

Extension of the SpanishWordNet

Clara Soler

Universitat Ramon Llull, C. Padilla 326–332, 08025 Barcelona, Spain
Email: `clarasp@blanquerna.ur1.es`

Abstract. WordNet divides adjectives in descriptives and relationals basically and they are represented in an enumerative way. The category was not introduced in EuroWordNet and Spanish adjectives in the SpanishWordNet are the translation of the English synsets. This paper describes a proposal of organizing and incorporating adjectives into the SpanishWordNet in terms of representing its polymorphic behaviour. The new organization would be made according to the adjectives taxonomy of MikroKosmos ontology. It results that the ontological approach can be used to explain adjectives polysemy. In the end a new adjectival classification appears in EuroWordNet, in terms of the three types of entities of the Top Ontology.

1 Introduction

This paper describes a proposal for incorporating and organizing Spanish adjectives into the lexical database SpanishWordNet by means of the MikroKosmos Ontology¹. Adjectives being currently displayed in the SpanishWordNet are a translation of the English adjectives contained in WordNet (version 1.5) into Spanish and Catalan languages. Thus, their semantic organization follows the model of the WordNet system.

It has been already suggested to extend EuroWordnet with language-neutral ontologies, such as CYC, MikroKosmos or Sensus [11]. In this case, in which adjectives are the focus, the procedure adapted to carry out the extension of the SpanishWordNet will be based on expanding the Top Concept Ontology of EuroWordNet with part of MikroKosmos ontology structure. The choice of MikroKosmos ontology is due to its lexical approach to represent a model containing information about types of things.

Section 2 outlines the classification and organization of adjectives in WordNet 1.5 and in the SpanishWordNet, as well as how adjective polysemy is considered in these databases. Section 3 proposes the extension of the SpanishWordNet by means of MikroKosmos; this will imply the incorporation and classification of adjectives according to ontological criteria, and will further imply a new classification of adjectives in terms of the three types of entities that constitute the Top Ontology of EuroWordNet, and finally the possibility of presenting the new approach to polysemy.

¹ *MikroKosmos Ontology* is one of the components of the MikroKosmos project on computational semantics, which is an automatic knowledge-based translation system. It is integrated by diverse microtheories whose objective is to describe the static meaning of all the lexical categories in different languages [10]

2 Adjectives Polysemy in WordNet, EuroWordNet, and the SpanishWordNet

2.1 In WordNet

Antonymy is the lexical relation that expresses in WordNet synsets which are opposite in meaning. It also divides and organizes adjectives in two main classes: the class of Descriptives, which have antonyms, and the class of Relationals, without antonyms². The first ones are organized into non-hierarchical synsets formed by one, or more, pairs of antonym adjectives. Relational adjectives are represented with pointers to the noun or verb from which they derive.

Apart from constituting a criterion to classify adjectives in WordNet, the relation of Antonymy helps to disambiguate polysemous nouns (as is stated in Fellbaum [5]). Descriptive adjectives express opposed values of attributes, most of which are bipolar. Some of these adjectives do not have direct antonyms but can acquire them indirectly via another semantic relation (Similar to). This relation distinguishes a peripheral or satellite adjective synset linked to the most central synset. This peripheral adjective (e.g. *moist*) may do not have a direct antonym, but via the Similar to Relation (*moist* is Similar to *wet*) acquires the indirect antonym *dry*. Now, as it is observed in Fellbaum [5] many of the less frequent and unusual adjectives (and less polysemous), are quite selective in relation to the noun they modify, and thus they constitute a class of adjectives that can be used to disambiguate the meaning of a polysemous noun. This verification suggests a division between a small set of highly polysemous common adjectives, such as *big, small, good, bad, new, old* etc., and a greater set of more discriminating, less interchangeable adjectives, like *academic* and *international*. A distinction can be made between those adjectives that can help to disambiguate a noun and those that cannot.

This difference is reflected between direct and indirect antonyms. Indirect antonyms are compatible with less nominal heads and are therefore less polysemous. They probably contribute to the disambiguation of the nouns that modify. The Relational ones are not organized in terms of sets of antonyms and are less polysemous.

2.2 In EuroWordNet

There are two main reasons why adjectives were not included in EuroWordNet. It is considered in Fellbaum [5] the information conveyed by an adjective, being a modifier, is less vital than the one expressed by nouns and verbs for understanding sentences in an NLP system. The other reason is the difficulty of its own semantics: adjectives are considered highly polysemous and that makes difficult to represent it in an enumerative lexicon like EuroWordNet, where is pretended to distinguish all senses of a word form.

However, if adjectives semantics was reconsidered, and perhaps their polysemy was not that high probably their inclusion in EuroWordNet would present less difficulty. In WordNet adjective polysemy is related to features such as its frequency, its compatibility with greater or smaller number of nominals, and noun disambiguation. This paper presents an approach to adjective polysemy based on completely different criterion.

² Apart from these types WordNet contains the participial adjectives file. These adjectives are considered a kind of Descriptives without antonyms, and are kept in a separated file.

2.3 In the SpanishWordNet

As previously stated, the Spanish and Catalan adjectives adopted by the SpanishWordNet are translations of the English word represented in WordNet 1.5. These adjectives are expressed in an enumerative and descriptive way, in correspondence with the structure of the semantic net. WordNet has already been noted for its excessive grain size, and sometimes the number of lexical entries exceeds those really necessary. This is especially the case of adjectives considered highly polysemous (Fellbaum [5]) e.g. *big, good, big*... This situation also occurs in the SpanishWordNet.

3 Extension of the SpanishWordNet

Adjective classification in WordNet is extensive enough and it takes account of both semantic and syntactic information. However there are some questions that remain with no answer: the classification does not give any account of the relation between different senses of the same adjectival form. What happens with Relationals? Are not they polysemous? Etc. We propose in this paper that the same classification can be made from other criterion. The fact that the Antonymy relation is a lexical relation between word forms and not concepts makes difficult to give an explanation of adjective polysemy. The proposed new criterion come from within the framework of the MikroKosmos project [9]. In this model, the lexicon mediates between a language of meaning representation and an ontology. Adjective meaning is explained according to this conceptual ontology. The lexical entries are instances of ontological types, and each one of them indicates a lexical connection of these units of the language to ontological concepts. In this framework adjectives are divided into Scalars, which are based on ontological concepts of Property and into Non-Scalars (Relationals). These are then subdivided into Denominals, that are based on ontological concepts of Entity, and in Deverbals, that are based on ontological concepts of Event. The basic criterion to establish a class of adjective is its association to a certain ontological type. This representation reflects the semantic structure of the adjective. An adjective always has either a noun or a verb as a reference. Ontological criterion supplies extra information to lexical criterion of WordNet classification and makes the analysis to become deeper and more comprehensive.

3.1 Treatment and Representation of Adjectives Polysemy

Adjective polysemy has become a subject increasingly studied within the area of NLP studies. The so called adjectival polymorphism is logically treated differently from different perspectives and points of view. It is not the objective of this paper to explain our own point of view about it, but we can outline the following:

- Polysemy always implies a change of meaning. It is possible to distinguish then between ambiguity and polysemy. In the case of ambiguity the adjective really does not suffer a change of meaning but acquires different shades. This is the case for instance of an adjective such as *good*, considered highly polysemous in several works (WordNet itself is an example). From our point of view in most of the cases is just ambiguity.

- Adjective polysemy can be explained within the framework of the lexical ontological semantics formulated in MikroKosmos. According to the ontological classification argued in [9], an adjective is polysemous when it is more of an ontological type. A casuistry then appears, examples of which are given next.

A Change of Ontological Type: from an Ontology of Entities or Events to an Ontology of Properties.

(1) Next Sunday I have a *familiar* meal.

The ontological nature of this adjective is of Entity.

(2) At work there is a very *familiar* atmosphere.

The ontological nature of this adjective is Evaluative (Scalar type).

A Change within the Same Ontological Type.

(3) El día *claro*³

Property: luminosity. Antonym: dark

(4) La crema *clara*⁴

Property: density. Antonym: thick

(5) *High* mountains

Property: height. Antonym: low

(6) *High* sea

In this case *high* is a synonym of *stormy*, which is a Denominal adjective. Antonym: calm.

The change takes place here within the Scalar adjectives. In (3), and (4) a change of scale takes place: from the property *brightness*, into the scale of the property *thickness*. It is a change that occurs at the same level, and apparently it seems difficult to predict which property comes first and which one derives from the other. Nevertheless, the change in (5) and (6) is different. The jump occurs from a scale of an objective property, such as *height*, to a Non-Scalar adjective. *High* is obviously a Scalar adjective, but probably in the case of *High sea* it becomes part of a collocation of a semi-compositive nature. Its real meaning becomes *stormy* which is a Denominal.

3.2 Extension of the SpanishWordNet

Having established our proposal concerning the polysemy of an adjective the next step is to establish the procedure to implement a representation of the adjectives in the SpanishWordNet which could express the new classification. This could be carried out by introducing some of the features of the MikroKosmos ontology to the Top Ontology of EuroWordNet. Let us outline the structure of both ontologies:

³ *El día claro* means *the bright day*

⁴ *La crema clara* means *the thin custard*

Top Ontology Structure EuroWordNet was founded upon two parts: the covering of a shared set of common Basic Concepts; and the extension of the lexical base from these Concepts using semiautomatic techniques. The Basic Concepts constitute a set of 1024, and the Top Ontology was created in order to classify them. It consist of 63 fundamental semantic distinctions used in various semantic theories and paradigms. These Top Concepts are organized by means of subtype and opposition relations. The ontology provides an independent structuring of the language to the Basic Concepts in terms of these semantic distinctions (which are considered to be more semantic features than common conceptual classes). These 63 semantic distinctions are classified by three types of entities: 1st-Order-Entities, 2nd-Order-Entities and 3rd-Order-Entities (following Lyons [6]). The 2nd-Order-Entities are those that can be denoted by any part of the speech: nouns, verbs, adjectives and adverbs. They represent any static or dynamic situation that cannot be grasped, seen, felt, or experienced as an independent physical thing. They are located in time and they can happen rather than exist. The Top Ontology is linked to the ILI (Inter-Lingual-Index), so are the word meanings into the local synsets (local wordnets such as the SpanishWordNet).

Mikrokosmos Structure MikroKosmos is organized according to a set of concepts. Each concept constitutes a collection of properties with partially specified values. The concepts are organized hierarchically. Semantically, the first difference between them occurs between ‘free concepts’ and ‘bounded concepts’. The ‘free concepts’ represent classes of objects and classes of events that have their corresponding instances in a TMR (Text Meaning Representation). The ‘bounded concepts’ represent classes of properties that categorize the objects and the events and that normally do not have instances but appear as values of the objects and instantiated events. The Concept Root is *ALL*, and the subclasses are *Events*, *Objects* and *Properties*.

Properties are the conceptual basics of the ontology. They help to define the concepts and can appear in the ontology in two different ways: As defined types of concepts or as values of the definitions of the objects and events. A value is the basic mechanism that represents relations between concepts. It is the fundamental metaontological mechanism. One of the subtypes of values is ‘relaxable-to’, which indicates the point at which the ontology allows violations of the restrictive selections giving rise to non literal uses such as the metaphor or metonymy. A proposal to extend SpanishWordNet is given next:

1. In the Top Ontology The adjectives must be classified under the 2nd-Order-Entities. Some of the Base Concepts belonging to these type of entities already refer to situations which can be denoted by adjectives (e.g. Social, Physical). It is therefore necessary to make an exhaustive verification in order to know which of those Basic Concepts refer to adjectives. It will then be possible to establish which are the lacking concepts.

2. In the MikroKosmos Ontology MikroKosmos Ontology establishes nine scales/properties of numerical type and four of literal type⁵ that identify Qualifying (descriptive) adjectives. It is therefore necessary to verify if all scalar adjectives are covered.

⁵ These scales come to be the ontological correlate of the different adjective taxonomies proposed in the framework of semantic studies of adjectives, being the one proposed by Dixon [4] the most relevant.

3. Joining both ontologies

- Incorporation of the adjectival synsets into the ILLI.
- Incorporation of the adjective ontological taxonomy of MikroKosmos into the Top Ontology. This will consist of the following: Incorporation of the new basic concepts that represent the adjective scales into the class of 2nd-Order-Entities. The scalar adjectives will be defined according to these concepts. Denominals will be defined according to the Basic Concepts (nominal) classified as 1st-Order-Entities and 3rd-Order-Entities, which are already introduced as they constitute EuroWordNet. Deverbals will also be defined according to basic concepts classified as 2nd-Order-Entities. Polysemy can be represented using the ‘relaxable-to’ relation, which can connect the different ontological concepts, from which the adjective derives its different meanings.

4. Later classification of adjectives The incorporation of these new basic concepts allows a double classification according to the three specified entity types. On the one hand, one classification of adjectives can be made according to the basic concepts understood as 2nd-Order-Entities, since any of them will be subsumed under these entities, and on the other hand they can be classified according to the ontological concepts from which they derive, that is to say, according to all three entity types, of first, second and third order.

4 Conclusions

We put forward a proposal to classify, to reorganize and to represent the semantic structure of the adjective in the SpanishWordNet. It can be proposed for EuroWordNet too. This will allow a more global understanding of its behaviour. This paper is focused on the paradigmatic aspect of the adjective and to study how to represent the syntagmatic aspect, which takes into account the adjective-noun combinations (of compositive, semicompositive and noncompositive character) constitutes one of the tasks to make next. Another one is to determinate the different polysemous types, and to study in depth the linguistic phenomenon of adjective polysemy.

References

1. Alonge, A., N. Calzolari, P. Vossen, L. Bloksma, I. Castellón, M.A. Martí and W. Peters: *The Linguistic Design of the EuroWordNet Database*. Kluwer Academic Publishers. Computers and the Humanities **32** 91–115(1998).
2. Apresjan, Ju. D.: *Regular Polysemy*. Linguistics **142** Mouton The Hague Paris (1974).
3. Bouillon, P. and E. Viegas: *The Description of Adjectives for Natural Language Processing: Theoretical and Applied Perspectives*. Atelier Thematique TALN, Cargèse, 12–17 juillet (1999).
4. Dixon, R. M. W.: *Where Have All the Adjectives Gone?* In: Robert M. W. Dixon, **Where Have All the Adjectives Gone? and Other Essays in Semantics and Syntax**. Berlin-Amsterdam-New-York: Mouton. 1–62 (1982).
5. Fellbaum, C.: *A Semantic Network of English: The Mother of All WordNets*. Computers and the Humanities **32** 209–220 Kluwer Academic Publishers (1998).
6. Lyons, J.: *Semantics*. Cambridge University Press. London (1977).

7. Miller, G., R. Beckwith, C. Fellbaum, D. Gross and K. Miller: *Five Papers on WordNet*. CSL Report **43**. Cognitive Science Laboratory. Princeton University (1990).
8. Nirenburg, S. and V. Raskin: *Ontological Semantics* (not published).
9. Raskin, V. and S. Nirenburg: *Lexical Semantics of Adjectives: A Microtheory of Adjectival Meaning*. NMSU CRL MCCA-95-288. Mahesh, K. and S. Nirenburg. Knowledge-Based Systems for Natural Language Processing. NMSU CRL MCCA-96-296 (1995).
10. Raskin, V. and S. Nirenburg: *Adjectival Modification in Text Meaning Representation*. Proceedings of the 17th International Conference on Computational Linguistics (1996).
11. Rodríguez, H., S. Climent, P. Vossen, L. Bloksma, W. Peters, A. Alonge, F. Bertagna and A. Roventini: *The Top-DoWordNet Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology*. Computers and the Humanities **32** 117–152. Kluwer Academic Publishers (1998).
12. Vossen, P.: *Introduction to EuroWordNet*. Computers and the Humanities. Kluwer Academic Publishers (1998).