

Document downloaded from:

<http://hdl.handle.net/10251/45259>

This paper must be cited as:

Periñán Pascual, JC. (2012). The situated common-sense knowledge in FunGramKB. *Review of Cognitive Linguistics*. 10(1):184-214. doi:10.1075/rcl.10.1.06per.



The final publication is available at

<http://dx.doi.org/10.1075/rcl.10.1.06per>

Copyright John Benjamins Publishing

The situated common-sense knowledge in FunGramKB

Carlos Periñán-Pascual

Universidad Politécnica de Valencia (Spain)

Abstract

It has been widely demonstrated that expectation-based schemata, along the lines of Lakoff's propositional Idealized Cognitive Models, play a crucial role in text comprehension. Discourse inferences are grounded on the shared generalized knowledge which is activated from the situational model underlying the text surface dimension. From a cognitive-plausible and linguistic-aware approach to knowledge representation, FunGramKB stands out for being a dynamic repository of lexical, constructional and conceptual knowledge which contributes to simulate human-level reasoning. The objective of this paper is to present a script model as a carrier of the situated common-sense knowledge required to help knowledge engineers construct more "intelligent" natural language processing systems.

Keywords: FunGramKB, common-sense knowledge, situated knowledge, script, natural language understanding

1. Introduction

FunGramKB¹ is currently being developed as a multipurpose lexico-conceptual knowledge base which has been carefully designed to be implemented in natural language processing (NLP) systems, and more particularly for those which require the capability of language comprehension in artificial intelligence (AI) settings. Among the many strengths of FunGramKB, we emphasize two features: knowledge organization and conceptual representation. In other words, FunGramKB is not only provided with a

a highly-structured organization of the lexical, grammatical and conceptual modules, whose theoretical underpinnings conform to the research results in the fields of cognitive science and linguistics, but it also uses such an expressive conceptual representation language that the FunGramKB knowledge can be actually transduced into different computational formalisms to simulate human reasoning (e.g. logic, production rules, conceptual graphs, etc).

The aim of this paper is to describe the benefits of FunGramKB as an NLP knowledge base, particularly when processing the situated common-sense knowledge in the form of scripts. The remainder of the paper is structured into five sections. Section 2 portrays the three levels of knowledge which make up the architecture of this knowledge base. Section 3 focuses on the cognitive model upon which the conceptual modules are built, presenting as well those theories which influenced the model. Section 4 deals with the description of the most outstanding common-sense knowledge databases in computational lexicography and AI. Section 5 approaches the FunGramKB script model from the perspectives of time representation, structure, enhancement, methodology and edition. Finally, section 6 highlights the relevance of a knowledge base such as FunGramKB in the scenario of a cognitive architecture for an AI system.

2. An overview to FunGramKB

FunGramKB (Periñán-Pascual and Arcas-Túnez, 2004, 2005, 2007a, 2010a, 2010b; Mairal-Usón and Periñán-Pascual, 2009; Periñán-Pascual and Mairal-Usón, 2009, 2010) is a multipurpose lexico-conceptual knowledge base for natural language understanding systems. FunGramKB is multipurpose in the sense that it is both multifunctional and multilingual. In other words, FunGramKB has been designed to be reused in various

NLP tasks (e.g. information retrieval and extraction, machine translation, dialogue-based systems etc) and with several natural languages.²

FunGramKB comprises three knowledge levels, consisting of several independent but interrelated modules:

Lexical level:

- The Lexicon stores morphosyntactic information about lexical units. This lexical model preserves the major linguistic assumptions of Role and Reference Grammar (RRG) (Van Valin and LaPolla, 1997; Van Valin, 2005), i.e. logical structures and the linking algorithm, although some important contributions have been introduced with the aim of building a robust knowledge base.
- The Morphicon helps our system to handle cases of inflectional morphology.

Grammatical level:

- The Grammaticon stores the constructional schemata corresponding to the four layers identified in the Lexical Constructional Model (LCM) (Mairal-Usón and Ruiz de Mendoza, 2009; Ruiz de Mendoza and Mairal-Usón, 2008), which helps Role and Reference Grammar to construct the semantics-to-syntax linkage.

Conceptual level:

- The Ontology is presented as a hierarchical catalogue of the concepts that a person has in mind, so here is where semantic knowledge is stored in the form of meaning postulates. The Ontology consists of a general-purpose module (i.e. Core Ontology) and several domain-specific terminological modules (i.e. Satellite Ontologies).
- The Cognicon stores procedural knowledge by means of scripts, i.e. schemata in which a sequence of stereotypical actions is organised on the basis of temporal continuity, and more particularly on the basis of Allen's temporal model (1983).
- The Onomasticon stores information about named entities and events, such as

Barack Obama or 9/11. This module stores two different types of schemata (i.e. snapshots and stories), because instances can be portrayed synchronically or diachronically.

The main consequence of this three-level design is that every lexical or grammatical module is language-dependent, whereas every conceptual module is shared by all languages. In other words, computational linguists must develop one Lexicon, one Morphicon and one Grammaticon for English, one Lexicon, one Morphicon and one Grammaticon for Spanish and so on, but knowledge engineers build just one Ontology, one Cognicon and one Onomasticon to process any language input conceptually. In this scenario, FunGramKB adopts a conceptualist approach, since the Ontology becomes the pivotal module for the whole architecture.

The development of multifunctional and multilingual resources such as FunGramKB heavily influences knowledge base design. With regard to multifunctionality, the current trend in many language engineering projects consists in developing *ad hoc* resources for a particular NLP application in a given domain. This *modus operandi* leads to great efficiency in knowledge representation, but the main drawback is the lack of flexible portability to other domains or tasks because of the inability to meet the new requirements of other applications (Lenci, 2000). Since building a large-scale NLP knowledge base is costly in time and effort, it is eagerly recommended to design reusable and updatable resources, so that they can be easily maintained or improved in different projects along the time (Floridi, 1999). Thus, multifunctional knowledge bases should integrate any information potentially relevant to any NLP task. However, the type of knowledge NLP systems require closely depends on the purpose of the applications themselves. For instance, spell checkers require very little lexical information. On the contrary, text understanding systems usually need to

process morphological, syntactic, semantic and pragmatic information of lexical units, as well as non-linguistic knowledge from the world model (Nirenburg and Raskin, 2004). Therefore, the most reasonable strategy to implement a multifunctional knowledge base is to make it conform to the requirements of the NLP task but letting other systems access additional information if required. On the other hand, with regard to multilinguality, Aguado de Cea et alii (2007) presented a thorough analysis of current strategies in knowledge-based systems. Among the many models reviewed, FunGramKB matches the model of knowledge representation in which links between the ontology and language resources are set. More particularly, the FunGramKB Lexicon is developed for every language handled by the NLP system, where the shared Ontology is able to relate lexical units from different languages.

3. The FunGramKB cognitive model

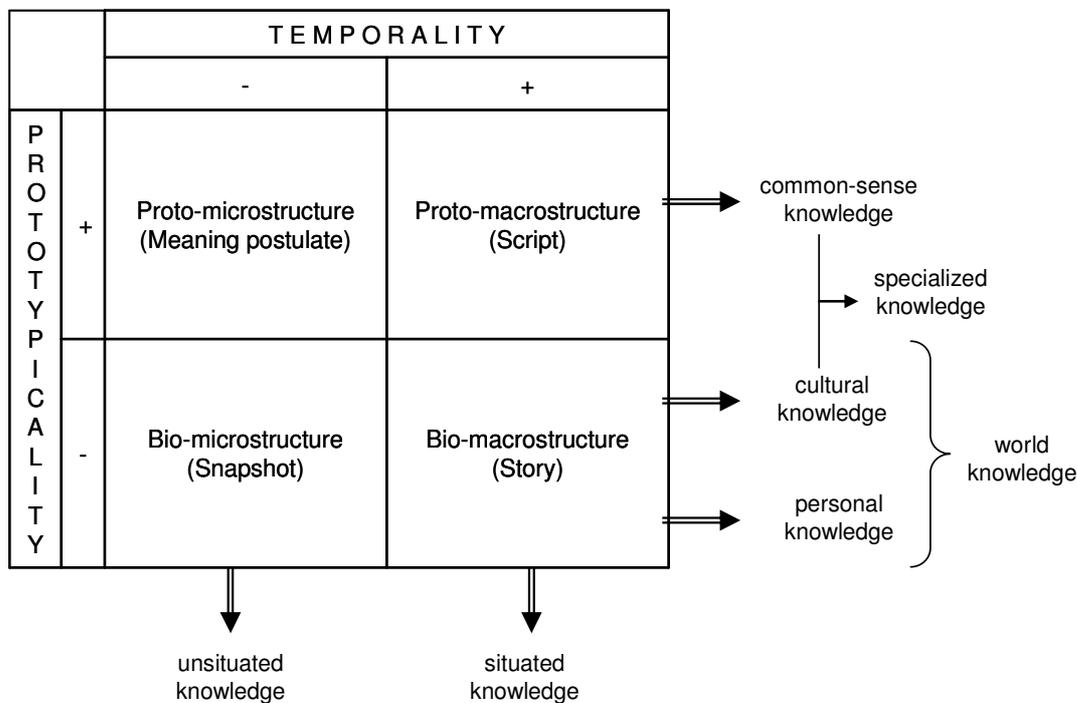
The model of “schema” originated in fields such as psychology, linguistics and artificial intelligence —resulting in a myriad of conceptual constructs such as “schema” (Barlett, 1932; Rumelhart, 1975), "frame" (Minsky, 1975), "script" (Schank and Abelson, 1977), "mental model" (Johnson-Laird, 1983), or "idealized cognitive model" (Lakoff, 1987), among others— has had a significant impact on the representation of conceptual knowledge in FunGramKB. According to the schema theory, a schema is a mental representation of a given type of entity or event, usually consisting of a set of expectations which are developed as more and more memories of similar experiences are accumulated. Typically, schemata contain generalized knowledge from our past experiences, facilitating thus the inference of information from our perception of the world. Future experiences are interpreted on the basis of the patterns constructed from past experiences, so these patterns serve to alleviate the cognitive overload. The

psychological status of mental schemata, i.e. that human beings process linguistic exchanges with respect to a set of expectations stored in our minds, is grounded on the fact that, for example, wrong predictions usually arise from the mismatch between perceptual information and the standardized model of the world. For instance, if we hear the statement *Mary usually goes to high school at 8:30*, we could infer that *Mary* is a student, according to the standardized knowledge about what type of people usually go to schools. If we later hear that *Last weekend Mary corrected a lot of exams*, then we will infer with a higher degree of certainty that she is a teacher. Moreover, mental schemata can sometimes distort our memories. For example, works such as Barlett (1932) and Brewer and Treyens (1981) demonstrated that human memories (e.g. stories which have been read, or images which have been seen) can be shaped by unawaresly replacing unusual components by more common ones in order to fit those memories into the standard model of the world. However, mental schemata turn out to be very useful as a way to explain how fragmented information can be retrieved, since these schemata supply the missing information to reconstruct the situation model of linguistic utterances.

FunGramKB conceptual schemata play a key role to draw inferences along the process of text understanding. In our knowledge base, schemata are classified according to two parameters: prototypicality and temporality. On the one hand, conceptual representations can store prototypical knowledge (i.e. proto-structures) or can serve to describe instances of entities or events (i.e. bio-structures). For example, the description of the meaning of *film* involves the construction of the proto-structure of its corresponding concept; however, if we want to provide information about the film *Jurassic Park*, then we should do it through a bio-structure. Therefore, whereas proto-structures help to model common-sense knowledge, bio-structures help to construct the

model of the world. On the other hand, knowledge within conceptual schemata can be presented atemporally (i.e. microstructures) or in a temporal framework (i.e. macrostructures). For example, the biography of Steven Spielberg requires a macrostructure, but a microstructure is sufficient to describe the profession of film director. In this respect, the availability of a spatio-temporal framework makes macrostructures be developed from a situated approach to conceptualization. As shown in Figure 1, the convergence of the values of these two parameters results in a typology of four different conceptual schemata which shape the FunGramKB conceptual level.

Figure 1. Typology of conceptual schemata in FunGramKB.



As shown in Figure 1, a critical issue in conceptual representation is to determine the influence of situational context upon the description of cognitive categories. Typically, meaning is thought to be a finite set of semantic features applicable to every context in which the lexical unit is used. According to this view, the conceptual representation is not linked to a given situation. Instead, the representation is the product of abstraction over different background contexts. Thus, I can simply define

CHAIR with the features FURNITURE, TO SIT ON and WITH BACK SUPPORT as a result of all the experiences I have had with different instances of this type of entity in the real world, and due to the perceived lack of contextual commonality, I can ignore other features referring to the number of legs, material, colour etc.

Although most systems adopt this unsituated approach to concepts, Barsalou (1982) devised a model to describe lexical meaning through (i) context-independent properties, which are activated in all background situations by the word(s) linked to the concept,³ and (ii) context-dependent properties, which are activated just by the relevant situations in which the words are encoded. Due to the remarkable impact of context-dependent properties on cognitive processing, Barsalou (2002, 2003) coined the term “situated conceptualization”, since people conceptualize the same cognitive category in a different way with respect to the active situations, being each conceptualization relevant just to the corresponding background situation. This model, which has been widely demonstrated in empirical studies,⁴ involves that conceptual knowledge should be arranged in a series of situations with the aim to facilitate processing. For example, background situations play a critical role in language comprehension. We use models of situation to represent text meanings: once the situation is identified, we focus on the most likely entities and events, which ultimately constitute the relevant knowledge in the current situation. In fact, knowledge organization around background situations makes cognitive processing more tractable. Thus, instead of scanning all categories in memory, the system points at the most relevant entities and events for the situation. In this fashion, it is easier to retrieve relevant information in tasks such as text comprehension, problem solving or reasoning, among many others (Yeh and Barsalou, 2006).

In this respect, FunGramKB organizes the temporality dimension of knowledge in such a way that it roughly corresponds to Barsalou's distinction (1985, 1991) between taxonomic and goal-derived categories. Whereas taxonomic categories are provided with hierarchically-organized context-independent representations through an ontology model (i.e. unsituated knowledge), goal-derived categories are conceptualized by means of representations which take into account the background situation (i.e. situated knowledge). Barsalou (1991) demonstrated that, in the case of both taxonomic and goal-derived categories, the speakers of a given community share prototype structures which are stable over time, playing a key role in knowledge representation and processing. Therefore, FunGramKB implements the approach to situated conceptualization through goal-derived categories, which take the form of macrostructures similar to Barsalou's view to mental models, i.e. schemata which must produce "quasi-continuous simulations of events", simulating its successive states over space and time (Barsalou 1992, p. 164).

On the whole, conceptualization in FunGramKB is actually in consonance with Lakoff's theory (1987: 284-287) on propositional Idealized Cognitive Models (ICM), i.e. those conceptual configurations which are not based on "imaginative devices" such as metaphor and metonymy. The various structures in which propositional ICMs can take place are present in FunGramKB, where the main difference lies in the capability of the latter to fully integrate the various types of schema. For example, the predicate-argument construct of propositional ICMs take the form of the FunGramKB thematic frames, which are used in turn to build the meaning postulates, being able to represent both feature bundles and radial categories linked to taxonomic categories in the Ontology. Finally, Lakoff's scenarios, which typically describe goal-oriented activities structured into an initial state, a sequence of events and a final state, correspond to the

FunGramKB scripts. Taking into account the situatedness criterion, however, FunGramKB schemata better correlates to the low-level cognitive models discussed in Ruiz de Mendoza (2007), i.e. “non-generic semantic configurations created by making well-entrenched, coherent links between elements of our encyclopedic knowledge store”. More particularly, our frame-like constructs building up the common-sense knowledge correspond to low-level non-situational models in the case of meaning postulates, and to low-level situational models in the case of scripts.

On the other hand, and as shown in Figure 1, FunGramKB stores four different types of knowledge according to their degree of prototypicality: common-sense, specialized, cultural and personal. These types of knowledge are represented through the same formal language, i.e. COREL (Conceptual Representation Language), so information sharing takes place more directly and effectively. Since the FunGramKB conceptual modules (i.e. Ontology, Cognicon and Onomasticon) use COREL as the “common language” for schema construction, natural language understanding systems will only require one “common reasoner”. Indeed, we are currently developing an automated cognizer with human-like defeasible reasoning powers which will be able to make inferences and draw conclusions from meaning postulates, scripts, snapshots and stories. Following the mainstream in psychology, FunGramKB adopts a mentalist approach with respect to schemata, so concepts are deemed to be as “structured mental representations”, or, according to the objectivist view to ICMs in cognitive semantics, our schemata mainly consist of “mental entities, not real things” (Lakoff, 1987: 285) which are linked by a set of relations.⁵ In line with “the language of thought hypothesis” (Fodor, 1975), COREL has a syntax and a semantics to describe formally these mental representations (cf. Perrián-Pascual and Arcas-Túnez, 2004). The FunGramKB schemata are actually modelled through propositional representations. Although there is

no empirical evidence about the device transducing language expressions into propositional representations, nor researchers agree about the abstract and arbitrary symbols used to construct these representations, predications have sufficient expressive power so as to access cognitive states directly, as well as having great potential to be implemented computationally (Barsalou et alii, 1993).

The remainder of this section serves to illustrate the different types of knowledge involved in FunGramKB. For instance, most researchers agree that common-sense knowledge represents a system of pre-scientific beliefs upon which our everyday cognitive activities rest (Smith, 1995), or in lay terms, “the knowledge that every person assumes his neighbors also possess” (Panton et alii, 2006, p. 1), as illustrated by (1):⁶

(1) Ice is frozen water.

+ (e1: past +FREEZE_00 (x1)Theme (x2: +WATER_00)Referent (f1: +ICE_00)Result)

Specialized knowledge consists of beliefs on scholarly, scientific or technical domains which are accepted to be true by the competent members of an expert community, as can be seen in (2):

(2) A batholith is a large geological formation, made of granite, quartz monzonite, or diorite.

+ (e1: +BE_00 (x1: \$BATHOLITH_00)Theme (x2: +GEOLOGICAL_FORMATION_00)Referent)
+ (e2: +BE_01 (x1)Theme (x3: +BIG_01)Attribute)

+(e3: +BE_01 (x1)Theme (x4: +GRANITE_00 | +QUARTZ_MONZONITE_00
| +DIORITE_00)Attribute)

Cultural knowledge consists of factual information about the past, present or future model of the world. This knowledge can be shared to a higher or lesser degree, depending on the individuals' education, as shown in (3):

(3) Paris is the capital of France.

*(e1: +BE_00 (x1: %PARIS_00)Theme (x2: \$CAPITAL_00)Referent)
+(e2: +BE_02 (x1)Theme (x3: %FRANCE_00)Location)

Finally, personal knowledge, whose beliefs are accepted to be true by the person who holds them, is typically not shared by other people, unless, for example, relatives or friends are told about it (Van Dijk, 2003). An example is shown in (4).

(4) My neighbour Tim is a bit hypochondriac.

*(e1: +BE_00 (x1: %TIM_00)Theme (x2: +NEIGHBOUR_00)Referent)
*(e2: +BE_01 (x1)Theme (x3: p \$HYPOCHONDRIAC_00)Attribute)

As said in the previous section, the Ontology plays a centroidal role in the conceptual processing within FunGramKB, where any kind of schema can resort to ontological units for the construction of common-sense, world-model or specialized knowledge (Figure 2).

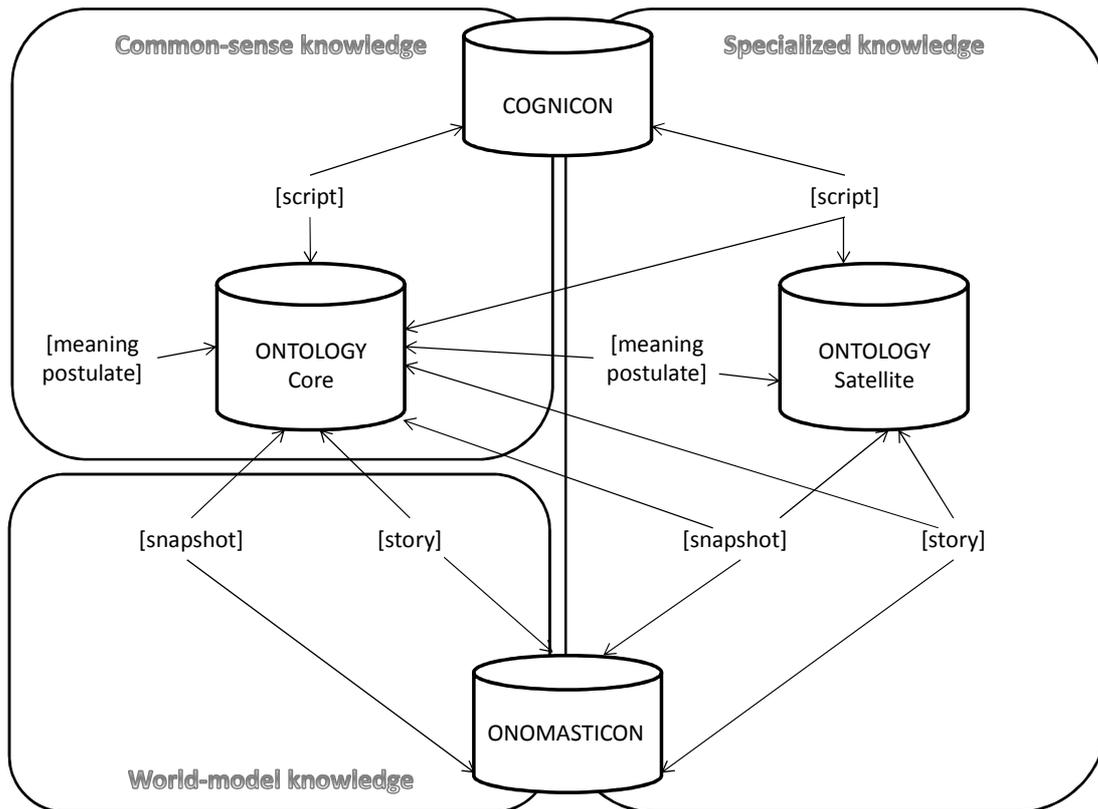


Figure 2. Conceptual modules and schemata.

4. Common-sense knowledge databases

The lack of a full-fledged machine-tractable repository of common-sense knowledge becomes the Achilles' heel of many NLP systems, a problem which steadily grows in AI. This section pinpoints the weaknesses of the most remarkable large-scale common-sense knowledge bases at present: a computational-lexicography product (i.e. FrameNet) and two AI-oriented knowledge repositories (i.e. CyC and ConceptNet).

4.1 FrameNet

The FrameNet project (Baker et alii, 1998; Ruppenhofer et alii, 2006), which is built upon the theory of Frame Semantics (Fillmore, 1976, 1982, 1985; Fillmore and Atkins, 1992), is intended to construct a lexical database where word senses are linked to hand-crafted semantic frames, thus becoming the notational devices for meaning description.

In other words, the semantic frame is a schematic arrangement of the frame elements which describe the scenario underlying the meanings of semantically-related words, e.g. the *Theft* frame consists of the core frame elements GOODS, PERPETRATOR, SOURCE and VICTIM, together with the peripheral frame elements MEANS, TIME, MANNER and PLACE. In addition, the frame elements in this collection of semantic frames are used to annotate corpus-extracted sentences manually. In this way, it is possible to retrieve automatically the inventory of syntactic patterns in which lexical units are involved. As a result, the lexical database has over 10,000 lexical units linked to about 800 hierarchically-related semantic frames, exemplified by means of over 130,000 annotated sentences.

The main drawback of the Frame Semantics model lies not only in the syntax-semantics interface, as explained in Jiménez-Briones and Luzondo-Oyón (s.d.), but also in the deceptively-deep approach to knowledge representation. Velardi et alii (1991) distinguished two well-defined strategies when describing meaning in computational lexicography: the cognitive content of a lexical unit can be described by means of semantic features or primitives (conceptual meaning), or through associations with other lexical units in the lexicon (relational meaning). Strictly speaking, the latter does not provide a real definition of the lexical unit, but it describes its usage in the language via “meaning relations” with other lexical units. It is certainly easier to state associations among lexical units in the way of meaning relations than formally describing the cognitive content of lexical units, but the inference power of conceptual meaning is undoubtedly stronger. According to this distinction, FrameNet is certainly “a large lexical databank which provides deep semantics” (Fillmore et alii, 2001), providing clear advantages over relational lexical databases such as WordNet (cf. Boas, 2005). The controversy arises when the description of meaning takes place in the conceptual

realm but it is restricted to a list of roles (e.g. frame elements) which work as binary semantic relations. In other words, between the poles of a deep approach (e.g. FunGramKB) and a surface approach (e.g. WordNet) to knowledge representation, a “shallow” approach implies that the cognitive content of a lexical unit is described by means of a simple feature-value matrix of conceptual relations (e.g. FrameNet). Consequently, surface and shallow models of natural language understanding are not sufficient for constructing efficient cognitive-based systems, since their expressive power is dramatically restricted (cf. Perrián-Pascual and Arcas-Túnez, 2007b). For instance, if you take into account the verb *forgive*, its *Forgiveness* frame, whose elements are EVALUÉE, JUDGE and OFFENSE, cannot fully state the meaning of the verb: “you stop being [1] angry [2] with someone you blamed [3] him or her”. That is, neither the semantic frame can represent aspectuality [1] or temporality [3], nor the frame elements can be conceptually qualified [2].

Moreover, although frame elements are deemed to be fine-grained roles, FrameNet ignores the *differentiae* of many verbs by lumping them together under the same semantic frame, resulting in coarse-grained meaning representations. For example, lexical units such as *steal*, *shoplift* and *snatch* are all linked to the *Theft* frame, so it fails to handle the Location and Manner *differentiae* in the meanings of *shoplift* and *snatch* respectively. Indeed, FrameNet researchers possibly opted for this excessive granularity in semantic roles in order to compensate for the deficiencies in this shallow model of lexical meaning. To this must be added that, with regard to the typology of conceptual schemata, FrameNet is restricted to the representation of unsituated common-sense knowledge, further impoverishing the language comprehension process.

4.2 CyC and ConceptNet

Two of the most prominent common-sense knowledge implementations for AI systems are CyC and ConceptNet. CyC (Lenat et alii, 1990; Lenat, 1995), whose name derives from “enCyClopedia”, includes about three million manually-constructed sentences which make over 300,000 concepts be interrelated. The inference engine is applied to these sentences following the first-order-predicate calculus with some extensions to handle second-order features. On the other hand, ConceptNet (Liu and Singh, 2004a, 2004b; Singh and Barry, 2003) is a semantic network of common-sense knowledge that contains 1.6 million edges connecting more than 300,000 nodes. The nodes represent concepts, which take the form of semi-structured English fragments (p.ej. “get to bed early”, “wake up in morning” or “eat breakfast”), and the edges represent the semantic relations linking the concepts (e.g. SubeventOf, UsedFor or EffectOf), where inferences are not based on logical deduction but on graph-based reasoning. ConceptNet is populated through the Open Mind Common Sense web site (Singh et alii, 2002), where the general public enter English sentences about common-sense knowledge. The semantic network is automatically built by applying NLP techniques and extraction rules to the semi-structured sentences in the corpus.

It is noteworthy to point out that both resources also present some weaknesses. With respect to CyC, many of the problems are related to the logical approach to reasoning. Mathematical logic is precise, allowing a systematic way to evaluate, maintain and prove the truth of statements. However, logical reasoning is not flexible enough to simulate human thinking. Logic is effective when applied to a simplified scenario, or “microworld”. However, nobody has still managed to face successfully the task of building a logical system with a realistically large set of propositions. Therefore, we can make logic systems work in simple tasks, but as these tasks become more realistic, many problems arise. For example, logical reasoning is deductive, so it has

much difficulty to formalize induction. Moreover, logical notation is extremely rigorous, so it has much difficulty to model the imprecise way that human categorize (Liu and Singh, 2004a). Another key problem of CyC is that concerning knowledge organization, whose model is not well-structured:

The distribution of CYC's half-million axioms in "knowledge space" showed no discernible pattern that would allow me to construct such a suite, short of simply picking particular axioms out of CYC's database and carefully phrasing questions around those axioms. (Pratt, 1994)

Apart from being poorly evaluated in terms of accessibility, the semantic encoding in CyC should also have been more specific, especially taking into account that the knowledge base has been “carefully” handcrafted. As Parmar (2001) noted, for example, CyC often lacks axioms that capture the effects of actions, so the representation does not support automated reasoning about change.

On the other hand, ConceptNet’s associative networks don’t require complete formal semantics (Liu and Singh, 2004a), but such a practical reasoning is not so expressive and precise as the CyC logical inferences, not to mention that noisy data resulting from mass collaboration inevitably spoils common-sense reasoning. However, the main weakness lies in the use of English as the language to represent common-sense knowledge. This lack of a formal language implies that it is not only more difficult to be precise, but natural language fragments may also be full of ambiguity.

Taking into account the FunGramKB cognitive model portrayed in section 3, together with the enhanced version of script as described in section 5, it can be

concluded that the deficiencies of FrameNet, CyC and ConceptNet can be largely overcome.

5. The FunGramKB scripts

Scripts have become a very popular schema to capture the overall meaning of texts. Indeed, the foundational work on knowledge-based systems for text understanding is traced back to Schank and Abelson (1977), illustrated with the well-known “restaurant” script. However, as these researchers recognized (Schank and Abelson, 1977: 61-62), not all scripts are of the same kind. Firstly, situational scripts reflect institutionalized behavior patterns of events as a result of people’s experiences on social interactions, e.g. eating out at restaurants. Secondly, instrumental scripts remind us to “operating manuals”, that is, they describe how to do things, e.g. driving a car. Finally, personal scripts only exist in the mental space of the agent, i.e. they describe situated knowledge which is not shared by other social participants. In the FunGramKB architecture, personal scripts do not belong to the Cognicon, but they take the form of stories in the Onomasticon, since only situational and instrumental scripts can embed an expected sequence of prototypical (or stereotypical) events.

5.1 Time representation

One of the crucial problems of script modelling is time representation within the sequence of events. In fact, from a psycholinguistic approach, an efficient organization of memory for event processing should involve the representation of temporal relations, so that the number of predictions can be dramatically reduced. In NLP, there are some systems which need to deal with temporality, what has been coined as “temporally-aware NLP”. In such cases, temporal reasoning allows the system to answer queries

involving the temporal relations between the events described in texts, e.g. whether a particular event preceded another one or how long the event lasted.

At present, the two most remarkable models to represent temporal information are temporal logic and the interval temporal model. On the one hand, logical models provide a robust way to model the dynamic world. Under the umbrella of temporal logic, the best-known formalisms are the Situation Calculus (McCarthy and Hayes, 1969) and the Event Calculus (Kowalski and Sergot, 1986). Whereas the Situation Calculus has proved to best fit hypothetical reasoning, the Event Calculus best addresses some time-related problems such as the frame problem, the specification of exact times and the description of simultaneous events (Goudey, 2007). However, similarities between the two models are such that there has been a lot of research to reconcile both formalisms (cf. Kowalski and Sadri, 1994; Van Belleghem et alii, 1995).

On the other hand, Allen's interval temporal model (1983) is based on a representation of time as a partially-ordered graph⁷ where nodes represent events and arcs are tagged with one or more relations of temporal ordering. An event E is treated as an interval consisting of a pair of time points (i, t) , i.e. the start time-point (i) and the end time-point (t). For example, supposing that an event occurs in the interval $E1(i_1, t_1)$ and another event occurs in the interval $E2(i_2, t_2)$, the interval relation $\text{Before}(E1, E2)$ is subject to the constraint $t_1 < i_2$. Some other interval relations from Allen's model are presented in Table 1.

	<u>Interval relations</u>	<u>Constraints</u>
1.	Before(E1, E2)	$(t_1 < i_2)$
2.	Meets(E1, E2)	$(i_2 = t_1)$
3.	Overlaps(E1, E2)	$(i_1 < i_2) \& (i_2 < t_1) \& (t_1 < t_2)$
4.	Starts(E1, E2)	$(i_1 = i_2) \& (t_1 < t_2)$
5.	During(E1, E2)	$(i_2 < i_1) \& (t_1 < t_2)$
6.	Finishes(E1, E2)	$(i_2 < i_1) \& (t_1 = t_2)$
7.	Equals(E1, E2)	$(i_1 = i_2) \& (t_1 = t_2)$

Table 1. Interval relations in Allen's temporal model.

Moreover, Allen devised an interval-based constraint propagation algorithm which computes all temporal relations taking place when a new event is added to the graph. Following the previous example, if event E3 is added, and E3 occurs during E2, then the system automatically infers that E1 is before E3.

Both the Event Calculus and the interval temporal model are suitable for common-sense reasoning in general and for natural language understanding in particular, because they allow NLP systems to perform a wide range of reasoning tasks, such as projection (or prediction) and explanation (or postdiction).⁸ However, the interval temporal model is far more expressive and natural than most temporal logic approaches used in AI. For this reason, we adopted Allen's model for the temporal framework of the FunGramKB scripts.

5.2 Structure

The FunGramKB script is structured into one or more predications within a linear temporal framework, where every predication is represented by a node in the graph. To

illustrate, the first eleven predications in the script @EATING_AT_RESTAURANTS_00 are presented in (5), together with the English equivalent in (6).

- (5) *(e1: +ENTER_00 (x1)Agent (x2: +CUSTOMER_00)Theme (x3)Location (x4)Origin (x5: +RESTAURANT_00)Goal (f1: (e2: +BE_01 (x2)Theme (x6: +HUNGRY_00)Attribute))Reason)
 *(e3: +TAKE_01 (x7:+WAITER_00)Agent (x2)Theme (x5)Location (x8)Origin (x9: +TABLE_00)Goal)
 *((e4: +SIT_00 (x2)Theme (x10: +CHAIR_00)Location)(e5: +BE_02 (x10)Theme (x9)Location (f2: m +NEAR_00)Position))
 *(e6: +TAKE_01 (x7)Agent (x11: \$MENU_00 | \$WINE_LIST_00)Theme (x5)Location (x12)Origin (x9)Goal)
 *(e7: +GIVE_00 (x7)Agent (x11)Theme (x13)Origin (x2)Goal)
 *(e8: +SEE_00 (x2)Theme (x11)Referent)
 *(e9: +CHOOSE_00 (x2)Theme (x14: +MEAL_00 & +BEVERAGE_00)Referent (f3: x11)Instrument)
 *(e10: +REQUEST_01 (x2)Theme (x14)Referent (x7)Goal)
 *(e11: +MOVE_00 (x15)Agent (x7)Theme (x5)Location (x9)Origin (x16: +KITCHEN_00)Goal)

- (6) A customer goes into a restaurant because he/she is hungry. A waiter goes with the customer to a table and the customer sits on a chair near the table. The waiter takes the menu and the wine list and gives them to the customer. The customer

sees the menu and the wine list, chooses the meal and the beverage, and places the order. The waiter goes to the kitchen.

Continuing with the same example above, the temporal relations between predications are stated in (7), which serves to generate automatically the graph in Figure 3.

- (7) e1 -> e3 [Before]
e3 -> e4 [Before ^ Meets]
e4 -> e6 [Before]
e6 -> e7 [Meets]
e7 -> e8 [Before ^ Meets]
e9 -> e8 [During]
e9 -> e10 [Before ^ Meets ^ Overlaps ^ Starts ^ Equals]
e10 -> e11 [Before ^ Meets]

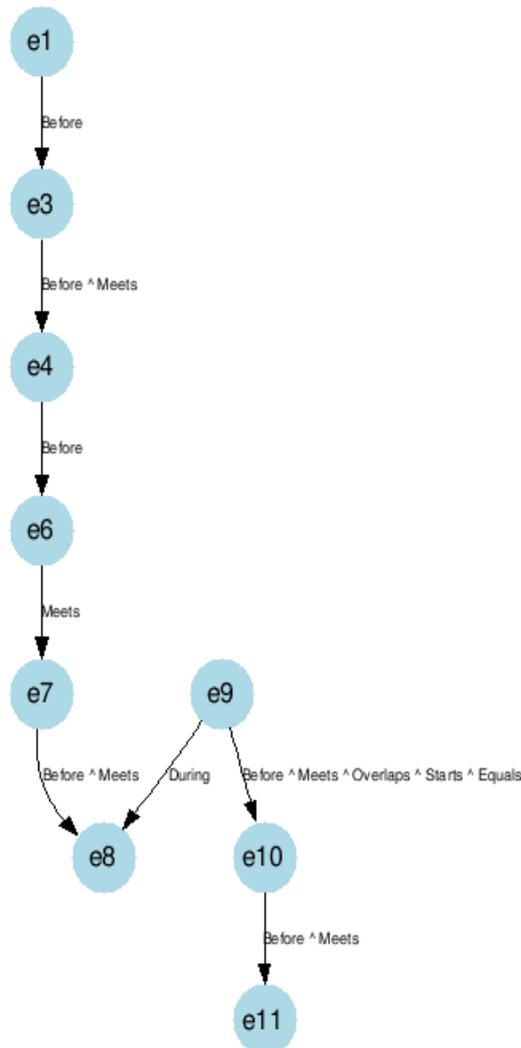


Figure 3. Temporal-knowledge representation in the FunGramKB scripts.

In the FunGramKB scripts, the nodes in the propositional networks can also represent script activators. For instance, in the last seven predications in @EATING_AT_RESTAURANTS_00, which are presented in (8), the predication (e30) can trigger either @WORKING_WITH_CASH_00 or @WORKING_WITH_CREDIT_CARD_00.⁹

- (8) *(e28: +PAY_00 (x2)Agent (x29: +MONEY_00 ^
 \$CREDIT_CARD_00)Theme (x30)Origin (x7)Goal)
 *(e29: +GIVE_00 (x7)Agent (x29)Theme (x31)Origin (x32: \$TILL_D_00)Goal)

*(e30: @WORKING_WITH_CASH_00 (x1: x32, x11: x7; x33: x9, x33: x10) ^
@WORKING_WITH_CREDIT_CARD_00 (x1: x32, f1: x7; x33: x11, x33:
x12))

*(e31: +GIVE_00 (x7)Agent (x33)Theme (x34)Origin (x2)Goal)

*(e32: +STORE_00 (x2)Theme (x33)Referent)

*(e33: +STAND_00 (x2)Theme (x35)Location)

*(e34: +LEAVE_00 (x2)Agent (x2)Theme (x5)Location (x10)Origin (x36:
+STREET_00)Goal)

Therefore, the scenario described by @EATING_AT_RESTAURANTS_00 can call another script describing the operation involved in the method of payment, which is also embedded, for example, in the scripts @GOING_SHOPPING_00 or @GOING_TO_THE_CINEMA_00, among many others. In this case, @EATING_AT_RESTAURANTS_00, @GOING_SHOPPING_00, and @GOING_TO_THE_CINEMA_00 become the host scripts for @WORKING_WITH_CASH_00, whose role is that of a guest script. Moreover, as shown in predication (e30) in (8), a single predication can include several script activators, but only one can be ultimately triggered in the conceptual processing, since all the scripts involved should be connected by the disjunctive logical operator.

A script activator takes the form of (9), as illustrated in (10):

(9) <script ID> (<outcoming mapping> ; <incoming mapping>)

(10) @WORKING_WITH_CASH_00 (x1: x32, x11: x7; x33: x9, x33: x10)

In other words, calling a script is not simply to retrieve more knowledge, but to have it properly integrated. Thus, the higher the number of participants shared between the host and the guest scripts, the higher the degree of conceptual integration. For example, the script @WORKING_WITH_CASH_00 is shown in (11).

- (11) *(e1: +OPEN_01 (x1: \$TILL_D_00)Agent (x2: \$TILL_00)Theme (x3)Location
(x4)Origin (x5)Goal)
*(e2: +ENTER_00 (x1)Agent (x6: +MONEY_00)Theme (x7)Location
(x8)Origin (x2)Goal)
*(e3: +LEAVE_00 (x1)Agent (x9: \$CHANGE_01)Theme (x7)Location
(x2)Origin (x8)Goal)
*(e4: +WRITE_00 (x2)Theme (x10: +RECEIPT_00)Referent)
*(e5: +GIVE_00 (x1)Agent (x9)Theme (x8)Origin (x11: +HUMAN_00)Goal)
*(e6: +GIVE_00 (x1)Agent (x10)Theme (x2)Origin (x11)Goal)
*(e7: +CLOSE_00 (x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal)

By “outcoming mapping” we mean the conceptual propagation from some participants in the host script to their corresponding participants in the guest script, whereas “incoming mapping” refers to the reverse process. Therefore, in the case of the script activator (10), the mappings between the host script @EATING_AT_RESTAURANTS_00 (5) and the guest script @WORKING WITH CASH_00 (11) are presented in Table 2.

	@GOING_TO_RESTAURANTS_00		@WORKING_WITH_CASH_00
\$TILL_D_00 (<i>cashier</i>)	x32	→	x1
+WAITER_00 (<i>waiter</i>)	x7	→	x11
\$CHANGE_01 (<i>change</i>)	x33	←	x9
+RECEIPT_00 (<i>receipt</i>)	x33	←	x10

Table 2. Mappings between host and guest scripts.

Therefore, the system can infer, for example, that the cashier typically returns any change with the receipt to the waiter, i.e. the predications (e5) and (e6) in the script (11), who in turn gives them to the customer, i.e. the predication (e31) in the script (8). As already explained, the modular architecture of FunGramKB supports the cognitive burden on a single conceptual language, so the various types of knowledge are able to become fully integrated into a whole network. Thus, the conceptual propagation of semantic knowledge into procedural knowledge allows the system to infer from the meaning of \$TILL_00 that the money customers pay at the restaurant is typically kept in the till, i.e. the predication (e2) in the meaning postulate (12).

- (12) +(e1: +BE_00 (x1: \$TILL_00)Theme (x2: +CALCULATOR_00)Referent)
+(e2: +STORE_00 (x1)Theme (x3: +MONEY_00)Referent (f1: +SHOP_00 ^
+RESTAURANT_00)Location)

5.3 Enhancement

The expectation-based model of script formulated by Schank (1975) and Schank and Abelson (1977) deeply influenced the theoretical foundation of the FunGramKB

Cognicon. However, we depart from this theory in three critical issues: organization, culturality and linguistic awareness.

Firstly, unlike Schank's dynamic memory (1982), where scripts are hierarchically organized in scenes, MOPs or TOPS, the Cognicon stores the scripts separately but interconnected through their internal script activators, resulting in a mesh-like graph of scripts. For example, all the scripts presented in the previous section can model the graph shown in Figure 4, where source nodes play the role of guest scripts and target nodes serve as host scripts.



Figure 4. A script-based graph.

Therefore, the role of script activators is two-fold: (i) keeping scripts interrelated, and (ii) enhancing the cognitive economy.

Secondly, FunGramKB is also intended to manage “cultural distinctiveness”, which is commonly found in procedural knowledge, e.g. social protocols. The possibility of calling a whole script within another script gives us the chance to introduce culturally-biased knowledge, since every script is assigned a geographical feature determining the continent, country, etc where that knowledge is typically true.¹⁰

Thus, in many respects, the shared knowledge of people in Western Europe, North America and South America transcends individual countries and different languages. If necessary, one could make a further distinction between different kinds of “national knowledge” as it is learned and presupposed by public discourse in different countries. (Van Dijk, 2003, p. 24)

Thirdly, Schank’s model of language comprehension is grounded on a formalism called Conceptual Dependency (Schank 1972), i.e. graphs which represent conceptually an input text on the basis of the semantic decomposition of verbs according to Fillmore’s Case Grammar (1968) and a set of eleven primitive ACTs (e.g. MOVE, PROPEL, ATRANS, PTRANS etc). The Conceptual Dependency theory downplayed the role of syntax and concentrated almost entirely on the construction of semantic representations. As demonstrated in the English Language Interpreter (Riesbeck 1975, Riesbeck y Schank 1978), the syntactic parsing took place only when it was required by the semantic analysis.

On the contrary, although COREL schemata can also be computationally considered as a combination of conceptual graphs and frames, the FunGramKB lexical and grammatical levels are grounded in sound linguistic theories, allowing the system to capture syntactic-semantic generalizations which are able to provide both explanations and predictions of language phenomena. This linguistic-aware approach, which is an uncommon practice in current NLP systems, is supported on the theoretical frameworks of RRG and LCM. On the one hand, RRG (Van Valin and LaPolla, 1997; Van Valin, 2005), one of the most relevant functional models on the linguistic scene

today, adopts a communication-and-cognition view of language, i.e. morphosyntactic structures and grammatical rules should be explained in relation to their semantic and communicative functions. RRG is a monostratal theory, since the semantic and the syntactic components are directly mapped without the intervention of abstract syntactic representations. Thus, the semantic and the syntactic components are directly mapped in terms of a linking algorithm, which includes a set of rules that account for the syntax-semantics interface. On the other hand, the LCM (Ruiz de Mendoza and Mairal-Usón, 2008; Mairal-Usón and Ruiz de Mendoza, 2009) has been developed in the last few years as a model which accounts for all those aspects involved in meaning construction, including those that go beyond core grammar, i.e. pragmatic implicature, illocutionary force and discourse coherence. Unlike some other existing linguistic models, the LCM is meant to provide a comprehensive description of meaning, which will serve as input for the syntactic apparatus, where the LCM follows the RRG linking algorithm in the sense that theoretical notions such as macroroles, privileged syntactic arguments etc are used for the description of the syntactic phase.

The clear-cut division between the cognitive and the linguistic levels becomes one of the central methodological axioms in FunGramKB. As a result, this distinction demands two different metalanguages, i.e. Conceptual Logical Structure (CLS) and COREL schema. The CLS, which is able to account for a wide range of linguistic phenomena within the RRG framework, serves as the pivot language between the text and the COREL representation, whereas the COREL schema, which provides the background knowledge from the FunGramKB conceptual modules, serves as the pivot language between the CLS and the automated reasoner. The following figure illustrates the relationship between syntactic and conceptual representations:

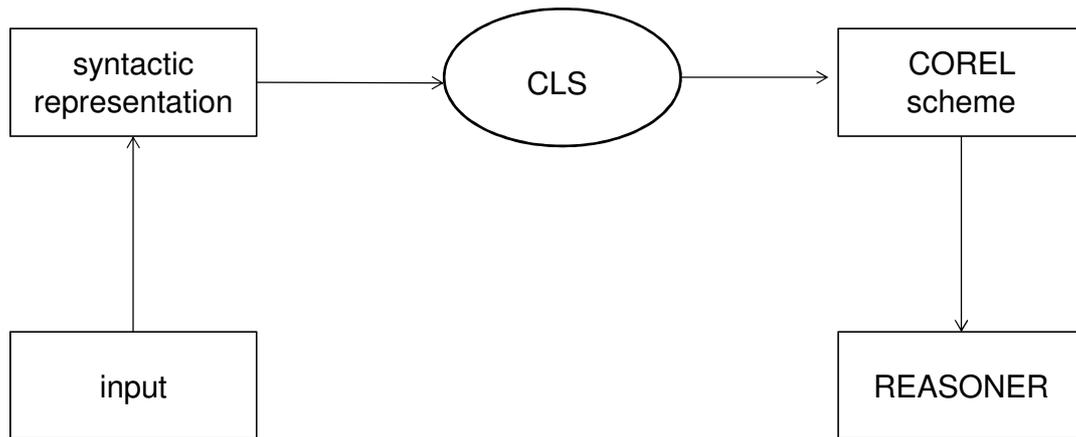


Figure 5. The syntactic-conceptual linkage in FunGramKB.

For example, the sentence *The bottle floated into the cave* can be automatically converted into the CLS (13), a task which requires the Core Grammar of the Lexicon, the RRG linking algorithm, and the conceptual knowledge stored in the Ontology.

(13) <_{IF} DECL <_{TNS} PAST < [do (+BOTTLE_00-Theme [+FLOAT_00 (+BOTTLE_00-Theme)])] (\$CAVE_00-Goal)>>>

In order to perform some reasoning with the input, the CLS (13) should be transduced into the COREL representation (14), so that it can be enriched by common-sense and world-model knowledge.

(14) +(e1: past +MOVE_00 (x1)Agent (x2: +BOTTLE_00)Theme (x3)Location (x4)Origin (x5: \$CAVE_00)Goal (f1: (e2: +FLOAT_00 (x2)Theme (x3)Location)))Manner)

As can be noted, the caused-motion construction involved in the input actively takes part in the modelling of the COREL schema by embedding the manner in which the event is performed. Therefore, both the Lexicon and the Grammaticon, together

with the Ontology and other conceptual modules (e.g. the Cognicon), help the NLP system to reconstruct the situation model underlying an utterance.

5.4 Methodology

When the FunGramKB scripts are being modelled, it is necessary to determine, first of all, the participants which are typically involved in the situation being described. These participants include entities such as people, objects and places which will take part in a time sequence of events. Afterwards, participants are organized along the time sequence of actions and situations, which are then represented by means of predications. Therefore, as in meaning postulates, the building blocks of scripts are predications made up of ontological units. As Newtonson (1976) demonstrated in episodic situations, the main qualitative changes in the configuration of individuals represent “breakpoints” in the sequence of events perceived by people. Consequently, since scripts are goal-derived categories, the introduction of a new predication typically involves that some entity changes the state of affairs with the aim to progress towards the goal. To build this type of knowledge, we often rely on “perceptual simulations” (Barsalou, 1999), i.e. when people describe the properties of a situation, they simulate an instance taking the form of a prototype, scan it mentally and identify the perceived properties. By representing the situation mentally, it is as if people “were there” with one of the instances; in other words, people create a modal simulation where they interact with the instance (Barsalou, 2002). Therefore, cognitive researchers fall back on their introspection through perceptual simulations to discover non-specialized procedural knowledge, because there is no text repository, such as dictionaries or corpora, from where this type of knowledge can be extracted.¹¹ At first sight, it could seem that the characteristics of the knowledge stored in the FunGramKB scripts are rather

idiosyncratic, closely connected with the knowledge engineer's previous experiences. However, Barsalou (1981, 1985, 1991) demonstrated that prototype structures generated for goal-derived situations are so stable between individuals as those generated for taxonomic categories, although in the former typicality is influenced not only by central tendency but mainly by ideals, i.e. features which should be included in the representation of a category in order to serve its purpose optimally.

5.5 Editor

FunGramKB is provided with a set of user-friendly tools to browse, edit and check the knowledge stored in the Cognicon. On the one hand, researchers can browse and edit the feature-value information about the Cognicon elements, e.g. ID, COREL script, temporal relations, domain and culture. On the other hand, a script validator checks that only consistent well-formed constructs are stored. More particularly, we developed lexical, syntactic and semantic parsers to examine several aspects of the scripts automatically. Firstly, the lexical analysis can verify the names of tokens by using deterministic finite automata represented by regular expressions. Secondly, the syntactic analysis checks that a sequence of various tokens is valid. For example, events must be followed by a parenthetical list of arguments together with their thematic roles. This parser is based on a stack automaton and a production-rule grammar (or recursive transition network). Thirdly, the semantic analysis identifies all those errors which are neither lexical nor syntactic. For instance, it has been programmed to check that in the script:

- every predication is headed by an event followed by its correct thematic frame, which is stored in the Ontology,

- the entities, events and qualities in the predications must be semantically defined in the Ontology, and
- the script activators include references to guest scripts which are stored in the Cognicon, as well as verifying that the mapping of participants is possible.

6. Cognitive architectures and knowledge bases

From the initial steps of our research project (cf. Perrián-Pascual and Arcas-Túnez, 2004, 2005), one of the main objectives was the development of a machine-tractable model of conceptualization which could simulate human-level reasoning. In such a scenario, the natural language understanding system should comprise the knowledge base together with a cognitive architecture, e.g. ACT-R (Anderson, 1993; Anderson and Lebiere, 1998), Soar (Laird, Rosenbloom and Newell, 1986; Newell 1990), Icarus (Shapiro and Langley, 1999) or Prodigy (Carbonell, Knoblock and Minton, 1990; Minton, 1990), just to mention a few. Cognitive architectures serve to determine the underlying infrastructure for the intelligent system, i.e. the cognitive mechanisms which are constant over time and across different applications. The typical components in cognitive architectures are the following (Langley et alii, 2009):

- (i) short- and long-term memories that store content about the agent's beliefs, goals and knowledge,
- (ii) a language that represents the memory contents in the form of cognitive constructs, and
- (iii) functional processes that operate on these constructs.

However, the psychological adequacy of intelligent agents has lately become a thorny issue, since most current studies on the development of natural language

understanding systems are based on the “cognitive black box” approach (Ball, 2006). In other words, computational mechanisms are applied to imitate human behaviour, but there are few attempts to model the insights of human language processing. Although early AI practitioners showed great interest in cognitive science research in order to build psychological-plausible models (cf. Minsky, 1975; Schank and Abelson, 1977), in the last two decades the AI community has been ignoring cognitive science and linguistics, an attitude motivated by three main factors (Langley, 2006):

- (i) Faster computer processors and larger memories have made possible the application of some complex statistical methods which require large amounts of computation.
- (ii) Mathematical tractability becomes a key concern, resulting in studies taking a more formal logic-oriented approach. The work focuses on problems that researchers can handle analytically, rather than on problems that people tackle heuristically.
- (iii) The commercial relevance of AI technologies has led many academics to focus on narrowly-defined problems, where acceptable results can be guaranteed, rather than on time-consuming complex tasks.

However, if AI research keeps being dominated by the goal of developing more efficient algorithms and formal theories, it will be difficult to achieve cognitive-plausible AI systems, so it is unlikely that this kind of systems can imitate human behaviour in such a complex domain as language processing (Ball, 2006). Therefore, the goal of simulating human-level intelligence should be achieved by implementing natural language systems within a cognitive architecture which supports the following capabilities (Langley et alii, 2009):¹²

- (a) Recognizing environmental situations as instances of known patterns.

- (b) Categorizing the objects and events involved in the situations.
- (c) Making decisions and choosing among alternatives.
- (d) Perceiving the world through different modalities.
- (e) Assessing the situation with respect to a model of the current environment.
- (f) Predicting the effects that actions have on the environment.
- (g) Monitoring the environment by relating sensing to prediction.
- (h) Solving problems to achieve the agent's goals.
- (i) Generating plans to achieve the goals.
- (j) Reasoning to draw conclusions from the agent's beliefs.
- (k) Deciding whether to maintain existing beliefs.
- (l) Interacting with other agents to obtain knowledge.
- (m) Communicating knowledge with a spoken or written language.
- (n) Remembering cognitive structures which describe external situations.
- (o) Reflecting on the agent's inferences, plans, decisions or actions.
- (p) Learning from experience.

As noted above, the knowledge base, including the agent's beliefs, goals and common-sense knowledge, should not be treated as part of the cognitive architecture, since working memory contents can change over time. The problem lies in the fact that most researchers are mainly concerned with the execution process of the cognitive architecture, but the conceptual knowledge base is equally important to cognition. Cognitive architectures use knowledge in the form of categories, but "they often relegate them to opaque symbols, rather than representing their meaning explicitly" (Langley et alii, 2009, p. 151). In this respect, a knowledge base such as FunGramKB, which is the product of research in the fields of cognitive science, linguistics and AI,

can allow the construction of enhanced computer-based systems for language comprehension.

7. Conclusions

This paper has demonstrated that it is feasible to design and develop an NLP knowledge base grounded on a psychological-adequate and linguistic-aware model to simulate human-level conceptualization. The parameters of prototypicality and temporality, along with the dimension of situatedness, serve to shape FunGramKB's cognitive space, facilitating the integration of a wide range of knowledge, such as common-sense, specialized, cultural and personal knowledge. Moreover, cognitive modeling by means of a common language of conceptual representation enables to fit more effectively common-sense knowledge into the situation model underlying the input text.

Unlike most NLP knowledge bases, FunGramKB reconciles perfectly taxonomic categories with goal-derived scenarios, also taking into account other structures pertaining to Lakoff's propositional ICMs. Therefore, FunGramKB becomes a robust knowledge base for a cognitive architecture to be implemented as an intelligent system, since it can store the agent's beliefs and goals as well as a deep-semantic representation of the common-sense knowledge. In fact, FunGramKB shouldn't be regarded simply as a knowledge repository, but rather as an integrated environment to model knowledge, where the lexical and grammatical levels rely on the theoretical apparatus of the RRG and the LCM, and the conceptual level takes a renewed approach to the long-standing schemata in cognitive science.

Acknowledgement

Financial support for this research has been provided by the DGI, Spanish Ministry of Education and Science, grants FFI2011-29798-C02-01 and FFI2010-15983. The research has been co-financed through FEDER funds.

Notes

¹ www.fungramkb.com

² English and Spanish are fully supported in the current model of FunGramKB, but we have also devised the lexical modules for languages such as German, French, Italian, Bulgarian and Catalan.

³ Barsalou and Billman (1989) called these context-independent properties as the “core attributes” of the frame, providing the underpinning to attribute systematicity across all contexts.

⁴ The cognitive plausibility of this approach has been demonstrated in many empirical experiments (cf. Yeh and Barsalou, 2006). For example, Wu and Barsalou (2009) showed that, although you ask people to generate the intrinsic properties (i.e. taxonomic and entity properties) of some objects, they tend to produce many other properties which describe background situations (i.e. situation properties) and subjective perspectives from these situations (i.e. introspective properties). Therefore, objects are usually not represented in an isolated way, but they are mentally located in their surrounding situations.

⁵ On the contrary, many philosophers adopt the Fregean approach to meaning, where concepts are dealt as abstract entities (cf. Peacocke, 1992; Rey, 1994).

⁶ The examples in this section are presented as COREL-formatted predications, together with their English equivalents.

⁷ A partially-ordered graph is a directed graph where the indegree and outdegree of each node (i.e. the number of edges entering and exiting from the node) can be greater than one.

⁸ In other words, for a given sequence of events, the system is able to determine the most likely results (i.e. projection) and the most likely reasons (i.e. explanation).

⁹ In fact, both `@WORKING_WITH_CASH_00` and `@WORKING_WITH_CREDIT_CARD_00` are embedded as script activators within the script `@WORKING_AS_CASHIER_00`.

¹⁰ The “default” value of the geographical feature states that a given script is universally applicable. For the issue of the universality and cultural distinctiveness in FunGramKB, we refer the reader to Periñán-Pascual and Mairal-Usón (fc).

¹¹ In order not to rely solely on introspection, García Garrido (2010) put forward a preliminary proposal to objectify the construction of low-level situational models by means of searches into repositories of knowledge, such as dictionaries or encyclopedias, and elicitation tests with native speakers.

¹² As noted by Langley (2006), cognitive architectures do not necessarily guarantee the construction of cognitively-plausible systems. In this regard, Soar is rather exceptional, since you can develop advanced AI systems for modelling and simulation, command and control, or robotics, among many other fields (e.g. Hanford and Long, 2011; Taylor et alii, 2011), whereas those intelligent agents are supported on sound cognitive theories in language comprehension (cf. Newell, 1990).

References

- Aguado de Cea, G., Montiel Ponsoda, E. & Ramos Gargantilla, J.A. (2007). Multilingualidad en una aplicación basada en el conocimiento. *Procesamiento del Lenguaje Natural*, 38, 77-97.
- Allen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832-843.
- Anderson, J.R. (1993). *Rules of the Mind*. Hillsdale: Lawrence Erlbaum.
- Anderson, J.R. & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah: Lawrence Erlbaum.
- Baker, C.F., Fillmore, C. & Lowe, J.B. (1998). The Berkeley FrameNet project. In *Proceedings of the COLING-ACL (pp. 86-90)*. Montreal.
- Ball, J.T. (2006). Can NLP systems be a cognitive black box? (Is cognitive science relevant to AI problems?). In *Proceedings of the AAAI Spring Symposium: Cognitive Science Principles Meet AI-hard Problems (pp. 1-6)*. Menlo Park: AAAI Press.
- Bartlett, F. (1932). *Remembering*. Cambridge: Cambridge University Press.
- Barsalou, L.W. (1981). *Determinants of Graded Structure in Categories*. Doctoral dissertation, Stanford University.
- Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, 10(1), 82-93.
- Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 629-654.
- Barsalou, L.W. (1991). Deriving categories to achieve goals. In G. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 27 (pp. 1-64)*. San Diego: Academic Press.

- Barsalou, L.W. (1992). *Cognitive Psychology: An Overview for Cognitive Scientists*. New York-London: Psychology Press.
- Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Barsalou, L.W. (2002). Being there conceptually: simulating categories in preparation for situated action. In N.L. Stein, P.J. Bauer & M. Rabinowitz (Eds.), *Representation, Memory and Development: Essays in Honor of Jean Mandler* (pp. 1-15). Mahwah: Lawrence Erlbaum.
- Barsalou, L.W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18, 513-562.
- Barsalou, L.W. & Billman, D. (1989). Systematicity and semantic ambiguity. In D. Gorfein (Ed.), *Resolving semantic ambiguity* (pp. 146-203). New York: Springer.
- Barsalou, L.W., Yeh, W., Luka, B.J., Olseth, K.L., Mix, K.S. & Wu, L.L. (1993). Concepts and meaning. In K. Beals, G. Cooke, D. Kathman, K.E. McCullough, S. Kita & D. Testen (Eds.), *Chicago Linguistics Society 29: Papers from the Parasession on Conceptual Representations* (pp. 23-61). Chicago: Chicago Linguistics Society.
- Boas, H.C. (2005). From theory to practice: Frame Semantics and the design of FrameNet. In S. Langer & D. Schnorbusch (Eds.), *Semantik im Lexikon* (pp.129-160). Tübingen: Gunter Narr.
- Brewer, W.F. & Treyens, J.C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, 13, 207-230.

- Carbonell, J.G., Knoblock, C.A. & Minton, S. (1990). PRODIGY: an integrated architecture for planning and learning. In K. VanLehn (Ed.), *Architectures for Intelligence* (pp. 241-278). Hillsdale: Lawrence Erlbaum.
- Fillmore, C.J. (1968). The case for case. In E. Bach & R.T. Harms (Ed.), *Universals in Linguistic Theory* (pp. 1-88). New York: Holt, Rinehart, and Winston.
- Fillmore, C.J. (1976). Frame Semantics and the nature of language. In *Conference on the Origin and Development of Language and Speech, volume 280 of Annals of the New York Academy of Sciences* (pp. 20-32). New York: New York Academy of Sciences.
- Fillmore, C.J. (1982). Frame Semantics. In Linguistic Society of Korea (Ed.) *Linguistics in the Morning Calm: Selected Papers from SICOL-1981* (pp. 111-137). Seoul: Hanshin.
- Fillmore, C.J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6, 222-254.
- Fillmore, C.J. & Atkins, B.T.S. (1992). Towards a frame-based lexicon: the semantics of RISK and its neighbours. In A. Lehrer & E.F. Kittay (Eds.) *Frames, Fields and Contrasts*. Hillsdale: Lawrence Erlbaum.
- Fillmore, C.J., Wooters, C. & Baker, C.F. (2001). Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Hong Kong.
- Floridi, L. (1999). *Philosophy and Computing: An Introduction*. London-New York: Routledge.
- Fodor, J.A. (1975). *The Language of Thought*. Cambridge: Harvard University Press.

- Garrido García, N. (2010). *Low-level Situational Cognitive Models within the Lexical Constructional Model and their Computational Implementation in a Knowledge Base*. Master thesis. Universidad de la Rioja.
- Goudey, B. (2007). A comparison of situation calculus and event calculus. In: http://www.cs.mu.oz.au/481/2007_projects/bwgoudey.pdf
- Hanford, S. & Long, L. (2011). A Soar agent for searching indoor environments using a mobile robot. In *Proceedings of the 2011 Conference on Behavior Representation in Modeling and Simulation (pp. 264-271)*. Simulation Interoperability Standards Organization.
- Jiménez-Briones, R. & Luzondo-Oyón, A. (s.d.). FrameNet and FunGramKB: a comparative account of computational resources for word meaning representation in Natural Language Processing.
- Johnson-Laird, P. N. (1983). *Mental Models*. Cambridge: Harvard University Press.
- Kowalski, R. & Sadri, F. (1994). The situation calculus and event calculus compared. In *Proceedings of the 1994 International Symposium on Logic programming (pp. 539–553)*. Cambridge: MIT Press.
- Kowalski, R.A. & Sergot, M.J. (1986). A logic-based calculus of events. *New Generation Computing*, 4, 67-95.
- Laird, J. E., Rosenbloom, P. S. & Newell, A. (1986). Chunking in Soar: the anatomy of a general learning mechanism. *Machine Learning*, 1, 11–46.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Langley, P. (2006). Intelligent behavior in humans and machines. Technical report. Computational Learning Laboratory, Stanford University.

- Langley, P., Laird, J.E. & Rogers, S. (2009). Cognitive architectures: research issues and challenges. *Cognitive Systems Research*, 10, 141-160.
- Lenat, D.B. (1995). CyC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11), 33-38.
- Lenat, D.B., Guha, R.V., Pittman, K., Pratt, D. & Shepherd, M. (1990). CyC: toward programs with common sense. *Communications of the ACM*, 33(8), 30-49.
- Lenci, A. (2000). Building an ontology for the lexicon: semantic types and word meaning. In *Workshop on Ontology-Based Interpretation of Noun Phrases*, Kolding.
- Liu, H. & Singh, P. (2004a). Commonsense reasoning in and over natural language. In *Proceedings of the 8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems. Lecture Notes in Artificial Intelligence (pp. 293-306)*. Berlin-Heidelberg: Springer.
- Liu, H. & Singh, P. (2004b). ConceptNet—a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4), 211-226.
- Mairal-Usón, R. & Periñán-Pascual, C. (2009). The anatomy of the lexicon within the framework of an NLP knowledge base. *Revista Española de Lingüística Aplicada*, 22, 217-244.
- Mairal-Usón, R. & Ruiz de Mendoza Ibáñez, F.J. (2009). Levels of description and explanation in meaning construction. In C. Butler & J. Martín Arista (Eds.), *Deconstructing Constructions (pp. 153–198)*. Amsterdam-Philadelphia: John Benjamins.
- McCarthy, J. & Hayes, P.J. (1969). Some philosophical problems from the standpoint of AI. *Machine Intelligence*, 4, 463-502.

- Minton, S.N. (1990). Quantitative results concerning the utility of explanation-based learning. *Artificial Intelligence*, 42, 363–391.
- Minsky, M. (1975). A framework for representing knowledge. MIT-AI Laboratory Memo 306.
- Newell, A. (1990). *Unified Theories of Cognition*. Cambridge: Harvard University Press.
- Newton, D. (1976). Foundations of attribution: the perception of ongoing behavior. In J.H. Harvey, W.J. Ickes, & R.F. Kidd (Eds.), *New Directions in Attribution Research, Vol. 1* (pp. 223-247). Hillsdale: Lawrence Erlbaum.
- Nirenburg, S. & Raskin, V. (2004). *Ontological Semantics*. Cambridge: MIT press.
- Panton, K., Matuszek, C., Lenat, D., Schneider, D., Witbrock, M., Siegel, N. & Shepard, B. (2006). Common sense reasoning: from CyC to intelligent assistant. In Y. Cai & J. Abascal (Eds.), *Ambient Intelligence in Everyday Life* (pp. 1–31). Berlin: Springer.
- Parmar, A. (2001). The representation of actions in KM and CyC. Technical report. Stanford University.
- Peacocke, C. (1992). *A Study of Concepts*. Cambridge: MIT Press.
- Periñán-Pascual, C. & Arcas-Túnez, F. (2004). Meaning postulates in a lexico-conceptual knowledge base. In *Proceedings of the 15th International Workshop on Databases and Expert Systems Applications* (pp. 38-42). Los Alamitos: IEEE.
- Periñán-Pascual, C. & Arcas-Túnez, F. (2005). Microconceptual-Knowledge Spreading in FunGramKB, In *Proceedings of the 9th IASTED International Conference on Artificial Intelligence and Soft Computing* (pp. 239-244). Anaheim-Calgary-Zurich: ACTA Press.

- Periñán-Pascual, C. & Arcas-Túnez, F. (2007a). Cognitive modules of an NLP knowledge base for language understanding. *Procesamiento del Lenguaje Natural*, 39, 197-204.
- Periñán-Pascual, C., Arcas-Túnez, F. (2007b). Deep semantics in an NLP knowledge base. In *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence* (pp. 279-288). Salamanca: Universidad de Salamanca.
- Periñán-Pascual, C. & Arcas-Túnez, F. (2010a). Ontological commitments in FunGramKB. *Procesamiento del Lenguaje Natural*, 44, 27-34.
- Periñán-Pascual, C. & Arcas-Túnez, F. (2010b). The Architecture of FunGramKB. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation* (pp. 2667-2674). Malta: ELRA.
- Periñán-Pascual, C. & Mairal-Usón, R. (2009). Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural*, 43, 265-273.
- Periñán-Pascual, C. & Mairal-Usón, R. (2010). La gramática de COREL: un lenguaje de representación conceptual. *Onomázein*, 21, 11-45.
- Periñán-Pascual, C. & Mairal-Usón, R. (fc) Universality and culturality in a natural language processing knowledge base.
- Pratt, V. (1994). CyC report. Technical report. Stanford University.
- Rey, G. (1994). Concepts. In S. Guttenplan (Ed.), *A Companion to the Philosophy of Mind* (pp. 185-193). Cambridge: Blackwell.
- Riesbeck, C.K. (1975). Conceptual Analysis. In R.C. Schank (Ed.), *Conceptual Information Processing*. Amsterdam: North-Holland.

- Riesbeck, C.K & Schank, R.C. (1978). Comprehension by computer: expectation-based analysis of sentences in context. In W.J.M. Levelt & G.B. Flores d'Arcais (Eds), *Studies in the Perception of Language*. New York: Wiley.
- Ruiz de Mendoza Ibáñez, F.J. (2007). High-level cognitive models: in search of a unified framework for inferential and grammatical behavior. In K. Kosecki (Ed.), *Perspectives on Metonymy* (pp. 11–30). Frankfurt-Main: Peter Lang.
- Ruiz de Mendoza Ibáñez, F.J. & Mairal-Usón, R. (2008). Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia Linguistica*, 42(2), 355–400.
- Rumelhart, D.E. (1975). Notes on a schema for stories. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and Understanding: Studies on Cognitive Science* (pp. 211-236). New York: Academic Press.
- Ruppenhofer, J., Ellsworth, M., Petruck, M., Johnson, C.R. & Scheffczyk, J. (2006). FrameNet II: extended theory and practice. Technical report. International Computer Science Institute, University of California at Berkeley.
- Schank, R.C. (1972). Conceptual Dependency: a theory of natural language understanding. *Cognitive Psychology*, 3(4), 532-631.
- Schank, R.C. (1975). *Conceptual Information Processing*. New York: Elsevier.
- Schank, R.C. (1982). *Dynamic Memory*. Cambridge: Cambridge University Press.
- Schank, R.C. & Abelson, R.P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale: Lawrence Erlbaum.
- Shapiro, D. & Langley, P. (1999). Controlling physical agents through reactive logic programming. In *Proceedings of the Third Annual Conference on Autonomous Agents* (pp. 386–387). Seattle: ACM Press.

- Singh, P. & Barry, B. (2003). Collecting commonsense experiences. In *Proceedings of the Second International Conference on Knowledge Capture*, Florida.
- Singh, P., Lin, T., Mueller, E.T., Lim, G., Perkins, T. & Zhu, W.L. (2002). Open Mind Common Sense: knowledge acquisition from the general public. In *Proceedings of the First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems. Lecture Notes in Computer Science (pp. 1223-1237)*. Berlin-Heidelberg: Springer.
- Smith, B. (1995). Formal ontology, common sense and cognitive science. *International Journal of Human-Computer Studies*, 43(5-6), 641-666.
- Taylor, G., Stensrud, B., Maddox, J. & Aycock, H.E. (2011). Formative evaluation of an IUI for supervisory control of CGFs. In *Proceedings of the 2011 Conference on Behavior Representation in Modeling and Simulation (pp. 185-191)*. Simulation Interoperability Standards Organization.
- Van Belleghem, K., Denecker, M. & De Schreye, D. (1995). Combining situation calculus and event calculus. In *Proceedings of the Twelfth International Conference on Logic Programming (pp. 83-97)*. Cambridge: MIT Press.
- Van Dijk, T.A. (2003). Specialized discourse and knowledge. A case study of the discourse of modern genetics. *Cadernos de Estudos Linguísticos*, 44, 21-55.
- Van Valin, R.D. (2005). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- Van Valin, R.D. & LaPolla, R. (1997). *Syntax, Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- Velardi, P., Pazienza, M.T. & Fasolo, M. (1991). How to encode semantic knowledge: a method for meaning representation and computer-aided acquisition. *Computational Linguistics* 17(2): 153-170.

- Wu, L.L. & Barsalou, L.W. (2009). Perceptual simulation in conceptual combination: evidence from property generation. *Acta Psychologica*, *132*, 173-189.
- Yeh, W. & Barsalou, L.W. (2006). The situated nature of concepts. *American Journal of Psychology*, *119*(3), 349-384.