Procesamiento del lenguaje Natural, nº 41 (2008), pp. 137-144 recibido 7-05-2008; aceptado 16-06-2008

A cognitive approach to qualities for NLP

Un enfoque cognitivo de las cualidades para el PLN

Carlos Periñán-Pascual, Francisco Arcas-Túnez
Universidad Católica San Antonio
Campus de los Jerónimos s/n
30107 Guadalupe - Murcia (Spain)
{jcperinan, farcas}@pdi.ucam.edu

Resumen: En muchas bases de conocimiento para el PLN prevalece actualmente un enfoque lexicista. En cambio, FunGramKB utiliza la ontología como módulo pivote entre los niveles léxicos y cognitivos, convirtiéndose así en el componente más importante. El propósito de este artículo es la descripción de los tipos semánticos que se asignan a las cualidades de FunGramKB y cómo el enfoque cognitivo que se adopta facilita tanto la estructuración de la base de conocimiento como el razonamiento en sistemas del PLN.
Palabras clave: Base de conocimiento, representación del conocimiento, ontología, postulado de significado, razonamiento, comprensión del lenguaje natural.

Abstract: Unlike most current NLP knowledge bases, where the lexicalist approach prevails, FunGramKB is ontology-oriented, since the ontology plays a pivotal role between the lexical and the cognitive levels. The objective of this paper is to describe the semantic types assigned to qualities in FunGramKB ontology and how the cognitive approach adopted facilitates the structuring of the knowledge base as well as the reasoning in NLP systems.
Keywords: Knowledge base, knowledge representation, meaning postulate, ontology, reasoning, natural language understanding.

1 FunGramKB ontology

FunGramKB (Periñán-Pascual and Arcas-Túnez, 2004, 2005, 2006, 2007) is a multipurpose lexico-conceptual knowledge base for natural language processing (NLP) systems. It is multipurpose in the sense that it is both multifunctional and multilanguage. In other words, FunGramKB can be reused in various NLP tasks (e.g. information retrieval and extraction, machine translation, dialogue-based systems, etc) and with several natural languages.[1]

FunGramKB is made up of five independent but interrelated modules: lexicon, morphicon, onomasticon, cognicon and ontology. The most important module is the ontology, since it is deemed as a pivotal component. For example, lexical units are assigned syntactic, pragmatic and collocational information in the lexicon, but their meaning representations are conceived as semantic properties in the ontology, so that every sense of a lexical unit is linked to a conceptual unit. FunGramKB ontology is presented as a hierarchical structure of all the concepts that a person has in mind when talking about everyday situations. This ontology distinguishes three different conceptual levels, each one of them with concepts of a different kind:

a. Metaconcepts, which constitute the upper level, are used as cognitive dimensions.
b. Basic concepts, which constitute the intermediary level, are used as defining units which allow the construction of

[1] Although the current version of FunGramKB deals with just two languages (i.e. English and Spanish), we intend to develop lexica for French, German and Italian.

ISSN: 1135-5948

interlinguistic meaning postulates for basic concepts and terminals.

c. Terminals constitute the leaf nodes of the ontology, so hierarchization at this level is practically non-existent. The borderline between basic concepts and terminals is based on their definitory potential to take part in meaning postulates.

Basic concepts and terminals are not stored as atomic symbols but are provided with a rich internal structure consisting of properties such as the semantic types and the meaning postulate. This paper focuses on the knowledge representation of qualities in FunGramKB ontology.[2] More particularly, section 2 gives an accurate description of their semantic types and section 3 presents the benefits of the cognitive approach adopted for NLP.

2 *The semantic types of qualities*

FunGramKB assigns four different semantic types to every concept linked to the metaconcept #QUALITY, i.e. the cognitive dimension for the qualities: intersective/subsective, dynamic/static, gradable/non-gradable and polar/serial/none. Although these types have been already used in other NLP models—namely, Generalized Upper Model (Bateman, Henschel and Rinaldi, 1995), Mikrokosmos (Raskin and Nirenburg, 1995, 1998), EAGLES (1999) and SIMPLE (Peters and Peters, 2000), a cognition-oriented approach is adopted in FunGramKB. Thus, lexical-syntactic substitution tests characteristic of traditional semantic analyses (Quirk et al., 1985; Cruse, 1986) are not suitable for the diagnosis of concepts. For instance, supposing that α and β are lexical realizations of a noun and an adjective respectively, the following validation tests have been usually employed to verify that the adjective is [i] dynamic or [ii] gradable:

[i] *α is being β*
[ii] *α is very β* or *How β is α?*

On the contrary, the semantic types of FunGramKB qualities are exclusively determined on the basis of semantic criteria, regardless of their surface realizations.[3]

2.1 Intersective/subsective

This parameter takes into account the speaker's standpoint on the truth value expressed by the quality, resulting in the dichotomy of intersectivity (i.e. absolute truth-value) and subsectivity (i.e. relative truth-value). For example, the concept NAKED,[4] understood as "not wearing any clothes", is shared by all people in such an identical way that it causes no disagreement when describing the same reality. On the contrary, not all individuals can perceive the same instance of an entity as INTERESTING, maybe because, for some speakers, the instance (e.g. a theory, a class, etc.) provides information that they already know. Therefore, NAKED and INTERESTING are intersective and subsective concepts respectively.

2.2 Dynamic/static

A quality is dynamic when, for the same instance of entity, the quality can be affected along the time—because of the nature of the entity itself or an action exerted by an external entity. Otherwise, the quality is static. For example, HOT describes a quality that can be temporally present in an instance of entity, so the concept is dynamic. On the contrary, GERMAN, understood as "born in Germany", is static, since it refers to a quality which will never be altered in an instance of entity.

2.3 Gradable/non-gradable

A quality is gradable (e.g. EXPENSIVE) when, for the same instance of entity, the quality can take varying degrees of intensity along the time. Otherwise, the quality is non-gradable (e.g. ALIVE).

2.4 Polar/serial/none

FunGramKB describes meaning oppositions by locating them in cognitive spaces, where positive and negative focal concepts are

---

[2] Currently, FunGramKB contains approximately 750 and 630 adjectives in the English and Spanish lexica respectively, organized into 320 full-featured concepts in the ontology.

[3] This is the reason why a careful distinction is made in this paper between the terms 'noun-adjective' (lexical labels) and 'entity-quality' (conceptual labels) respectively.

[4] Although their names are represented by English words in block letters, concepts are not language dependent. Thus, lexical units such as *naked* and *nude* (English), or *desnudo* and *en cueros* (Spanish), are all linked to NAKED.

determined. Here terms such as "positive" and "negative" are not applied to refer to a kind of meaning connotation, but to the presence or not of the negation operator in the meaning representation. In other words, the negative focal concept is defined as the negation of the positive one: e.g. *false* means *not true*. Evidently, if A is the opposing concept of B, then there is no need to state that B is the opposing concept of A. *A priori*, any of the two focal concepts in a dimension is liable to be deemed as positive. However, FunGramKB knowledge engineers follow the arbitrary criterion of taking as positive the concept to which the lexical unit with the highest frequency index[5] is linked.[6] Figure 1, which has been captured from FunGramKB Suite,[7] illustrates a case of non-gradable polarity.

If there is gradation within a semantic dimension, the concepts involved are described around the two focal concepts, which are determined by comparing the frequency indices of the lexical units linked to all those concepts. More particularly, the positive one is selected on the basis of the highest index, and the negative one follows the same criterion but taking into account just those concepts located in the opposing side of the dimension. For the remaining concepts, quantifying operators *m* (many/much) and *p* (few/little) are used to describe different degrees of intensity around the focal concepts. Figure 2 displays *FunGramKB* framework for the semantic representation of oppositions involving qualities. As can be seen, a cognitive dimension in which qualities are involved in a meaning opposition can be divided up to seven semantic zones, where the central one results from the negation of both focal concepts. Indeed, the canonical structure of these dimensions is determined by the combination of two semantic types (Table 1).

zones	type	examples[8]
7	gradable + serial	BEAUTIFUL, HAPPY, INTELLIGENT
6	gradable + polar	ANGRY, SENSITIVE, TIRED
4	quasi-gradable[9] + polar	DIRTY, NOISY, OPEN
2	non-gradable + polar	ARTIFICIAL, MALE, WRONG

Table 1: Structuring meaning oppositions.

One of the key features of these semantic zones is their "cognitive feasibility", what does not necessarily imply "lexicalization". In other words, every semantic zone can be represented by a concept, but it is possible for a particular language to have no lexical realization for that concept. In fact, the difference between series and polarities lies on the cognitive feasibility of the central semantic zone, regardless of the possibility lexicalization in that zone. For example, Figure 3 illustrates the dimension of size. Although not all their semanitic zones are lexicalized in English or Spanish, it is treated as a series, since any zone is "potentially lexicalizable" when embedding other natural languages in the knowledge base.

2.5 Final remarks

Since FunGramKB approach is remarkably conceptualist, validation tests based on the linguistic behaviour of adjectives are substituted for questions concerning the conceptual perception of qualities in the real world:

a. Do all individuals share the same view of the quality when present in an instance of entity? (Y: intersective/N: subsective)

[5] This frequency index is obtained from WordNet.

[6] In order to facilitate meaning representation, this criterion is not applied in the case that standard dictionaries use a less frequent concept to define the opposing one. Some examples are *natural-artificial* and *different-similar*, where the first adjective is more frequent but the second one is preferred as defining word.

[7] FunGramKB Suite has been developed in order to assist engineers in the acquisition and maintenance of the knowlege base. In this case, the tool can automatically reconstruct a whole semantic dimension from the meaning postulate of a particular quality.

[8] Each example is represented by the positive focal concept of a cognitive dimension.

[9] A cognitive dimension is "quasi-gradable" if one side is gradable but the other isn't. For example, an instance of entity can be open in different degrees (i.e. gradable quality), but that instance can only stay in one position if it is closed (i.e. non-gradable quality).

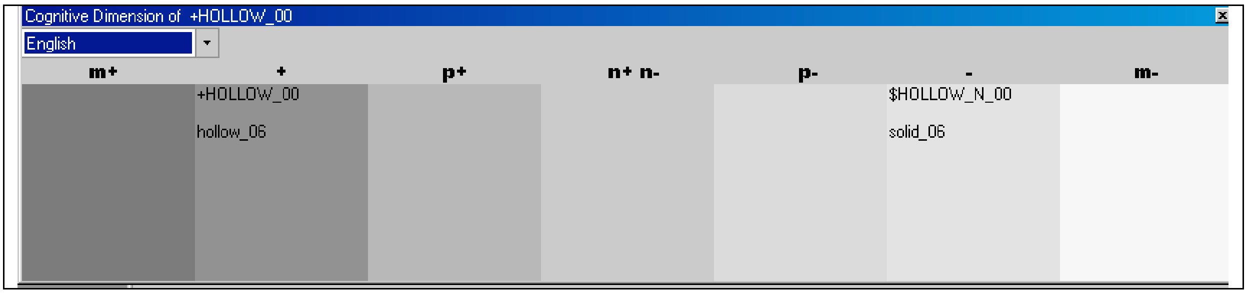

Figure 1: A sample of non-gradable polarity.

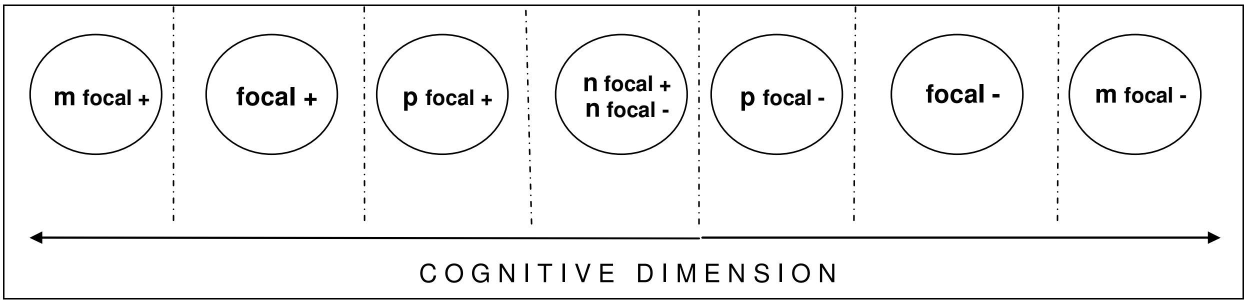

Figure 2: Cognitive framework for the representation of meaning oppositions.

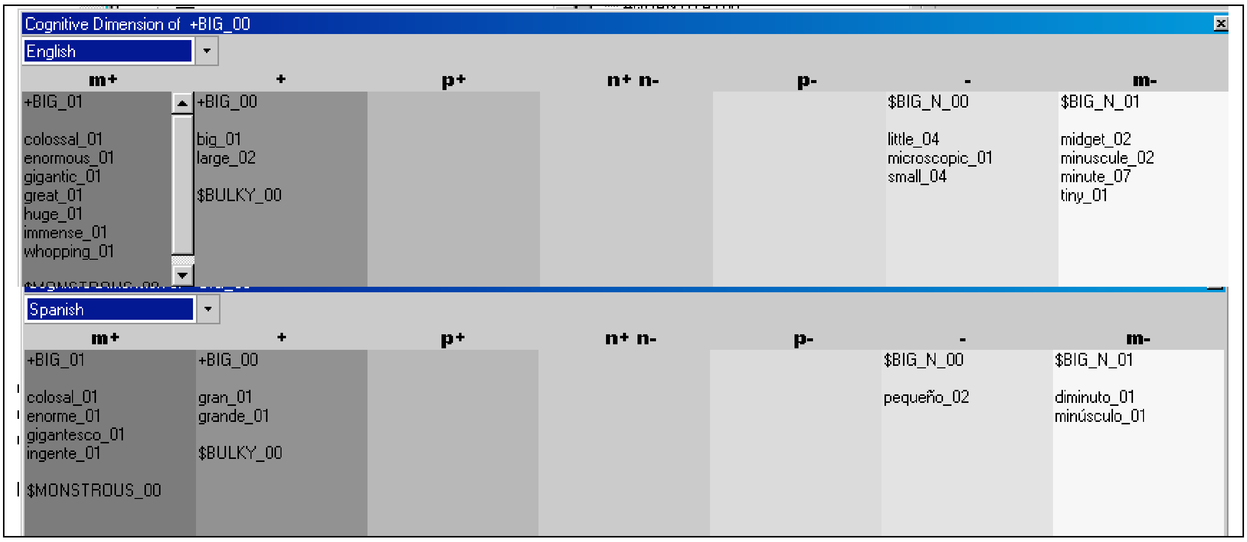

Figure 3: A sample of gradable series.

b. Can the quality be affected along the time when present in an instance of entity? (Y: dynamic/N: static)
c. Can the quality take varying degrees of intensity along the time when present in an instance of entity? (Y: gradable/N: non-gradable)
d. Does the quality take part in a meaning opposition? (Y: next question/N: none)
   Does the absence of that quality in an instance of entity necessarily imply the presence of the opposing quality? (Y: polar/N: serial)

3 NLP and the cognitive approach to qualities

This view of the semantic types of qualities facilitates the structuring of the NLP knowledge base (e.g. ontological modelling and composition of meaning postulates) as well as the reasoning in natural language understanding systems, among other advantages.

3.1 Ontological modelling

On the one hand, the distribution of basic qualities within meaning oppositions (i.e. polarity or series) made engineers re-arrange concepts between the basic and terminal levels of the ontology. The source of the inventory of FunGramKB basic concepts was the defining vocabulary in *Longman Dictionary of Contemporary English* (1988).[10] However, some of these concepts were finally demoted to terminals, because of their place in the cognitive dimension. For example, adjectives *big* and *small* are included in the *Longman* vocabulary, so initially they were going to be treated as basic concepts (i.e. +BIG and +SMALL). However, since +SMALL plays the role of negative focal concept, it was finally stored as terminal concept. But the problem raised when building the meaning representation of the concept to which adjectives *midget*, *minuscule*, *minute* or *tiny* are linked, since it is impossible to describe them accurately without using the concept SMALL. The solution to this and other similar problems in the conceptualization of the *Longman* vocabulary lied in the adoption of the following protocol:

a. All concepts in a polarity or series are stored as terminal, except for the positive focal concept. For example, in the dimension of size + BIG_00 is the positive focal concept and $BIG_N_00 is the negative one, which has been demoted to terminal concept (Figure 3).
b. However, if any terminal concept in the dimension serves as a co-superordinate of another concept, then the former is promoted to basic concept. The reason is that only basic concepts can be used as defining units, and superordinates appear in the meaning postulate of subordinate concepts. For example, +BIG_01 was promoted to basic concept, which is one of the superordinates of $MONSTROUS_00 (Figures 3 and 4).
c. The names of all terminal concepts in a dimension will be formed out of that of the positive focal concept plus the infix _N_ (i.e. not). The reason is that the only case in which FunGramKB permits terminal concepts to be included in meaning postulates occurs when the terminal concept owns the same name as the *definiendum*. In this way, both focal concepts can be used as descriptors in gradable cognitive dimensions. For example, $BIG_N_00 is used to describe the meaning of $BIG_N_01 (Figure 3).

On the other hand, the gradation and polarity parameters shape the IS-A taxonomy of qualities. Firstly, both focal concepts are siblings. Secondly, neighbouring concepts around the focal ones should be subsumed by their corresponding focal concept. For example, Figure 4 displays the hierarchical structure of the dimension of size (Figure 3).

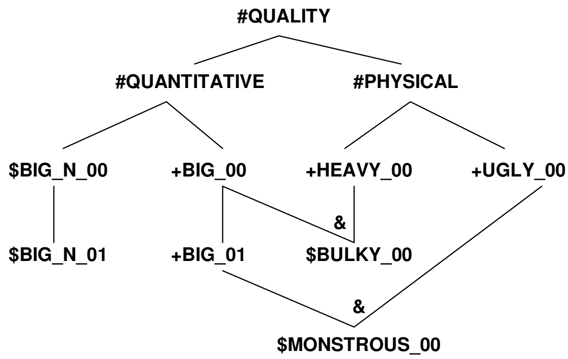

Figure 4: A sample of the quality taxonomy.

Therefore, these two parameters help to formally describe qualities around two complementary axis: a horizontal axis in which meaning oppositions are organised internally and a vertical axis in which qualities are related by subsumption.

3.2 The composition of meaning postulates

A meaning postulate is a set of one or more logically connected predications (e_1, e_2... e_n), which are cognitive constructs carrying the generic features of the concept.[11] Concepts, and not words, are the building blocks for the formal description of meaning postulates, so a meaning postulate becomes a language-independent semantic knowledge representation. To illustrate, some predications in the meaning postulate of STRONG are presented:

[10] In the conceptualization phase, lexical units from the *Longman* vocabulary were mapped into basic cognitive units.

[11] The formal grammar of well-formed predications for meaning postulates in FunGramKB is described in Periñán-Pascual and Arcas-Túnez (2004).

+STRONG_00
*(e1: +BE_01 (x1: +HUMAN_00 ^ +ANIMAL_00)Theme (x2: +STRONG_00)Attribute)
*(e2: +HAVE_00 (x1)Theme (x3: m +ENERGY_00)Referent (f1: ((e3: pos +MOVE_00 (x1)Agent (x4)Theme (x5)Location (x6)Origin (x7)Goal) (e4: +BE_01 (x4)Theme (x8: +HEAVY_00)Attribute)))Result)

These predications have the following natural language equivalents:

> A person or animal that is strong has a lot of physical power to that they can move heavy things.

In FunGramKB, semantic types of qualities (mainly, gradation and polarity parameters) help to determine the structural pattern of predications in their meaning postulates. More particularly, the canonical layout structure is as follows:

a. The first predication provides information about prototypical entities to which the quality is usually assigned. This predication is structured as follows:

(e_1: +BE_01 (x_1: *<entity>*)$_{\text{Theme}}$ (x_2: *<quality>*)$_{\text{Attribute}}$)

For example:

+ROUND_00
*(e1: +BE_01 (x1: +CORPUSCULAR_00)Theme (x2: +ROUND_00)Attribute)

b. In case of cognitive dimension, the second predication can explicitly state the entity which best describes that dimension. This predication is structured as follows:

(e_2: +BE_00 (x_2: *<quality>*)$_{\text{Theme}}$ (x_3: *<entity>*)$_{\text{Referent}}$)

For example:

+RED_00
+(e2: +BE_00 (x2: +RED_00)Theme (x3: +COLOUR_00)Referent)

c. In case of meaning opposition, the third predication describes the quality in relation to a focal concept. This predication can be structured by one of the following patterns:

[i] (e_3: n +BE_01 (x_1: *<entity>*)$_{\text{Theme}}$ (x_4: *<quality>*)$_{\text{Attribute}}$)

[ii] (e_3: +BE_01 (x_1: *<entity>*)$_{\text{Theme}}$ (x_4: m *<quality>*)$_{\text{Attribute}}$)

[iii] (e_3: +BE_01 (x_1: *<entity>*)$_{\text{Theme}}$ (x_4: p *<quality>*)$_{\text{Attribute}}$)

For example:

[i] $INTELLIGENT_N_00 [*daft, idiotic, imbecile, silly, stupid*]
*(e2: n +BE_01 (x1: +HUMAN_00)Theme (x3: +INTELLIGENT_00)Attribute)

[ii] $INTELLIGENT_01 [*brainy, brilliant, gifted*]
*(e2: +BE_01 (x1: +HUMAN_00)Theme (x3: m +INTELLIGENT_00)Attribute)

[iii] $INTELLIGENT_N_01 [*simple*]
*(e2: +BE_01 (x1: +HUMAN_00)Theme (x3: p $INTELLIGENT_N_00)Attribute)

The choice of one particular type of pattern depends on the location of the *definiendum* within the meaning opposition. In other words, [i] is used when describing the negative focal concept—where x_4 is the positive one, and [ii]-[iii] for the description of any other concept in the dimension except for the focal ones—indeed, one of the focal concepts should be referenced by x_4. When the *definiendum* plays the role of positive focal concept (i.e. +INTELLIGENT_00), there is no need to include this third predication, so that redundancy is minimized. On the contrary, the concept located in the central zone of the opposition requires two predications for the negation of both focal concepts:

$MATURE_00
*(e3: n +BE_01 (x1: $MATURE_00)Theme (x4: +YOUNG_00)Attribute)
*(e4: n +BE_01 (x1)Theme (x5: +YOUNG_N_00)Attribute)

d. Further features about the *differentiae* can be described in other predications of the meaning postulate.

This description of the structural pattern of meaning postulates is aimed at qualities with a single parent node in the ontology. In case of multiple inheritance (i.e. multi-parent qualities), the pattern is slightly different, since the second predication is used to list all superordinate concepts, being connected by means of logical operators:

(e_2: +BE_01 $(x_1)_{Theme}$ (x_3: <*superordinate*$_1$> <*op*> <*superordinate*$_2$> <*op*>… <*superordinate*$_n$>$)_{Attribute}$)

For example:

$CLAMMY_00
*(e1: +BE_01 (x1: +CORPUSCULAR_00)Theme (x2: $CLAMMY_00)Attribute)
*(e2: +BE_01 (x1)Theme (x3: +HOT_N_00 & +WET_01)Attribute)

3.3 Reasoning in Natural Language Understanding

Some NLP systems, e.g. machine translation or dialogue-based systems, attempt to "understand" the input text by translating it into a formal language-independent representation. This approach requires a knowledge base with conceptual representations which reflect the structure of human beings' cognitive system. However, this type of knowledge-based systems should also be provided with a reasoning engine. To this respect, semantic types of FunGramKB qualities (mainly, intersectivity and dynamism) can enhance reasoning results in AI systems.

A key issue in natural language understanding is the treatment of non-monotonicity. In FunGramKB, each predication taking part in a meaning postulate is preceded by a reasoning operator in order to state if the predication is strict (+) or defeasible (*). FunGramKB inference engine handles predications as rules, allowing monotonic reasoning with strict predications, and non-monotonic with defeasible predications. Strict predications are law-like rules, which have no exceptions: e.g. whales are mammals, circles are round. On the other hand, defeasible predications can be defeated by contrary evidence: e.g. birds typically fly.[12]

The intersectivity parameter of qualities can determine the choice of the reasoning operator in their meaning postulates. Indeed, in the case of subjective qualities, all predications in the meaning postulate are headed by the defeasible operator. The reason lies on the fact that, because the truth value of some qualities is subject to speakers' interpretation, it cannot be really assured that all individuals will share the same truth value when perceiving the same instance of entity. To illustrate, the meaning postulate of HEAVY is presented:

+HEAVY_00
*(e1: +BE_01 (x1)Theme (x2: +HEAVY_00)Attribute)
*(e2: +WEIGH_00 (x1)Theme (x3: +MUCH_00)Attribute (f1: (e3: pos +MOVE_00 (x4: +HUMAN_00)Agent (x1)Theme (x5)Location (x6)Origin (x7)Goal (f2: +DIFFICULT_00)Manner))Result)

When you say that an instance of entity is heavy, it is supposed to weigh a lot, but what do you mean by 'a lot'? The answer is directly dependent on the individual's physical strength and/or the weight of other instances of that entity. In other words, the assessment of weight is conditioned by the context of the world model, leading to the relativism of the truth value of the predications in the above example. Therefore, strict operators are not appropriate for this case.

With regard to temporal reasoning in natural language understanding, FunGramKB also contributes to mitigate the effects of the persistence problem in temporal projection. Until the 1980s, much debate was raised over the "frame problem" (McCarthy and Hayes, 1969), i.e. the construction of a logic-based model to efficiently represent the things which remain the same as actions are performed. Many solutions were then proposed, and the classical problem was solved. Most of these proposals were based on the "common sense law of inertia", whereby "a fluent[13] is assumed to persist unless there is reason to believe otherwise" (Shanahan, 1997). However, when reasoning over time with dynamic domains, many of those solutions proposed to the classical frame problem do not work for the persistence problem, i.e. the difficulty of determining which things remain the same in a changing world (Morgenstern, 1996). The problem is that some properties change even when no event occurs that interrupts them. In

[12] An accurate account of how these operators work in common-sense reasoning is described in Periñán-Pascual and Arcas-Túnez (2004).

[13] In temporal logics, a fluent is usually understood as anything whose truth value is subject to change over time.

FunGramKB, the dynamism parameter helps to predict which qualities are liable to change and which ones remain the same in a given situation, particularly valuable in understanding unexpected changes.

4 Conclusions

Currently most NLP systems adopt a relational approach to represent lexical meanings, since it is easier to state associations among lexical units in the way of meaning relations than describing the cognitive content of lexical units formally. Although large-scale development of deep-semantic resources requires a lot of time and effort, the expressive power of conceptual meanings is much more robust (Periñán-Pascual and Arcas-Túnez, 2007). Within this framework, describing semantic properties of qualities according to the syntactic criteria of traditional lexical semantics would have been an inconsistent decision. Moreover, the cognitive approach to semantic types benefits the construction of a sound NLP knowledge base. On the one hand, parameters such as gradation and polarity improve the cognitive economy in conceptual description and organization, as well as facilitating the efficient management and maintenance of the knowledge base. On the other hand, parameters such as intersectivity and dynamism contribute to improve the performance of reasoning engines in natural language understanding systems.

Bibliography

Bateman, J.A., R. Henschel, and F. Rinaldi. 1995. The Generalized Upper Model 2.0. Technical report. IPSI/GMD, Darmstadt.

Cruse, D.A. 1986. *Lexical semantics*. Cambridge University Press, Cambridge.

EAGLES. 1999. Preliminary recommendations on lexical semantic encoding. Technical report EAGLES LE3-4244.

McCarthy, J. and P.J. Hayes. 1969. Some philosophical problems from the standpoint of AI. *Machine Intelligence*, 4: 463-502.

Morgenstern, L. 1996. The problem with solutions to the frame problem. K.M. Ford and Z.W. Pylyshyn (eds.) *The robot's dilemma revisited: the frame problem in AI*. Ablex, Norwood: 99-133.

Periñán-Pascual, C. and F. Arcas-Túnez. 2004. Meaning postulates in a lexico-conceptual knowledge base. *Proceedings of the 15th International Workshop on Databases and Expert Systems Applications*. 38-42, IEEE, Los Alamitos.

Periñán-Pascual, C. and F. Arcas-Túnez. 2005. Microconceptual-knowledge spreading in FunGramKB. *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*. 239-244, ACTA Press, Anaheim-Calgary-Zurich.

Periñán-Pascual, C. and F. Arcas-Túnez. 2006. Reusing computer-oriented lexica as foreign-language electronic dictionaries. *Anglogermánica Online*, 4: 69-93.

Periñán-Pascual, C. and F. Arcas-Túnez. 2007. Cognitive modules of an NLP knowledge base for language understanding. *Procesamiento del Lenguaje Natural*, 39: 197-204.

Periñán-Pascual, C. and F. Arcas-Túnez. 2007. Deep semantics in an NLP knowledge base. *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence*. 279-288, Universidad de Salamanca, Salamanca.

Peters, I. and W. Peters. 2000. The treatment of adjectives in SIMPLE: theoretical observations. *Proceedings of the Second International Conference on Language Resources and Evaluation*, Athens.

Quirk, R., S. Greenbaum, G. Leach, and J. Svartvik. 1985. *A comprehensive grammar of the English language*. Longman, London.

Raskin, V. and S. Nirenburg. 1995. Lexical semantics of adjectives: a microtheory of adjectival meaning. Technical report MCCS-95-288, Computing Research Laboratory-New Mexico State University.

Raskin, V. and S. Nirenburg. 1998. An applied ontological semantic microtheory of adjective meaning for natural language processing. *Machine Translation*, 13: 135-227.

Shanahan, M. 1997. *Solving the frame problem: a mathematical investigation of the common sense law of inertia*. MIT Press, Cambridge (MA).