (with their application adapted to paradigmatic structure)	
ANTI	Antonym. This LF also combines with other LFs to negate them.
BON	Good (expression of praise)
CAUS	Cause
CONT	Continuity/duration
CULM	The highest point of []
DEGRAD	To get worse
FACT	Be realized
INCEP	The beginning of []
INSTR	Instrument
INVOLV	Subactivities implied by the predicate
LOC_{ad}	Spatial location with directionality “to”
LOC_{in}	Spatial location with directionality “in”
LOC^{temp}	Temporal location which can have arrows marking past (←), present (↔) or future ().
MAGN	intense(ly), very [intensifier], to a very high degree
MINUS	less of []
OBSTR	to function with difficulty
PERM	permit
PLUS	more of
SYMPT	physical symptoms
Additional lexical functions used in this article	

INTENT	intentionality
POSS	possibility
PROB	probability
PURP	purpose