Adjectival Semantics and the Legal Domain: a study for ontology improvement

Anderson Bertoldi[1], Rove Luiza de Oliveira Chishman[2], Isa Mara da Rosa Alves[3]

[1,2] Programa de Pós-Graduação em Lingüística Aplicada
Universidade do Vale do Rio dos Sinos (UNISINOS)
São Leopoldo, Brazil
[3] Programa de Pós-Graduação em Lingüística e Língua Portuguesa
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP/Ar)
Araraquara, Brazil
andersonbertoldi@yahoo.com, rove@unisinos.br, isamralves@gmail.com

Abstract

This paper carries out a descriptive study on Portuguese adjectives. Our aim is to describe the semantics of the legal domain adjectives in order to construct an ontology which may improve Information Retrieval Systems. For this, we present an approach based on valency and semantic relations. The ontology proposed here is a first step aiming to build a legal ontology based on top-level concepts.

Keywords: Adjectival semantics, Ontologies, Information retrieval.

1. Introduction

The research reported here presents partial results from a master's thesis project developed in the scope of SEMANTEC, DIRPI and LOIS projects. SEMANTEC project aims at improving natural language processing and information retrieval by exploring semantic information from lexicon. The choice for studying legal domain adjectives and legal ontologies within SEMANTEC project was motivated by an international project of cooperation, the DIRPI project.

DIRPI (Desenvolvimento e Integração de Recursos para Pesquisa de Informação) brought together Brazilian and Portuguese researchers from four different universities: Universidade de Évora, Universidade Nova de Lisboa, Pontifícia Universidade Católica do Rio Grande do Sul and Universidade do Vale do Rio dos Sinos. Because of the DIRPI project we became more involved with the University of Evora and the Portuguese side of the LOIS project. LOIS (Lexical Ontologies for Legal Information Sharing) was an investigation project supported by European Commission within the e-Content program. The aim of LOIS was to build a European legal wordnet for legal information retrieval.

The aim of this work is to evaluate the descriptive potential of adjectives for ontology improvement. In this work we (i) review some computational applications that codify adjectives; (ii) and present a preliminary study on legal domain adjectives, in which we apply a combinatorial description based on valency and semantic relations. In the development of the applied part, we used 6 legal documents extracted from the *Institute of the Technologies of Information in Justice[1]* web site.

---

[1] *www.dgsi.pt*

The same documents had been studied in [1]. The reuse of the same corpus had the objective of reviewing that study, expanding its results. Our corpus presents 66 different adjectives, totaling 132 occurrences. The theoretical approach for semantic description was based on [3].

In order to carry out this research, we chose 5 adjectives, according to frequency and value of information to represent. Our purpose was to describe the semantics of these adjectives within these documents and represent them into the ontology editor Protégé, creating a sample ontology of the legal domain. For the future, the commitment will be the extension of this corpus, a new extraction of adjectives and the creation of an ontology based on top-level concepts with adjectival information.

The paper is structured as follows: in section 2, we briefly present some computational lexicons and ontologies that deal with the representation of the adjectives. In section 3, we deal with the theoretical points that base our analysis. In 4, we present the data of the corpus. Then, in section 5, we present our proposal of representation of these adjectives using the ontology editor Protégé. Finally, section 6 presents the final remarks.

2. Adjectival Semantics in Lexicon and Ontologies

In this section, we present, in general terms, some projects which make the representation of adjectival semantics. In order to do it, we will consider three tools, namely: WordNet, SIMPLE and MikroKosmos. Through this study, we can anticipate some theoretical approaches for adjectival semantic representation. The projects described here can give us some guidelines to semantic description and adjectival representation to be applied in our work, which will be presented in section 5. Below, we deal with the description of the aspects that interest our research.

WordNet is a semantic lexicon of cognitive orientation that groups adjectives into two great classes: descriptive and relational adjectives [6]. Descriptive adjectives are prototypical adjectives. They assign attributes to nouns. Antonymy and synonymy are the main semantic relations attributed to this class. Therefore, an adjective as *fast* has as synonyms *quick*, *rapid,* etc. The antonym of *fast* can be *slow*, which, in its turn, would have *sluggish,* *delaying,* etc. as synonyms. This class also allows gradation.

Relational adjectives compose a large class of adjectives, which are morphologically derived from nouns. This kind of adjective is different from the former because they do not attribute a property to nouns which they are related to and because they are not gradable as well. Most of relational adjectives do not fit comfortably into the synonym – antonym separation proposed by WordNet. When relational adjectives possess bipolar directions, for instance, *mental* and *physical*, in "*physical health*" and "*mental health*", they are grouped with descriptive adjectives. The other relational adjectives are kept in a separate archive, with pointers to nouns that they are related.

SIMPLE is an ontology which divides the adjectives into two great classes: extensional and intensional [4]. The process of division is based on logical semantics. Extensional adjectives attribute a property to the noun they modify; for example, *American* in the following sentence, *Harry Truman was an American president.*

Intensional adjectives do not attribute properties to nouns. These adjectives are, in general, related to modal, emotional and temporal aspects of the predication. It is the case of *current*, in *Bush is the current president of the United States*, in which *current* does not attribute any property to the name; however, it highlights the temporal aspect that is involved in predication. The adjectives can be both extensional and intensional. For example, *poor*, which is used as an extensional adjective in the sentence, *Mário is a poor man*, but it can also be used as an intensional adjective, as can be seen in, *Poor Mário! He lost his job*. In this last case, although *poor* does not attribute a property to the name, it emphasizes an emotional aspect, which is involved in the predication.

Extensional adjectives are divided into intersective and subsective. The former attributes the same property to both the hyponym and the hyperonym of the noun. A *red car* is a *red vehicle*; however, a *small elephant* is not a *small animal*. When an adjective attributes a property to a noun, which is not the same property of the hyperonym, this adjective is called subsective. Then, *small* is a subsective adjective.

MikroKosmos divides the adjectives into five classes, according to their morphological and

semantic features: *scalar gradable*, *non-scalar gradable*, *proper non-scalars*, *event-related non-scalars* and *true relative non-scalar* [5]. In MikroKosmos, gradable and scalar are not synonyms. Gradable adjectives can be used in comparisons, such as, *good*, in *Pedro is good / Pedro is better than Paulo*. On the other hand, scalar adjectives are all those which can receive a value, in a scale from 0 to 1.0, gradable or not. Therefore, *big*, receives a 0.75 as a value, in a range that includes adjectives as *minuscule*, *small*, *medium-size*, *enormous* and *giant*.

From then on, we can understand the division of the adjectives shown above. Scalar gradable adjectives are those that can be used in comparisons and can also receive a numerical value in a scale of values, like *big*. Non-scalar gradable adjectives cannot receive a numerical value, though they can be used in comparisons. In fact, these adjectives are denominal, which accept comparisons, for instance, *administrative*, in the sentence: Maria *has a more administrative style than Paulo*.

Nationality adjectives like Brazilian, Portuguese, French, etc, are considered to be proper non-scalar adjectives. This kind of adjective is neither scalar nor gradable; that is why they seem to have the same characteristics of deverbal and denominal adjectives. Event-related non-scalars are adjectives that come from verbs, like *eatable*, *lovable*, *recyclable*, etc. True-related non-scalar adjectives come from nouns and, different from the non-scalar gradables, such as *administrative*, they cannot be used in comparisons, for example, *civil*, *cardiac*, *ministerial*, etc.

In general terms, WordNet divides adjectives into descriptive and relational, SIMPLE classifies them into intensional and extensional, and MikroKosmos divides adjectives into scalars and non-scalars. Dividing the adjectives into descriptive and relational is a good way to distinguish prototypical adjectives, which attribute properties to beings, from denominal adjectives, which are related to the nouns that originate themselves. However, not all the descriptive adjectives attribute properties to nouns preceding them. This is the case of the Portuguese adjective *doce*. Does the adjective *doce* in *um beijo doce* (a sweet kiss) or *uma bebida doce* (a sweet drink) have the same semantic information that in *água doce* (fresh water)?[2] We cannot claim that

sweet attributes the same property to *water*, once it just classifies *água doce* as a kind of water (in this case, *water from rivers* and lakes) in contrast to *água salgada* (sea water).

A division of adjectives into intensional – extensional could solve this kind of difficulty. In SIMPLE, adjectives are classified according to a logic-formal approach instead of their morphosyntactic features. The adjectives that attribute properties to the noun preceding them are called extensional; if there is no property to be assigned, it is an intensional adjective. This approach is able to deal with delicate semantic phenomena, like intersection and subsection among properties and entities in a hierarchy. For example, *a red car* is *a red vehicle*; there is an intersection of attribute between the hyponym *car* and the hyperonym *vehicle*. However, the same case does not happen to the adjective *big*, since a *big rat* is not a *big animal*.

The division between scalars and non-scalars adopted in MikroKosmos is linked to the ontology purpose, which is automatic translation. As scalar adjectives are not numerous in our corpus; the approach adopted by MikoKosmos is not the most appropriate to deal with the data we have in our corpus. Then, in section 3, we present a proposal that we believe is the most appropriate to treat legal domain adjectives.

3. A Semantic Description for Adjectives

In this exercise, adjectives are classified into *qualifying* and *classifying* [3]. The major reason for this theoretical choice is related to the legal domain adjectival nature. In a previous analysis of the adjectives in our corpus, we verified that most of them had the specific function of separating nouns into classes. For instance, the Portuguese adjectives *penal* and *civil* put the noun *código* into a specific class: legal documents (*código civil* and *código penal*). Therefore, we needed a theoretical approach which helped to describe and explain this phenomenon.

---

[2] *The Portuguese examples shown here do not have a correspondent in English. .The adjective "doce", "sweet"*

in English, can be used to qualify or classify a noun. In Portuguese the noun phrase "água doce", the adjective "doce" can be interpreted in two ways. In one interpretation "doce" is a qualifying adjective; in this case the noun phrase means "sweet water". In the second possibility, as a classifying adjective, it means "fresh water". When it is used as a classifying adjective, "doce" do not attribute a property to noun it precede.

Qualifying adjectives are grouped with nouns and represent a subjective evaluation (see example 1). Classifying adjectives separate nouns into classes; their role is to classify the entities in the world (see example 2).

(1) interpretação *inadmissível* (*inadmissible* interpretation)
(2) código *civil* (*civil* code)

Qualifying adjectives present five main features that distinguish them from classifying adjectives [3]. These features are listed below.

(i) Possibility of attributive and predicative uses:

(3) a. interpretação *inadmissível*
     A interpretação é *inadmissível*. (The interpretation is *inadmissible*.)
   b. código *civil*
     *O código é *civil* (*The code is *civil*.)

(ii) Transformation from adjective into noun:

(4) a. interpretação *inadmissível*
     a *inadmissibilidade* da interpretação (the inadmissibility of the interpretation)
   b. código *civil* /
     *a *civilidade* do código (*the civility of the code)

(iii) Gradation:

(5) a. A interpretação é extremamente inadmissível. (The interpretation is extremely inadmissible.)
     A sua interpretação da lei é a mais inadmissível de todas. (Your law interpretation is the most inadmissible of all.)
   b. *O ataque foi *muito* cardíaco. (*The arrest was very cardiac.)
     *Esse ataque foi *mais* cardíaco que o outro. (*This arrest was much more cardiac than the other one.)

(iv) Association with evaluation verbs:

(6) a. Eu acho essa música muito bonita. (I find this song very beautiful.)
   b. *Eu *acho* este ataque cardíaco. (*I find this arrest cardiac.)

(v) Exclamatory constructions:

(7) a. Que interpretação *inadmissível*! (What an *inadmissible* interpretation!)
   b. *Que código *civil*! (*What a *civil* code!)

All these possibilities are grammatically impossible with classifying adjectives. The examples marked with "*" show the wrong use of Portuguese classifying adjectives.

We assume here that relational adjectives correspond, in general terms, to classifying adjectives. Therefore, we believe that the semantic description adopted by WordNet [6], SIMPLE [4] and Mikrokosmos [5] to describe relational adjectives could be improved by adding circumstantial information, as proposed in [3].

Circumstantial adjectives come from an underlying adverbial function. These adjectives can express *location* (8), *time* (9), *instrument* (10), *causation* (11), etc.

(8) traumatismo *craniano* (*cranial* traumatism) (that occurred in the cranium)
(9) publicação *mensal* (*monthly* publication) (that is done monthly)
(10) conversa *telefônica* (telephone talk) (that happens over the telephone)
(11) doença *tropical* (*tropical* disease) (that is caused by tropical climate)

Adverbial information can be formalized into a legal ontology by semantic relations. The semantic relation that connects *craniano* to *traumatismo* is *ocorre em*. The relation *feita por* connects *mensal* to *publicação*. The relation *por meio de* connects *telefone* to *conversa* and *causada por* connects *tropical* to *doença*. Formalizing this kind of semantic information could extend the *pertain to* or *related to* relations traditionally adopted by ontologies and lexicons to describe the meaning of relational adjectives.

Classifying adjectives into qualifying and classifying adjectives is fruitful to ontology construction. Classifying adjectives emphasize the homonymic relations that organize ontologies. As these adjectives separate nouns into classes, they help us to recognize the different levels of ontological entities.

If classification is the strong point of classifying adjectives, qualifying adjectives can assume predicative position, thus having the same status of verbs. When adjectives are predicators they can be

formalized in the same way as verbs are, that is, they do not necessarily need to be subordinated to a noun. The problem of this theoretical approach [3], which will be presented in section 4, is the difficulty in distinguishing classifying from qualifying adjectives.

So far, we have seen that the distinction between classifying and qualifying adjectives is clear. The syntactic behavior of both adjectives is the most important feature to classify an adjective. But, what can we say about an adjective like *ilegal*? In [2], *ilegal* is a classifying adjective, then it would occur only in attributive position. On the contrary, this adjective occurs only in predicative position, which means that the linguistic theory chosen for this application has some explanatory limitations.

Other data that contradict the syntactic division between classifying and qualifying adjectives adopted in this theory will be explored in section 4. Among the 6 that occurred in predicative position, 4 were classifying, only 2 of them were qualifying. According to the theory applied to this work [3], attributive use only is one of the most important feature of the classifying adjectives. Thus, a question still remains: What differentiates classifying adjectives from qualifying adjectives? Certainly, the syntactic feature is not the most important. The benefits of adopting a classifying-qualifying division to formalize adjectives into a legal ontology crash into the difficulty in grouping them only into two categories. In section 4 we expose some of the difficulties discussed here.

4. Portuguese Adjectives of Legal Domain

In this section we present the analysis of the adjectives. We chose the five most frequent adjectives to do the semantic analysis. According to what we have said before, our corpus is composed by 6 legal documents. They were collected via web from the *Institute of the Technologies of Information in Justice*, a Portuguese legal database. As this experiment is a first step aiming to build a legal ontology based on top-level concepts, the extraction of adjectives was done manually. The syntactic criterion was the first classification. Adjectives were divided into attributive (if they occur only in attributive position) and predicative (if they assume both attributive and predicative position). The syntactic behavior already points to adjectival semantic classification.

In our corpus we found 66 adjectives, considering only the types. These adjectives are distributed in this way: 61 in attributive position; and 6 in predicative position. Only one adjective occurs both in attributive and predicative position. Our theoretical approach to adjectival semantics is based on [3]. According to this approach, adjectives are classified into classifying and qualifying. Seeking a rigorous classification, not based on intuition, we followed [2].

	All Adjectives		Adjectives in Attributive Position		Adjectives in Predicative Position	
	nº	%	nº	%	nº	%
Classifying	39	59,09	36	59,02	4	66,67
Qualifying	27	40,91	25	40,98	2	33,33
Total	66	100	61	100	6	100

Figure 1. Adjectival Semantic Organization

The figure above shows one of the limitations of the theory adopted. According to [3], qualifying adjectives are predicators; classifying adjectives are attributive only. In fact, our corpus suggests the opposite. We can see that classifying adjectives are the most frequent ones in the legal domain: 59.09%, against 40.91% of qualifying. The unexpected is that classifying adjectives are more frequent even in predicative position. Considering the predicative position, 66.67% of adjectives were classifying, against 33.33% of qualifying adjectives.

The five most frequent adjectives were selected to the applied part of our research. They are: *civil*, *legal*, *supremo*, *patrimonial* and *sumário*. Three of them are classifying and two are qualifying. All of them occur in attributive position. The semantic analysis for ontological representation is presented below.

Civil occurred 27 times in our corpus in 4 different contexts: *Código civil* (17 occurrences); *Código processual civil* (05 occurrences); *Responsabilidade civil* (04 occurrences); and *Construção civil* (01 occurrence). Each one of these contexts presents distinct semantic information, as can be seen in figure 2.

Civil: [Classifying] 1. que diz respeito às relações dos cidadãos entre si 2. relativo ao cidadão considerado em circunstancias particulares dentro da sociedade

Context 1: Código *civil*
Hyperonym: *código*.
Co-hyponyms: código *penal*; código *comercial*; código *administrativo*; código *processual civil*.
Civil is related to *cidadão*.

Context 2: Código processual *civil*
Hyperonym: *código*.
Co-hyponyms: código *penal*; código *comercial*; código *administrativo*; código *civil*.
Civil is related to *cidadão*.

Context 3: Responsabilidade *civil*
Hyperonym: *responsabilidade*.
Co-hyponyms: responsabilidade   *objetiva*; responsabilidade *subjetiva*.
Civil is related to *cidadão*.

Context 4: Construção *civil*
Hyperonym: *construção*.
Co-hyponyms: construção *naval*.
Civil is related to *cidadão*.

Figure 2. Relations to Adjective *Civil*.

The adjective *legal* appears in second place. With 08 occurrences in 4 contexts: *Taxa legal* (03 occurrences); *Juros legais* (02 occurrences); *Presunção legal* (02 occurrences); and *Regime legal* (01 occurrence).

Legal : [Classifying] 1. prescrito pela lei.

Context 1: Taxa *legal*
Hyperonym: *taxa*.
Co-hyponyms: taxa *ilegal*; taxa *abusiva*.
Legal is related to *lei*.

Context 2: Juros *legais*
Hyperonym: *juros*.
Co-hyponyms: juros *ilegais*; juros *abusivos*.
Legal is related to *lei*.

Context 3: Presunção *legal*
Hyperonym: *conclusão*.
Co-hyponym: conclusão *de direito*.
Legal is related to *lei*.

Context 4: Regime *legal*
Hyponym: regime *legal* do arrendamento urbano; regime *legal* da adoção.
Legal is related to *lei*.

Figure 3. Relations to Adjective *Legal*.

The adjective *supremo* occurred 7 times, in only one context, *Supremo Tribunal de Justiça*, which is a proper name.

Supremo [Qualifying] 1. extremo 2. último, derradeiro 3. máximo 4. do mais alto grau.

Context: *Supremo* Tribunal de Justiça
Hyperonym: *tribunal*.
Co-hyponyms: tribunal *da primeira instância*; tribunal *da relação*.
Supremo qualifies *tribunal*.

Figure 4. Relations to Adjective *Supremo*.

The adjective *patrimonial* occurred 6 times in our corpus, in 2 contexts: *danos não patrimoniais*, with 4 occurrences; and *danos patrimoniais*, with 2 occurrences.

Patrimonial [Classifying] 1. de ou relativo a patrimônio 2. que se baseia na propriedade.

Context 1: Danos não *patrimoniais*
Hyperonym: *danos*.
Synonym: danos *morais*
Opposites: danos *físicos*; danos *materiais*.
Co-hyponyms: danos *patrimoniais*; danos *indenizáveis*.
Patrimonial is related to *patrimônio*.

Context 2: Danos patrimoniais
Hyperonym: *danos*.
Synonyms: danos *físicos*; danos *materiais*.
Opposites: danos não *patrimoniais*; danos *morais*.
Co-hyponyms: danos *não patrimoniais*; danos *indenizáveis*.
Patrimonial is related to *patrimônio*.

Figure 5. Relations to Adjective *Patrimonial*.

The adjective *sumário* has 4 occurrences in our corpus.

Sumário: [Qualifying] 1. formulado sem formalidades 2. simples; preliminar

Context 1: Ação *sumária*
Hyperonym: *processo declarativo comum*.
Synonyms: forma *sumária*; processo *sumário*.
Co-hyponyms: ação *ordinária*; ação *sumaríssima*.
Sumário qualifies *ação*.

Context 2: Forma *sumária*
Hyperonym: *processo declarativo comum*.
Synonyms: ação *sumária*; processo *sumário*.
Co-hyponyms: forma *ordinária*; forma *sumaríssima*.
Sumário qualifies *forma*.

Context 3: Processo *sumário*
Hyperonym: *processo declarativo comum*.
Synonyms: ação *sumária*; forma *sumário*.
Co-hyponyms: processo   *ordinário*;   processo *sumaríssimo*.
Sumário qualifies *processo*.

Figure 6. Relations to Adjective *Sumário*.

These 4 occorences appeared in 3 contexts, but with meaning related to law. They are: *Ação sumária* (01 occurrence); *Forma sumária* (01 occurrence); and *Processo sumário* (01 occurrence)

In section 5 we deal with adjective formalization. We chose Protégé for this exercise.

5. Legal Domain Adjectives in Protégé Representation

Protégé is an integrated software tool used by system developers and domain experts to develop knowledge-based systems. Its architecture allows ontology edition and its conversion to be used in various knowledge-based applications. The main advantage of using this tool is the possibility to convert data into languages like Ontology Web Language (OWL); it makes the ontology compatible with the Semantic Web. Using this editor makes our proposal more relevant, since it organizes such knowledge and allows it to be reused. In this exercise we start from an ontology proposed to legal events in [1], improving it with adjectival information.

Protégé architecture and ontology organization is based on class and subclass relation. We created the super class *Propriedades* (Properties) to group adjectives. *Propriedades* is considered as an abstract role class in this tool, since it is not applied to real data corpus, but it is just used to organize data in the ontology. The same occurs to categories *Qualificadores* (qualifying adjectives) and *Classificadores* (classifying adjectives), which are considered *abstract classes*. Figure 7 illustrates adjectival hierarch in Protégé.

The hyponymic relation is, once more, approached by classifying adjectives inclusion. Nominal units composed by noun + classifying adjective, like *civil code*, *civil responsibility*, are entered in Protégé as subclasses of nouns. Besides hyponymic nouns, classifying adjectives can establish other relations. We mention, for instance, as a SLOT, the relation *is_related_to* to explain the link between the adjective *civil* (civil) and the noun *cidadão* (citizen).
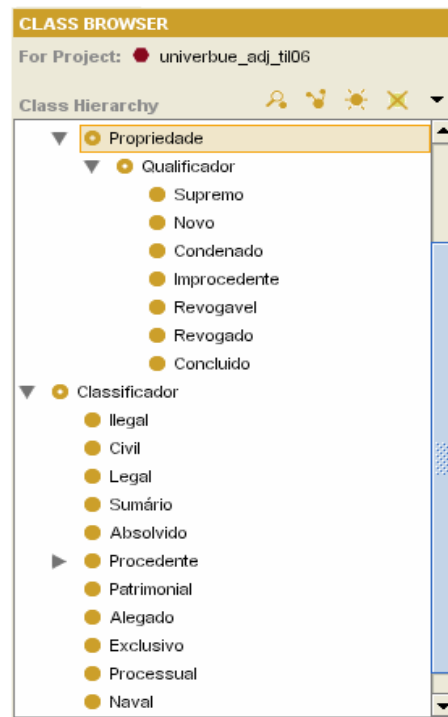

Figure 7. Adjectival Hierarch

Qualifying adjective semantics allows typical relations to adjectives, such as synonym and antonym. The adjectives *supreme* and *summary* are good examples of these possibilities.

6. Final Remarks

Even though the theoretical study we presented in this paper has a preliminary role, it made us notice important information linked to adjectival semantics. We realized that choosing the best approach and taxonomy according to research proposal is not an easy task, since there are several approaches available and each one has a different focus on language representation. We explored only one perspective that starts from a syntactic feature and subdivides adjectives into classifying and qualifying. However, there are other distinctions which have to be investigated: the gradable and ungradable distinction, intensionality and extensionality, and adjectival polissemy. These research points have to be addressed when we extend this work.

Regarding the applied aims of this research, we

managed to develop a first representation for adjectives in Protégé; however, it is an initial enterprise and needs to be expanded. It was possible to represent the characteristics of each subclass: classifying and qualifying adjectives. As we used a small sample of adjectives, not all the relations illustrated in section 3 could be explored in the applied part of this research.

Concluding, we emphasize, as a positive point, the power of Computational Lexical Semantics to build computational tools.

Acknowledgements

References

[1] Isa Mara da Rosa Alves. *O uso da semântica verbal em sistemas de extração de informação: a construção de uma ontologia de domínio jurídico*. Dissertação de Mestrado. São Leopoldo: UNISINOS, 2005.

[2] Francisco da Silva Borba. *Dicionário de usos do português do Brasil*. São Paulo: Ática, 2002.

[3] Francisco da Silva Borba. *Uma gramática de valências para o português*. São Paulo: Ática, 1996.

[4] Ivonne Peters e Win Peters. The treatment of adjectives in SIMPLE: theoretical observations. *Proceedings of LREC2000*, Athens, 2000. http://www.ub.es/gilcub/SIMPLE/reports/papers /Adj_Peters.pdf

[5] Victor Raskin e Sergei Nirenburg. An Applied Ontological Semantic Microtheory of Adjective Meaning for Natural Language Processing. *Machine Translation*, 13(2-3):135-227, 1998.

[6] Katherine J Miller. Modifiers in WordNet. In: *WordNet:* an electronic lexical database. Christiane Fellbaum (ed.). Cambridge, MA: MIT Press, 1999.