Adjectives in WordNet.PT

Sara Mendes
Group for the Computation of Lexical and Grammatical Knowledge
Center of Linguistics, University of Lisbon
Avenida Professor Gama Pinto, 2
1649-003 Lisboa, Portugal
sara.mendes@clul.ul.pt

Abstract

Most authors agree that adjective semantic analysis and representation is far from being a trivial issue. Since the semantic organisation of adjectives seems to be unlike that of nouns and verbs, as noted by Fellbaum et al (1993) and Miller (1998), this paper focuses on the encoding of adjectives in wordnets. We discuss the strategies used in WordNet.PT. Our proposal aims at mirroring adjectives definitional features in the database, allowing adjective classes to emerge from the relations expressed in the network. In order to do so, we use some of the semantic relations introduced in the Princeton WordNet, but we also propose some new pointers.[1]

1 Introduction

Adjectives, perhaps more than any other part-of-speech, show a non trivial semantic behaviour, namely in what concerns sense change depending on the linguistic context. Therefore, adjective semantic analysis and representation is far from being a trivial issue. Fellbaum et al (1993) and Miller (1998) state that the semantic organisation of adjectives in the Princeton WordNet (Miller (1990); Fellbaum (1998)) is unlike that of nouns and verbs, as these modifiers do not show a hierarchical organisation in WordNet.

Encoding adjectives in wordnets is the main issue discussed in this paper. We present the strategies put at work in WordNet.PT in order to mirror, in the database, adjectives main features, namely those that are on the basis of the definition of adjective classes. We will show that it is possible to make these classes emerge from the relations expressed in the network without having to encode them artificially by listing adjectives in separate files, as it is done in the Princeton WordNet. In order to do so we use some of the semantic relations introduced in the Princeton WordNet, but we also propose some new pointers. In section 1 we briefly describe some of the main adjective classes. Section 2 discusses the way adjectives have been encoded in wordnets, in particular in the Princeton WordNet and in Germanet, and section 3 presents our proposal for adjective encoding in WordNet.PT.

2 Adjective Classifications

There are several possible classifications of adjectives - semantic based classifications, syntactic based classifications, classifications regarding the relation holding between the adjective and the modified noun, and so on. Nonetheless, as our work on this issue progresses, it has become clear that only a combination of syntactic and semantic criteria can offer interesting insights concerning adjective linguistic behaviour. In fact, the main interest of making classifications is to identify relevant common features, characterising lexical items. In this section we will make a brief description of the adjective classifications we consider relevant for the sake of our work.

Regarding the way adjectives relate to the noun they modify, we consider two classes: property ascribing adjectives (in (1)), which add a new restriction to the properties the modified nominal introduces; and reference modifying adjectives (in (2)), which behave like a semantic operator, taking the reference of the modified nominal as its argument[2].

(1) o livro azul
'the blue book'

(2) o diamante falso
'the fake diamond'

While property ascribing adjectives, such as *azul* (blue), attribute properties to the modified nouns, reference modifying adjectives like *falso* (fake) show how a concept applies to a particular object, i.e. these adjectives do not deal with real objects nor with referential objects, they deal with concepts.

We will focus on property ascribing adjectives, since reference modifying adjectives constitute a closed class of a relatively small number of adjectives with very particular properties, which makes them somewhat close to semantic operators.

Demonte (1999) proposes a classification of property ascribing adjectives based on their intrinsic meaning. This classification combines syntactic and semantic criteria to determine which adjectives belong to which class. Two subclasses are considered: descriptive adjectives and relational adjectives. Each of these classes displays specific semantic and syntactic properties. Below we briefly present the distinguishing properties of these two adjective classes relevant for Portuguese.

[1]This research was supported by Fundação para a Ciência e a Tecnologia (grant SFRH/BD/8524/2002).

[2]This distinction between *property ascribing adjectives* and *reference modifying adjectives* is basically equivalent to the one used in the SIMPLE project (*extensional* vs. *intensional adjectives*, following Chierchia & McConnel-Ginet (1990)) to address the semantics of adjectives. This distinction is also included in the EAGLES recommendations for a semantic typology of adjectives.

Petr Sojka, Key-Sun Choi, Christiane Fellbaum, Piek Vossen (Eds.): GWC 2006, Proceedings, pp. 225–230.

In Portuguese, descriptive adjectives can occur both in attributive and predicative contexts, while relational adjectives occur almost exclusively in attributive contexts[3]. Both prenominal and postnominal positions are possible for descriptive adjectives in attributive contexts. Relational adjectives, on the contrary, can only occur in postnominal position, as shown in the examples below, adapted from Casteleiro (1981). Finally, descriptive adjectives are gradable, i.e. they can co-occur with degree adverbs, which is not the case for relational adjectives.

(3) a. Adoro as paisagens calmas.
'I love the landscapes calm'
b. Adoro as calmas paisagens.
'I love the calm landscapes'
c. Adoro as paisagens muito calmas.
'I love the landscapes very calm'
d. Adoro as paisagens que são calmas. (As paisagens são calmas.)
'I love the landscapes which are calm' ('the landscapes are calm')

(4) a. Adoro as casas rurais.
'I love the houses rural'
b. *Adoro as rurais casas.
'I love the rural houses'
c. *Adoro as casas muito rurais.
'I love the houses very rural'
d. ?*Adoro as casas que são rurais. (*As casas são rurais.)
'I love the houses which are rural' ('the houses are rural')

As the criteria presented above are not always sufficient to make a clear-cut distinction between relational and descriptive adjectives, Demonte (1999) proposes some additional criteria in order to determine which adjectives belong to each of these classes in a more accurate way: their occurrence in comparative structures, and the formation of polarity systems.

(5) a. O sabor desta laranja é mais doce do que o daquela.
'this orange taste is sweeter than that one's'
b. o rapaz alto / o rapaz baixo
'the tall boy / the short boy'

(6) a. *Este sabor é mais mineral do que aquele.
'this taste is more mineral than that one'
b. o sabor mineral / *o sabor amineral
'the mineral taste / the amineral taste'

But most of all, and besides all the syntactical contrasts presented above, there is a clear contrast in the way these two adjective classes relate to the noun they modify. Descriptive adjectives, on the one hand, ascribe a single property, setting a value for an attribute. Relational adjectives, on the other hand, introduce a set of properties.

(7) o prédio alto
'the high building'

(8) a indústria alimentar
'the alimentary industry'

Looking at (7) and (8), we see that, while *alto* (high) sets the value of the **height** attribute of *prédio* (building) to **high**, *alimentar* (alimentary) does not ascribe a single property, but a set of properties to *indústria* (industry). Moreover, this set of properties corresponds to the main features describing another noun, *alimento* (food) in the example above. In fact, the way in which properties are ascribed to the modified nouns in (7) and in (8) are quite different. Ascribing a singular property usually corresponds to an incidence relation of this property in the nominal referent[4], while ascribing sets of properties usually entails more complex and diversified semantic relations. Actually, relational adjectives establish a link between the modified noun and other domains which are exterior to them. Let us look at (9) and (10).

(9) o vestido vermelho
'the red dress' (there is an X which is a dress and which is a red object)
$\text{dress}(X) \wedge \text{red}(X)$

(10) o cais marítimo
'the sea quay' (there is an X which is a quay and which has a relation R1 with the sea)
$\text{quay}(X) \wedge \text{R1}(X,\text{sea})$

Ideally, the syntactic and semantic properties we have been discussing would be encoded in lexical models such as wordnets. The SIMPLE project also addresses the semantics of adjectives, identifying a set of features that are relevant for classifying and describing adjective behaviour. Adjectives are organised in terms of semantic fields and then associated with the partially instantiated feature structure corresponding to its class. Nonetheless, as noted by Peters & Peters (2000), even though similarities exist, "adjectives belonging to the same semantic class may differ from each other in numerous ways". Thus, the classes proposed in SIMPLE are not homogeneous.

In the following sections, we will see how adjectives have been encoded in wordnets, namely in the Princeton WordNet ("the mother of all wordnets", in Fellbaum's (1998) words) and in GermaNet (Hamp & Feldweg (1997)), and we will present the options taken in WordNet.PT.

3 Adjectives in wordnets

In the Princeton WordNet (Miller (1990); Fellbaum (1998)), descriptive and relational adjectives are distinguished, first of all, by being encoded in separate files, and second, by the relations holding between synsets.

[3] As shown in (4), predicative contexts with relational adjectives are generally ruled out. Nonetheless, some specific contexts, like contrastive contexts, for instance, seem to license predicative uses of relational adjectives :
(I) As próximas eleições são autárquicas, não são presidenciais.
'next election will be autarchic, not presidential'

[4] Some cognitive semanticists like Gardenförs, on the basis of the analysis of expressions such as *red skin* or *red wine*, would like to deny the analyticity of expressions like *red dress*, in (9), as the conjunction of two predicates, arguing for a refined relational treatment of the property **red**. Recognising that the problematic data correspond to crystallised expressions, allows us to keep the analyticity hypothesis addressing adjectives like *red*.

The semantic organisation of descriptive adjectives is entirely different from that of nouns and verbs. Descriptive adjectives are organised in clusters of synsets, each cluster being associated by semantic similarity to a focal adjective which is linked with a contrasting cluster through an antonymy relation. Therefore, antonymy is the basic semantic relation used in the Princeton WordNet to encode descriptive adjectives. This is motivated if we recognise that these adjectives function regards the expression of attributes, and that an important number of attributes are bipolar. Besides, this cluster organisation of adjectives seems to mirror psychological principles (Miller (1998)).

Nothing like the hierarchies of hyponymic (in the semantic organisation of nouns in WordNet) and troponymic relations (in the semantic organisation of verbs) is available for adjectives. Even if, as the example in Fig. 1 shows, it is possible to find some small local hierarchies, the hyperonymy/ hyponymy pointer is far from being the crucial semantic relation in the organisation of adjectives in relational lexical databases such as wordnets.

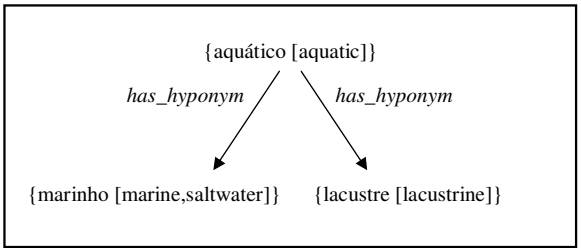

Figure 1: Hyponymy relations of the synset {aquático[aquatic]}: local hierarchical organisation of adjectival synsets

In GermaNet, Hamp & Feldweg (1997) discuss this organisation of adjectives. They abandon the cluster organisation of the Princeton WordNet in favour of a hierarchical structuring of adjectives, arguing for a uniform treatment of all POS. Even though taxonomic chains of adjectives yield rather flat in comparison to those of nouns and verbs, GermaNet authors claim to derive more structural information from these small taxonomies than from clusters, as they seek to eliminate what they consider to be the 'rather fuzzy concept of indirect antonyms'.

Even though the concept of indirect antonymy is not completely clear, it is not obvious to us why this fact should entail that adjectives must show a hierarchical organisation instead, as claimed by the authors of GermaNet. Further below we will discuss the options made in WordNet.PT in order to encode descriptive adjectives.

Relational adjectives, on the other hand, do not have antonyms. Therefore, they cannot be organised in opposite clusters. As pointed out by Levi (1978), the intrinsic meaning of these adjectives is something along the following lines: 'of, relating/pertaining to, associated with' some noun. The way these adjectives are encoded in the Princeton WordNet mirrors this as it links relational adjectives to the nouns they relate to.

In GermaNet a distinct treatment of relational and descriptive adjectives is abandoned, as the distinction between these two classes is considered to be 'not at all clear'. Nonetheless, the Princeton WordNet strategy for distinguishing between different adjective classes is maintained: listing lexical items in different files[5].

As we have pointed out in the previous section, even if the distinction between these two classes is not always clear-cut, testing adjectives against the set of syntactic and semantic criteria presented in section 1 allows us to distinguish descriptive from relational adjectives. In WordNet.PT we consider that this distinction can be mirrored in the database via the semantic relations expressed in the network, adjective listing in different files not being therefore necessary. Such an approach has the advantage of coping with adjective representation in a lexical semantic database without using strategies external to the lexical model, such as *a priori* semantic classes or separate files corresponding to different classes. We will argue for this option in the following sections.

4 Encoding adjectives in WordNet.PT

WordNet.PT (Marrafa (2001); Marrafa (2002)) is a lexical conceptual database for European Portuguese, developed in the general approach of EuroWordNet (Vossen (1998)). In this section we will present and discuss the options made in WordNet.PT with regard to the encoding of adjectives in the database.

4.1 Structuring relations

Following Demonte (1999), we have distinguished descriptive from relational adjectives, showing that they differ in terms of their intrinsic meaning, as well as with regard to their syntactic and semantic behaviour. In this section we will present our approach to reflect these adjectives main properties in WordNet.PT.

In WordNet.PT, all adjectives are encoded in the same file and the classification of adjectives emerges from the relations encoded in the network.

As shown above, descriptive adjectives typically introduce an incidence relation of a single property in the nominal referent. Put somewhat simplistically, descriptive adjectives ascribe a value of an attribute to a noun. In WordNet.PT we introduce a new pointer linking each descriptive adjective to the attribute it modifies. This pointer links nouns to adjectives as follows:

(11) a. alto *caracteriza quanto a* altura
'tall *characterises with regard to* height'
b. altura *é caracterizável por* alto
'height *can be characterised by* tall'

[5]GermaNet classifies the adjectives into 15 semantic classes, following the classes proposed by Hundsnurscher and Splett (1982), with some minor changes: perceptional, spatial, temporality-related, motion-related, material-related, weather-related, body-related, mood-related, spirit-related, behaviour-related, social-related, quantity-related, relational and general adjectives. One special class is added for pertainyms.

This semantic relation is very close to the *is a value of*/*attributes* pointer used in the Princeton WordNet. In order to make it more straightforward to the common user, we changed this relation label in WordNet.PT. Just as *alto* (tall), in the example above, an adjective such as *elevado* (high) or *baixo* (short), will also be linked to *altura* (height). Thus, instead of linking adjectives amongst themselves by a similarity relation, we argue that it is more informative and intuitive to link all adjectives modifying the same attribute to the lexicalisation of this attribute. In this way, without having to encode it directly – and somewhat artificially – in the network, we obtain the cluster effect, Fellbaum et al (1993) and Miller (1998) argue to be the basis of the organisation of adjectives.

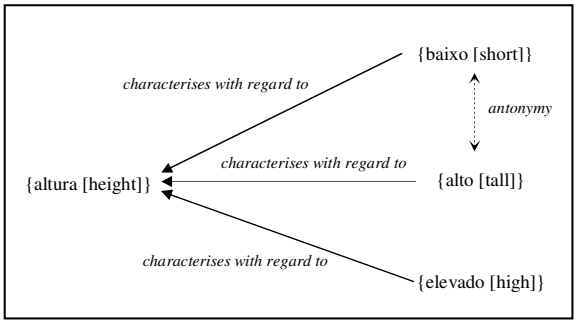

Fig. 2 – Adjective cluster around the attribute synset

Figure 2: Adjective cluster around the attribute synset {altura [height]}: representation in WordNet.PT

As shown by results obtained with word association tests, *antonymy* is also a basic relation in the organisation of descriptive adjectives. Nonetheless, as Fellbaum et al (1993) points out, this relation does not correspond to conceptual opposition. *Antonymy* holds between word forms and not synsets, i.e. word meanings. Here we will argue that conceptual opposition does not have to be explicitly encoded in a wordnet, as it is possible to make it emerge from the combination of the *synonymy* relation and the *antonymy* relation. Let us look at the example presented in Fig. 3.

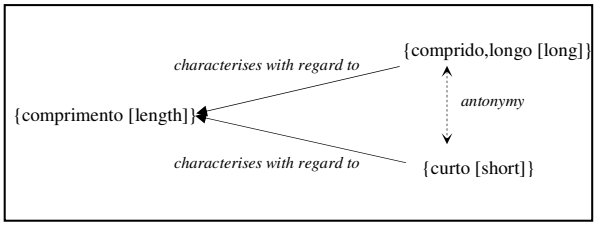

Figure 3: Conceptual opposition between *curto* (short) and *longo* (long): representation in WordNet.PT

In this case we have the antonym pair *curto* (short) / *comprido* (long), both linked to the noun *comprimento* (length) via the *characterises with regard to* relation. Also, *comprido* (long) belongs to the same synset as *longo* (long), i.e. they are synonyms. The expression of these relations entails that *longo* (long) is a conceptual opposite of *curto* (short), but it is not its antonym.

Also, we do not use the *indirect antonymy* relation in WordNet.PT. In fact, as we pointed out in section 2, the concept of *indirect antonymy* is not very clear. Besides, it is used in the Princeton WordNet to define adjective clusters, whereas in WordNet.PT we can cope without this relation, as we obtain the cluster effect via the *antonymy* and the *characterises with regard to*/*can be characterised by* relations. This strategy is, in fact, more intuitive and descriptively adequate, since many attributes are not bipolar, but can take many values along a continuum.

At the beginning of this section we stated that in WordNet.PT all adjectives were encoded in the same file. On the basis of such an option is the observation that the distinction between descriptive adjectives and relational adjectives is such an important and obvious one that the way it is encoded in lexical models such as wordnets should allow these classes to emerge. Moreover, and since not all adjectives behave exactly in the same way with regard to the criteria presented in section 1, it is more adequate to keep it a continuum than to encode them statically in different files.

As mentioned above, relational adjectives ascribe sets of properties to the noun they modify. Being property ascribing adjectives, along with descriptive adjectives, relational adjectives usually entail more complex and diversified semantic relations between the set of properties they introduce and the modified noun than descriptive adjectives. In fact, relational adjectives establish a link between the modified noun and a domain exterior to it, as the intrinsic meaning of these adjectives generally points to the denotation of a noun. In WordNet.PT, as in the Princeton WordNet, we will encode this relation through the *is related to* pointer.

(12) a. aquático *está relacionado com* água
'aquatic *is related to* water'
b. água *está relacionado com* aquático
'water *is related to* aquatic'

Being, to some extent, 'pointers' to an exterior domain, typically lexicalised by a noun, relational adjectives usually do not have antonyms. Nonetheless, these adjectives sometimes have semantic opposites, expressed in the database via the *near-antonymy* relation. This relation should not be confused with the *indirect antonymy* relation used in the Princeton WordNet, as we use it to encode different phenomena in the database[6].

(13) terrestre *é quasi-antónimo de* aquático **and**
aquático *é quasi-antónimo de* terrestre
'terrestrial *is near_antonym of* aquatic **and**
aquatic *is near-antonym of* terrestrial'

Thus, descriptive adjectives in WordNet.PT are encoded via *antonymy* and *characterises with regard to*/*can be char-*

[6]Ongoing research is being developed regarding this semantic relation. Preliminary results seem to indicate a dependency between human perspective and lexical opposition, as speakers oppose, for instance, *terrestre* (terrestrial) to *aquático* (aquatic) and *aéreo* (aerial), on the one hand, but do not oppose *aquático* (aquatic) to *aéreo* (aerial), on the other.

acterised by relations, while relational adjectives are linked in the database through the *is related to* pointer. These semantic relations allow us to encode the basic characteristics of these adjectives in the database, on the one hand, and, on the other, as they mirror the distinctive and contrastive linguistic behaviour of these adjective classes, they make it possible for us to cope with the distinction between descriptive and relational adjectives without having to decide beforehand to which class each adjective belongs. Rather, membership to these classes emerges from the relations expressed in the database.

4.2 Additional relations

In the previous section, we have introduced the basic structuring relations used in WordNet.PT to encode adjectives. We will now discuss some additional semantic relations. In WordNet.PT, we have introduced a new pointer which allows us to encode typical and salient characteristics of nouns expressed by adjectival synsets. This relation is stated as follows:

(13)a. carnívoro *é característica de* tubarão
'carnivorous *is a characteristic of* shark'
b. tubarão *tem como característica* carnívoro
'shark *has as a characteristic* carnivorous'

This relation allows us to encode the most salient – definitional – features, contributing to making the definition of the sense of each synset clearer and richer. Many would consider this information to be outside the domain of lexical knowledge. We do not intend to make any strong claims about the limits of lexical knowledge and meaning here, as this remains an open question in the Lexical Semantics domain. What we claim is the distinctive role this relation may play in a lexical database like WordNet.PT. For instance, the possibility of negating this relation allows us to encode contrasting definitional features of certain nouns in the database, as shown, for instance in Fig. 4.

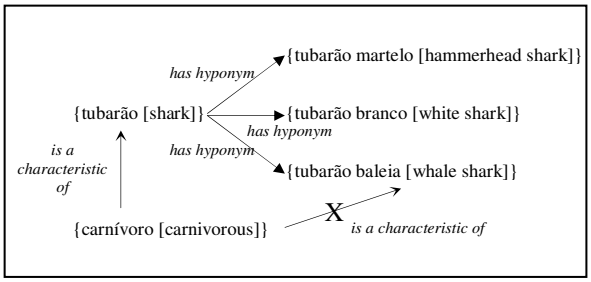

Figure 4: ***is a characteristic of*** relation around the synset {carnívoro[carnivorous]}: representation in WordNet.PT

In fact, what is prototypical of *tubarão* (shark) is that it is *carnívoro* (carnivorous), nonetheless, a hyponym of *tubarão* (shark) such as *tubarão baleia* (whale shark) has as one of its distinctive features the fact that it is not *carnívoro* (carnivorous), as this fact differentiates it from its sisters. Being able to express this in the database is therefore interesting, not only because it meets speakers' intuitions, but also because it can constitutes crucial information for many wordnet-based applications, namely all those using inference systems.

As mentioned in section 2, in GermaNet, adjectives are split up in 15 semantic classes, generally corresponding to the semantic domains adjectives apply to. The relation we have just presented also makes it possible to deduce semantic domains from the database, without having to determine them *a priori*. In fact, if for instance we analyse the *is a characteristic of* links of an adjective like *carnívoro* (carnivorous), we verify that it typically modifies living things. Therefore, if synsets are encoded in this fine-grained way, it will be possible to deduce the typical semantic domains of application of adjectives. A study on the classes and semantic domains emerging from the relations expressed in the database is still exploratory. Future work should include an extraction of these classes and a comparative study of the results with the classes defined both in the SIMPLE project and by Hundsnurscher & Splett (1982).

4.3 Statistics

Since the linguistic resources available for Portuguese are not suitable enough for automatically building a wordnet, WordNet.PT has been developed mainly on the basis of manual work. The first phase of the project (1999-2003), encoded a selection of semantic domains covering concepts with high productivity in daily life communication, mostly nouns. In mid 2004 a new phase of the project started, aiming at an integrated increment of the database, concerning several new semantic domains and all the main POS, including adjectives[7].

Table 1 presents the number of adjectives, as well as the total number of entries, encoded in the last version of WordNet.PT.

Table 1: Statistics on the last version of WordNet.PT

	adjectives	**total**
number of synsets	1034	12630
number of variants	1260	15438

Final remarks

We have shown how adjectives have been encoded in WordNet.PT and how the strategies used allow us to deduce adjective classes from the relations expressed in the network. To encode adjectives in WordNet.PT, we use some of the semantic relations introduced in the Princeton WordNet, but we also propose some new pointers.

Our approach differs from the Princeton WordNet in that we do not distinguish between different adjective classes by encoding them in different files; we do not use the indirect antonymy relation of the Princeton WordNet, basing the cluster organisation of adjectives in WordNet.PT on a set of relations; and artificial antonyms are not introduced in the database.

[7] For more detailed information on the present status of the WordNet.PT project, see Marrafa et al (2005).

Acknowledgements

I wish to thank Palmira Marrafa and Raquel Amaro for all their reviews and insightful comments. I also wish to thank the reviewers for their suggestions for improvement of this work.

References

Adjectives in GermaNet, `http://www.sfs.nphil.uni-tuebingen.de/lsd/Adj.html`.

Casteleiro, J. M. (1981) *Sintaxe Transformacional do Adjectivo – regência das construções completivas*, Instituto de Investigação Científica, Lisboa.

Chierchia, G. & S. McConnel-Ginet (1990) *Meaning and Grammar: an Introduction to Semantics*, Cambridge, MA: The MIT Press.

Demonte, V. (1999) *El Adjectivo: classes y usos. La posición del adjectivo en el sintagma nominal*, in Bosque, I. & V. Demonte (orgs.) "Gramática Descriptiva de la Lengua Española", Madrid: Espasa, vol. I.

EAGLES Lexicon Interest Group (1998) *Preliminary Recommendations on Semantic Encoding Interim Report*.

Fellbaum, C., D. Gross & K. Miller (1993) *Adjectives in WordNet*, in Miller et al. "Five Papers in WordNet", Technical Report, Cognitive Science Laboratory, Princeton University, pp. 26–39.

Fellbaum, C. (1998) *A Semantic Network of English: The Mother of all WordNets*, in Vossen, P. (ed.) "EuroWordNet: A Multilingual Database with Lexical Semantic Networks", Dordrecht: Kluwer Academic Publishers, pp. 137–148.

Hamp, B. & H. Feldweg (1997) *GermaNet – a Lexical Semantic Net for German*, in Proceedings of ACL workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications, Madrid.

Levi, J. N. (1978) *The Syntax and Semantic of complex nominals*, New York: Academic Press.

Marrafa, P. (2001) *WordNet do Português: uma base de dados de conhecimento linguístico*, Lisboa: Instituto Camões.

Marrafa, P. (2002) *Portuguese WordNet: general architecture and internal semantic relations*, D.E.L.T.A., 18.

Marrafa, P., R. Amaro, R. P. Chaves, S. Lourosa, C. Martins & S. Mendes (2005) *WordNet.PT new directions*, paper accepted at the GWA06 Conference.

Miller, G. A. (1990) *WordNet: an on-line Lexical Database*, in Special Issue of International Journal of Lexicography, Vol. 3, n^o 4.

Miller, K. J. (1998) *Modifiers in WordNet*, in Fellbaum, C. (ed.) "WordNet: an electronic lexical database", Cambridge, MA: The MIT Press, pp. 47–68.

Peters, I. & W. Peters (2000) *The Treatment of Adjectives in SIMPLE: Theoretical Observations*, in Proceedings of LREC 2000.

Vossen, P. (1998) (ed.) *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Dordrecht: Kluwer Academic Publishers.

Vossen, P. (1999) *EuroWordNet General Document*, University of Amsterdam.