IJCLA VOL. 4, NO. 1, JAN-JUN 2013, PP. 11–27
RECEIVED 07/12/12 ACCEPTED 11/01/13 FINAL 14/06/13

Increasing Density through New Relations and PoS Encoding in WordNet.PT

RAQUEL AMARO,[1] SARA MENDES,[1,2] AND PALMIRA MARRAFA[1]

[1] *Universidade de Lisboa, Portugal*
[2] *Universitat Pompeu Fabra*

ABSTRACT

This paper reports research developed in the scope of building a wordnet for Portuguese (WordNet.PT), particularly focusing on the impact the results obtained have in the density of the network of relations and, thus, on its usability for NLP tasks. Following from basic research on different linguistic phenomena and on strategies for modeling them in relational models of the lexicon, the implementation of these results amounts to a richer resource, with new cross-PoS relations and information on event and argument structures, thus crucially contributing to accurately modeling all the main PoS in the database. We also define a way to integrate prepositions in wordnets and discuss the motivations and modeling strategies used to do so. Based on this work, we show how our contributions augment the coverage and the accuracy of WordNet.PT, by increasing the density of the network of relations, thus making it more usable for NLP applications.

KEYWORDS: *wordnets; cross-PoS lexical semantic relations; network density; linguistic coverage*

1 Introduction

WordNet.PT [1] (WN.PT) ([1],[2]), a wordnet for Portuguese developed according to the approach of EuroWordNet (EWN) ([3]), presents distinctive properties concerning the extension of the set of relations used and the strategies employed for attaining lexical coverage.

The initial strategies employed for building WN.PT had as main concern the accuracy of the resulting resource, rather than its extension. This, together with a strong focus on research, motivated the option for the manual selection, description and encoding of all WN.PT data, resulting in a smaller but much more reliable lexical resource, compared with automatically and semi-automatically constructed databases. The enlargement of the database has followed the semantic domains approach, involving the integration of lexical items from different PoS, which motivated the need for enriching the model with more information, namely information on selectional properties and new PoS, and for encoding new relations, in particular cross-PoS relations.

In this paper we present research developed in the scope of building WN.PT, particularly focusing on the impact the results obtained have in the density of the network and, thus, on its usability for NLP tasks. In Section 2 we present and discuss research on different linguistic phenomena, particularly regarding new relations, with a special focus on cross-PoS relations, introduced in WN.PT to model all the main PoS in the database and to encode information on argument structures. Section 3 is dedicated to the impact the contributions and modeling strategies implemented in WN.PT have on the density of the network. Section 4 concludes this paper with our final remarks and considerations regarding future work.

2 WordNet.PT Relations: Innovation and Coverage

WN.PT adopts almost entirely the set of relations defined in EWN, exception being the DERIVED, PERTAINS and BE IN STATE relations. The first two, besides being morphological relations, are somewhat complementary to the set of relations used in EWN (see [3]:37): the relation DERIVED is only used when there is a morphological link between two synsets and a lexical-conceptual relation already stands; the relation PERTAINS fulfills a void,

[1] http://www.clul.ul.pt/clg/wordnetpt

whenever there is a clear morphological link between a given noun and a given adjective and no other relation clearly stands. Given the lack of a clear stable conceptual relation holding between word forms linked via these relations, we do not use them. BE IN STATE relation is addressed further below.

Interesting enough, these relations are mostly used to relate nouns and adjectives, and can be seen as a way of linking adjectives in the lexicon, given that hyperonymy is not a structuring relation in the case of this PoS and that it does not hold for many adjectival synsets. Fundamental research on event structure and on adjectives developed within WN.PT ([4],[5]) has led to the definition of new semantic relations that further support discarding the ones mentioned above.

2.1 *Adjectives in WordNet.PT*

Following research on adjectives and their modeling in relational lexica ([6],[4]), in WN.PT we defined the following set of relations – CHARACTERIZES WITH REGARD TO, SETS VALUE TO, IS BY DEFINITION RELATED TO, IS A CHARACTERISTIC OF and IS TELIC SUBEVENT OF –, dealing with various complex lexical semantics phenomena regarding adjectives in a general and systematic way.

Although HYPERONYMY is the main structuring relation in wordnets, the semantic organization of adjectives is considerably different ([7]): nothing like the hierarchies of hyponymic relations is available for adjectives. Also, descriptive and relational adjectives[2] differ in terms of intrinsic meaning and of syntactic and semantic behavior (see [4]:53-76 for a detailed discussion on this issue). In WordNet ([8], [9]), descriptive adjectives are organized in clusters of synsets, an organization that mirrors psychological principles of the organization of the lexicon ([7]).

As argued in detail in [6] and [4], descriptive adjectives typically apply an incidence relation of a single property to the denotation of the noun they are related to in context. Put somewhat simplistically, they assign a value of an attribute to a noun. These values can be of different types: Boolean values, scalar values, and values that are neither one nor the other. Encoding this information in wordnets contributes to a more accurate lexical

[2] Wordnets leave out non-restricting adjectives. This option is based on the fact that, as pointed out by different authors ([10], [11], [4], etc.), non-restricting adjectives are a small class with a very particular semantic contribution, closer to semantic operators than to other adjectives.

representation of this PoS. In view of these properties, in WN.PT we use a small set of conceptual relations to represent descriptive adjectives, some of which inherited from the general EWN framework.

In WN.PT we use a semantic relation corresponding to the ATTRIBUTE relation of WordNet to encode the relation between adjectives and attributes, which we label as CHARACTERIZES WITH REGARD TO/IS CHARACTERIZABLE BY for the sake of transparency for non-specialist users:

1. $\{\text{tall}\}_{Adj}$ CHARACTERIZES W.R.T $\{\text{height}\}_N$/$\{\text{height}\}_N$ IS CHARACTERIZABLE BY $\{\text{tall}\}_{Adj}$

Naturally, our claims regarding this semantic relation are not related to the label used to encode it, but rather to the way it is used in WN.PT. In WordNet 3.0[3], in each adjective cluster, only focal adjectives are linked to an attribute. This is counter-intuitive, since the relation holding between *cold* and *temperature* is just as strong as the relation linking *gelid* and *temperature*, for instance. Moreover, the information regarding which attribute is associated to a given adjective – which is just as relevant for focal adjectives as for any other adjective in the cluster – can only be obtained in WordNet 3.0 if a mechanism for navigating the network of relations is developed in order to extract information expressed for focal adjectives and assign it to non-focal adjectives, where appropriate. Another crucial difference regards the relations used for the definition of adjective clusters: in WordNet 3.0 adjectives are associated by semantic similarity to a focal adjective to form clusters, and linked to a contrasting cluster through ANTONYMY. Instead of using a similarity relation that clearly poses problems (see [4]:95 and ff.), we claim that all adjectives ascribing values of the same attribute are linked to this attribute and thus related amongst themselves. This way, without having to encode it directly and somewhat artificially in the network, the clusters argued to be on the basis of the organization of adjectives are obtained: not around pairs of opposite adjectives, but around a common attribute, overcoming the need to define focal adjectives for each cluster.

At the same time that it overcomes the shortcomings mentioned above, it can be argued that this strategy results in loss of information, as the relation between adjectives associated to close values of a given attribute is not explicitly encoded in the network. This is particularly relevant in the case of scalar adjectives, as these adjectives determine values that are organized

[3] http://wordnet.princeton.edu/wordnet/

relatively to each other[4]. [12] state that gradation is in fact a semantic relation organizing lexical memory for adjectives. However, it is not encoded in WordNet because it is rarely lexicalized in English. But besides this individual organization relatively to each other, scalar adjectives are also organized around areas of a scale: typically two extremes and a middle value. Despite the relevance of continuing to develop research on how to model adjective scales ([13]), we start with a coarser modeling of this adjective subclass, which encodes the area of the appropriate scale to which the attribute value assigned by a given adjective belongs. To accomplish this we use a new semantic relation to link the adjective and the lexicalization of the value it assigns, typically an adverb: SETS VALUE TO/IS THE VALUE SET BY.

2. $\{tall\}_{Adj}$ SETS VALUE TO $\{plus\}_{Adv}$/$\{plus\}_{Adv}$ IS THE VALUE SET BY $\{tall\}_{Adj}$

This relation overcomes the information loss mentioned above: through the combination of the CHARACTERIZES WITH REGARD TO and the SETS VALUE TO relations we are able to obtain the cluster organization of adjectives, without the need for using fuzzy similarity relations or for defining *a priori* pairs of focal adjectives. Moreover, we can use the same strategy for encoding descriptive adjectives which do not assign scalar values. Adjectives like *dead* and *alive*, for instance, assign Boolean values, associated to the presence or absence of an attribute in the modified noun, i.e. a **yes** or **no** value of the relevant attribute. To encode this, we also use the SETS VALUE TO relation, linking such adjectives to $\{yes\}_{Adv}$ or $\{no\}_{Adv}$.

With regard to relational adjectives, things are considerably different, as these adjectives are not organized in opposite clusters. The meaning of relational adjectives is something like ‘of, relating/pertaining to, associated with’ some noun. In WordNet and EWN, relational adjectives are encoded as pertainyms of the nouns they are morphologically associated to. In EWN the PERTAINS relation is basically a morphological link (which is not always the case: e.g. *water* and *aquatic*), associated to a fuzzy semantic relation: it holds when no other relation clearly stands. In contrast to what is claimed in [3]:37, we argue that far from being meaningless 'themselves', relational adjectives involve sets of properties and introduce a relation between these sets of properties and the noun modified ([6], [4]). These adjectives establish an underspecified relation, which is specified in con-

[4] For a discussion on adjective scales and WordNet adjective clusters see [13].

text, between the modified noun and a domain that is exterior to it. In WN.PT, we use a very underspecified link to encode this relation: the relation IS BY DEFINITION RELATED TO. The salience of the semantic relation holding between relational adjectives and the lexicalization of the set of properties they are associated with, independently of any morphological link between them, motivates the creation of this new relation, which is exactly the opposite of what is stated about the PERTAINS relation in EWN, which focuses on the morphological link. Also, this broader relation allows for linking relational adjectives even when the set of properties involved is not lexicalized by a noun, but by a lexical item from another PoS, like in the case of *sedative*$_{ADJ}$ and *sedate*$_{V}$, for instance.

This way the main relations used for encoding descriptive and relational adjectives in WN.PT are: ANTONYMY, CHARACTERIZES WITH REGARD TO, and SETS VALUE TO, for the former; and IS BY DEFINITION RELATED TO, for the latter. These semantic relations allow us to encode the basic definitional characteristics of these adjectives in a linguistically motivated way, at the same time making it possible for membership to these classes to emerge from the network of relations encoded.

But adjectives are also relevant for the codification of salient properties of other lexical items. EWN uses the BE IN STATE relation to encode "links between nouns that refer to anything in a particular state expressed by an adjective" ([3]:37), recognizing the role adjectives can play in the characterization of nominal synsets. However, the definition and scope of application of this relation is too narrow: it cannot be used with relational adjectives, which are associated to sets of properties and not to a single state. Inspired by these observations and in order to broaden the domain of application of the link used in EWN, we introduce the new relation IS A CHARACTERISTIC OF/HAS AS A CHARACTERISTIC, in (3).

3. {carnivorous}$_{Adj}$ IS A CHARACTERISTIC OF {shark}$_{N}$ *reversed*
   {shark}$_{N}$ HAS AS A CHARACTERISTIC {carnivorous}$_{Adj}$

This relation allows us to express the most salient – and definitional – features of nouns in the network, contributing to richer and more clearly defined synsets. The possibility of ascribing, but also of negating this relation allows us to encode contrasting definitional features of certain nouns, in a similar way to the features encoded by some meronymy relations[5].

[5] One of the prototypical features of *shark*, in (3), is *carnivorous*. In contrast, one of the distinctive features of *whale shark*, hyponym of *shark*, is the fact that

Being able to express this is therefore very relevant, not only because it mirrors speakers' lexical knowledge, but also because it can provide crucial information to wordnet-based applications using inference systems.

Finally, in WN.PT adjectives are also used to encode definitional properties of verbs. Following [14] and further work on the representation of complex predicates in wordnets, verb telicity is also encoded in WN.PT.

4. $\{sadden\}_V$ HAS TELIC SUBEVENT $\{sad\}_{Adj}$/$\{sad\}_{Adj}$ IS TELIC SUBEVENT $\{sadden\}_V$ *reversed*
5. $[_T\ [_P\ act(x,y)\ and \sim Q(y)],\ [_e Q(y)]]$
   T: transition, P: process, e: event, Q: atomic event
6. a. He made Mary sad./b. *He made Mary.
7. a. *He saddened Mary sad./b. He saddened Mary.

The semantics of telic verbs involves a change of state of their theme argument, i.e. the subevent that closes the whole event is an atomic event, (a state) that affects its theme and is different from its initial state. By default, these verbs are associated to an LCS (Lexical Conceptual Structure) like the one in (5).

When syntactically realized, in contexts with LCS deficitary telic verbs ([14]), for instance, the telic subevent generally corresponds to an adjectival constituent (see 6a), whereas in the general case the telic state is incorporated in the verb, hence the ill-formation in (7a). In WN.PT the telicity of these verbs is captured through the relation HAS TELIC SUBEVENT/IS TELIC SUBEVENT (see (4)). This relation is different from the SUBEVENT relation in EWN as the latter only stands for lexical entailment involving temporal proper inclusion, therefore not accounting for the geometry of the event (see (5)). This is not the case of the TELIC SUBEVENT relation which regards the atomic subevent that is the ending point of the global event denoted by the verb, thus not properly included.

---

it is not. Moreover, this is the specific difference that distinguishes it from its sisters. This example makes apparent that this relation between nouns and adjectives expresses information just as crucial as the one encoded by some MERONYMY relations: *caffeine* IS A MERONYM OF *coffee*, and the negation of this MERONYMY relation is the specific difference of *decaf*, for instance.

2.2 *Prepositions in WordNet.PT*

Besides being syntactic markers, prepositions are also regarded as a kind of relation operator, relating concepts such as space, temporality or causality, and have been described according to their conceptual properties ([15], [16], [17], among others). Studies such as these, along with the identification of the need to account for arguments introduced by prepositions for a fine-grained codification of predicates in relational models of the lexicon, motivated the integration of prepositions in WN.PT ([5]).

As other PoS, prepositions can be related by SYNONYMY [6], HYPERONYMY and ANTONYMY relations, although the criteria for establishing whether these relations hold or not between two prepositions require slight adjustments of the test formulae used for pinpointing these relations, in order to consider the preposition plus the element with reference potential it combines with ((8), (9), (10)).

8. $Prep_1$ IS SYNONYM OF $Prep_2$ in a given Context iff: if $Prep_1$ then $Prep_2$ and if $Prep_2$ then $Prep_1$ (*over* IS SYNONYM OF *on top of*)
9. $Prep_1$ IS HYPERONYM OF $Prep_2$ iff: $Prep_2$ is $Prep_1$+ (space/time/direction…) but not the converse ({toward} IS HYPERONYM OF {downward} (*toward* + direction))
10. $Prep_1$ IS ANTONYM OF $Prep_2$ iff: i) $Prep_1$ and $Prep_2$ are hyponyms of $Prep_3$; ii) $Prep_1+XP_i$ is the opposite of $Prep_2+XP_i$ and $Prep_2+XP_i$ is the opposite of $Prep_1+XP_i$; therefore if $Prep_1+XP_i$ then not $Prep_2+XP_i$ and if $Prep_2+XP_i$ then not $Prep_1+XP_i$ ({above} IS ANTONYM OF {below})

Interestingly, the linguistic tests for HYPONYMY show that prepositions denoting source and goal locations, for instance, are not hyponyms of a preposition denoting location ([5]). In fact, there is a strong semantic relation between the concepts of location, and source and goal locations, but it is a causality relation rather than a specification relation: moving something to a goal location causes that something to be in that location (see (11)), just as moving something from a source location causes that some-

[6] Although it exists, SYNONYMY is not very productive for this PoS. This fact is probably not independent of prepositions being a closed-class, and seems to be conversely proportional to the highly polysemic behavior of prepositional expressions.

thing not to be in that location[7]. This way, prepositional nodes can also be related by CAUSE relations.

11. $Prep_1$ CAUSES $Prep_2$ iff: $Prep_1$+XP_i CAUSES/HAS AS CONSEQUENCE $Prep_2$+XP_i, but not the converse ({to} CAUSES/HAS AS CONSEQUENCE {at})

The integration of prepositions in wordnets, besides allowing to explicitly state subcategorization properties of predicates, contributes to compensate some shortcomings of mainstream wordnets, namely in terms of distinguishing word senses based on the relations encoded in the database. In section 2.3, we discuss these aspects in detail, in relation with a proposal for encoding selectional properties of predicates in wordnets.

2.3 *Encoding Selection Information*

Among the cross-PoS relations available in EWN, there is a set of relations concerning the role (or function) of entities in events. As stated in [3]:29, ROLE relations are based on thematic role assignment, and are correlated with the argument structure of verbs. However, the nodes related by ROLE relations often are not coincident with the selection restrictions of verbs. In addition, in many cases, ROLE relations are only definitional to the meaning of the participant. For instance, a *passenger*/*customer* is defined as one that *travels/buys*, but the event denoted by *travel*/*buy* is not defined as an event having a *passenger/customer* as an agent.

Following research on verbal predicates ([5]), we define three new relations to account for selection information, based on the argument structure as defined in the Generative Lexicon (GL) ([18]). Argument structure in GL allows for specifying the number and semantic type of arguments of a given predicate, also including information on how these arguments behave syntactically in general, namely with regard to specific restrictions on their overt realization in context, distinguishing between true, shadow and default arguments ([18]:63 and ff.).

[7] PPs introduced by the preposition *at*, indicator of location, correspond to the resulting state of the movement *from* or *to* to a given location. Prepositional phrases headed by this item can replace state denoting items such as adjectives, providing evidence for this claim (see [5]:155): *John is* ***tired***./*John is* ***at the door***.

Briefly, the relation SELECTS/IS SELECTED BY refers to true arguments, i.e arguments that have to be syntactically realized (or whose omission has to be licensed by syntactic or pragmatic contexts); the INCORPORATES/IS INCORPORATED relation refers to shadow arguments, strictly incorporated in the lexical predicate, which means they cannot be overt arguments unless they are further specified; and the HAS AS DEFAULT ARGUMENT/IS DEFAULT ARGUMENT OF refers to participants in the event structure of the predicate that are mostly null, since the semantics of the predicate allows for a default interpretation (for further discussion on these relations, see [5]). Also, taking advantage of the inheritance mechanism in the WordNet model, the relation SELECTS accounts for the overt realization of the target node of this relation or any of its direct or indirect hyponyms, see (12).

12. $\{\text{die}\}_V$ SELECTS $\{\text{living being}\}_N$: All living beings / birds / men / insects / ... die.

The implementation of these relations in WN.PT takes advantage of the possibility of relating either variants or synsets, and from the conjunction operator, available in the EWN framework. The first allows for stating different selection restrictions for the members of a synset, in (13) below. If nothing is stated, the relation applies to all the elements. Otherwise, the variant-to-variant restriction has to be activated, and the two elements related explicitly identified. As to the conjunction operator, it allows for simultaneously linking the elements of complex arguments, as it is the case, for instance, of arguments introduced by a preposition, illustrated in (14):

13. $\{\text{voltar, regressar}\}_V$ [≅ return, come back]
    SELECTS $\{\text{para}\}_P$ (**variant to variant**: *voltar - para*);
    SELECTS $\{\text{a}\}_P$
14. $\{\text{engarrafar}\}_V$ [≅ bottle; put in a bottle]
    INCORPORATES $\{\text{em}\}_P$ [≅ in] (**conjunctive 1**)
    INCORPORATES $\{\text{garrafa}\}_N$ [≅ bottle] (**conjunctive 2**)

An inheritance mechanism drastically reduces the work involved in specifying this information, since selection information relations are inherited through hyponymic chains, as mentioned above. However, selectional information is not always completely inherited by hyponyms, as made apparent by the case of incorporated arguments, further motivating a mechanism of lexical inheritance by default: hyponyms inherit all the information that characterizes their hyperonyms if nothing is stated otherwise.

ROLE and selection information relations are not always coincident, even when considering definitional properties of predicates only. WN.PT data shows that ROLE and selection information relations are typically coincident in the case of agents ({dress}$_V$ INVOLVED_AGENT/SELECTS {person}$_N$), the same not being necessarily true when other participants are at stake. In (15), the instrument used in an event like *selar* (seal) is identified through a ROLE relation, but this relation does not allow us to know that, in the specific case of this verb, this is an incorporated argument, and, as such, only syntactically realized under strict constraints.

15. {selar}$_V$ [seal, ≅ close with a seal]
    INVOLVED_INSTRUMENT {selo}$_N$ [≅ seal]
    INCORPORATES {com}$_P$ [≅ with] (conj. 1); INCORPORATES {selo}$_N$ [≅ seal] (conj.2)

According to the literature ([9]), the specification of the manner in which events occur has a special significance in the determination of verbal meaning. This specification, when a lexicalization of the manner is available, is encoded through the IN_MANNER relation ([3]:36), linking verb and adverb synsets, such as {run}$_V$ IN_MANNER {fast}$_{Adj}$. In a similar way, when no lexicalization of manner is available, but this information is incorporated in the verbal predicate, we claim that the INCORPORATES relation can be used, as shown in (16).

16. {puxar}$_V$ [≅ move with traction, pull]
    INCORPORATES {com}$_P$[≅ with] (conj. 1); INCORPORATES {tracção}$_N$ [≅ traction] (conj. 2)

The introduction of selection information relations allows for distinguishing and representing different levels of information in the WordNet model, increasing the amount of information that can be expressed in it: ROLE and IN_MANNER relations (existing in the EWN framework) – conceptual properties; SELECTS, INCORPORATES and HAS AS DEFAULT ARGUMENT relations – selectional properties and syntactic restrictions. Selection information relations coherently complement the existing relations, resulting in a more accurate description of lexical items and linking synsets which otherwise would not be associated. Our goal is not to provide complete syntactic frames for each synset, but to make available richer descriptions of lexical-conceptual units, following the assumption that selection information reflects semantic and syntactic relations ([19]). Consider, for instance, the verb *pôr* (≅ put; to move to a location), that selects an argument introduced by the preposition *em* (≅ indicator of location). On the one hand, the specific information that distinguishes this verb from its hyperonym is directly related to the determination of a final location. On the other hand, there is no node in the lexicon suitable to be linked through INVOLVED_LOCATION to the verb *pôr*, since almost any lexicalization of a surface, object, area, body part, etc., can occur with *pôr*. The integration of prepositions in WN.PT, together with the selection information relations described, result in a richer and more accurate characterization of the semantic properties of lexical items, explicitly modeling semantic content and co-occurrence information reflecting semantic properties, and enabling the establishment of word senses based on the information expressed in the network ([19]):

17. {pôr}$_V$ [≅ put] HAS_HYPERONYM {mover}$_V$ [≅ move]
    INCORPORATES {para}$_P$ [≅ indicator of goal location]
    SELECTS {em}$_P$ [≅ indicator of location]

Although the preposition *em* (≅ in) does not impose strong selection constraints on the NP it combines with (e.g. *in the table / bedroom / fridge / field / shoe / air…*), the incorporated meaning component of *pôr* referring a goal location is now accounted for through the INCORPORATES relation, and the concept of location is indicated by the prepositional nodes selected.

These more precise and richer descriptions of synsets allow for contributing to the distinction of word senses based on the relations encoded in the network. In fact, selection information alone can be used to straightforwardly distinguish between word senses, as illustrated in (18). This is particularly relevant since, as argued in [20], information on the co-occurrence of words is easily available in texts – and thus easily accessible for NLP tasks based on corpora analysis and statistic oriented models –, whereas information on the co-occurrence of meanings is harder to extract from raw data and requires complex strategies involving text disambiguation.

18. {tratar}$_V$ [≅ treat, heal]
    SELECTS {animal}$_N$ [≅ animal]
    {tratar}$_V$ [≅ take care]
    SELECTS {de}$_P$ [≅ of] (conj.1); SELECTS {animal}$_N$ [≅ animal] (conj. 2)

3 WordNet.PT Data: Informational Richness and Density

The density of wordnets is specifically significant considering that in this model word senses are represented in terms of relationships between synsets. Also, WordNet has been used to solve primary barriers in the development of reliable information retrieval, machine translation, summarization and language generation systems, or word-sense disambiguation applications, for which rich language resources are crucial.

Particularly in the case of word sense disambiguation, the density of WordNet has been considered limited ([21]). Thus, augmenting the density of relational language resources is decisive from the point of view of their usability, whether we consider inference-based applications, where the richer the connectivity in the database, the more inference is possible ([22]), or applications that draw on measurements of semantic relatedness between concepts, since higher relational density provides shorter average paths between lexical objects ([23]). For these reasons, several strategies have been put forth in order to augment the density of wordnets, such as those depicted in [19], [20], [21] or [23], to name a few, and further developments resulting in the increase of network density continue to be welcomed. In this section, we show how the new relations proposed under the scope of our work accomplish just that.

Table 1 compares the density (number of relations per synset) of WN.PT, after the implementation of the new lexical-conceptual relations described, with the density of WordNet 1.5[8], regarding adjectives and a subset of verbs.

Comparing WN.PT with WordNet 3.0 instead of WordNet 1.5 (see footnote 8), particularly considering PoS differentiation, could provide different numbers. Nonetheless, it is possible to show how density increases as a result of using the new selection information relations with regard to the verbs tested. The same occurs when comparing EWN relations and

[8] Developed in the general framework of EWN, WN.PT was originally implemented with Polaris, which determined its mapping with WordNet 1.5 data, the mapping of Inter-Lingual links to WordNet 3.0 being still ongoing. For this reason, and given the fact that WordNet 3.0 statistics do not cover the number of relations in total or by PoS, the data considered here for purposes of comparison are those of WordNet 1.5. Based on the statistics available (http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html) and data offered by [24]:374, the density of WordNet 3.0 considering the total number of synsets and relations is 2.0.

Table 1. Network density for adjectives and a sample of verbs (change of location verbs) in WordNet 1.5, WN.PT only with EWN relations, and with its new specific relations

	WordNet 1.5	EWN relations	WN.PT relations
Adjectives	1.48	–	**4.48**
Change of location verbs	2.13	4.12	**5.68**

new relations implemented in WN.PT in general. The increase in density of WN.PT with regard to WordNet 1.5 is quite substantial: about 200% for adjectives and more than 165% for the verbs tested. This increase is not as high when comparing WN.PT using EWN relations only with WN.PT using the full set of relations proposed. However, it is still quite significant (37%).

Adding to the linguistic motivation, these results further sustain the use of these new relations in wordnets. Besides the importance of having a denser network from the point of view of wordnet-based applications, increasing the density of wordnets is a crucial aspect for relational models of the lexicon themselves since the meaning of each unit is determined by the set of relations it holds with other units: a denser network of relations results in richer and more appropriately defined synsets.

4 Final Remarks

The implementation of new relations and the integration of new PoS in WN.PT decisively contribute to enhancing its density, consistency and coverage. The new relations allow for more accurate and motivated descriptions but also for the integration of new PoS, enhancing the usability of the database in different types of computational applications. This has been tested in several applications, both in terms of the contribution to the treatment of different linguistic phenomena (such as co-occurrence restrictions of co-hyponyms and contrasts in Aktionsart values within troponymic chains ([5]), word-sense disambiguation ([25], [26]), or usability in Language Engineering applications ([27])).

However, several issues require further attention. First, some promising results emerge from the work already developed. With regard to adverbs, studying to which extent the comprehensive treatment of event modifying adjectives can contribute to the treatment of this category and how the set of properties identified is mirrored in wordnets are open questions that can contribute to a deeper treatment of this PoS. Further research and testing on the selection properties of predicates is also due, specifically in the case of underspecified arguments that correspond to high nodes in the hierarchy.

Regarding this, using the information available in WN.PT to establish selection features might provide a solution. Consider, for instance, the following Portuguese pair of verbs: *enjaular* (≅ cage; put inside a large and resistant cage, typically made of metal, animals of considerable size), as in *the hunter caged the lion*; and *engaiolar* (≅ cage; put inside a cage, small animals, typically birds or small mammals), as in *the child caged the canary*. Setting *animal* as argument of these two verbs overgenerates, since many hyponyms of *animal* cannot be arguments of either one or the other of the two verbs. The solution might be to consider features expressed by other available relations, such as HAS AS CHARACTERISTIC $\{\text{grande}\}_{\text{Adj}}$ (big) or $\{\text{pequeno}\}_{\text{Adj}}$ (small), for the arguments of *enjaular* and *engaiolar*, respectively.

References

1. Marrafa, P.: The Portuguese WordNet: General Architecture and Semantic Internal Relations, *DELTA* (2002).
2. Marrafa, P.: WordNet do Português - Uma base de dados de conhecimento linguístico. Lisbon: Instituto Camões (2001).
3. Vossen, P.: EuroWordNet General Document. EuroWordNet Project LE2-4003 & LE4-8328 report, University of Amsterdam (2002)
4. Mendes, S.: Syntax and Semantics of Adjectives in Portuguese: analysis and modeling. PhD dissertation. University of Lisbon (2009)
5. Amaro, R.: Computation of Verbal Predicates in Portuguese: relational network, lexical-conceptual structure and context. PhD dissertation. University of Lisbon (2009)
6. Mendes, S.: Adjectives in WordNet.PT. 3rd Global WordNet Association Conference, pp. 225–230. Jeju Island, Korea (2006)
7. Miller, K. J.: Modifiers in WordNet. In: Fellbaum, C. (ed.) WordNet: an electronic lexical database, pp. 47–68. Cambridge, MA: The MIT Press (1998)
8. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.J.: Introduction to WordNet: an On-line Lexical Database. International Journal of Lexicography, vol. 3, number 4 (1990).

9. Fellbaum, C.: A Semantic Network of English Verbs. In: Fellbaum, C. (ed.) WordNet. An Electronic Lexical Database, pp. 69–104. Cambridge, MA: The MIT Press (1998)
10. Kamp, H.: Two theories about adjectives. In: Keenan, E. (ed.) Formal Semantics of Natural Language, pp. 123-155. Cambridge: Cambridge University Press (1975)
11. Chierchia, G., McConnel-Ginet, S.: Meaning and Grammar: an Introduction to Semantics. Cambridge, MA: The MIT Press (1990)
12. Fellbaum, C., Gross, D., Miller, K. J.: Ajectives in WordNet. In Five Papers on WordNet. Princeton, USA (1993).
13. Sheinman, V., Tokunaga,T.: AdjScale: Differentiating between similar adjectives for language learners. 1st International Conference on Computer Supported Education, pp. 229–235 (2009)
14. Marrafa, P.: Modelling constituency and predication in Portuguese. Revista PaLavra, vol. 12 (special issue: Linguística Computacional), pp. 106–118 (2004)
15. Jensen, P., Nilsson J. F.: Ontology-Based Semantics for Prepositions. ACL-SIGSEM workshop. Institut de Recherche en Informatique de Toulouse (2003)
16. Saint-Dizier, P.: PrepNet: a framework for describing prepositions: preliminary investigation results. 6th International Workshop on Computational Semantics ITK, pp. 25–34. Tilburg (2005)
17. Mcshane, M, Beale S., Nirenburg S.: Disambiguating Homographous Prepositions and Verbal Particles In An Implemented Ontological Semantic Analyzer. Working Paper 01-05. ILIT, University of Maryland Baltimore County (2005)
18. Pustejovsky, J.: The Generative Lexicon. Cambridge, MA: The MIT Press (1995)
19. Agirre, E., Martinez, D.: Integrating selectional preferences in WordNet. 1st International WordNet Conference, pp. 1–9. Mysore (2002)
20. Bentivogli, L., Pianta E.: Extending WordNet with Syntagmatic Information. 2nd International WordNet Conference, pp. 47–53. Brno, Czech Republic (2004)
21. Boyd-Graber, J., Fellbaum, C., Osherson D., Schapire R.: Adding dense, weighted connections to WordNet. 3rd Global WordNet Meeting, pp. 29–35. Jeju Island, Korea (2006)
22. Harabagiu, S., Moldovan D.: Knowledge Processing on an Extended WordNet. In: Fellbaum, C. (ed.) WordNet. An Electronic Lexical Database, pp. 353–378. Cambridge, MA: The MIT Press (1998)
23. Lemnitzer, L., Wunsch H., Gupta P.: Enriching GermaNet with Verb-noun Relations - a Case Study of Lexical Acquisition. LREC 2008, pp.156–160. Marrakech, Morocco (2008)

24. Agirre, E., Montse, C., German R., Soroa, A.: Exploring Knowledge Bases for Similarity. LREC 2010, pp. 373–377. Valletta, Malta (2010)
25. Marrafa, P., Mendes, S.: Using WordNet.PT for translation: disambiguation and lexical selection decisions. International Journal of Translation, vol. 19. Bahri Publications (2007)
26. Marrafa, P., Amaro, R., Freire N., Mendes S.: Controlled Portuguese: coping with ambiguity. In: Kuhn, T., Fuchs N. E. (eds.) CNL 2012, LNCS 7427, pp. 152–166. Springer-Verlag Berlin Heidelberg (2012)
27. Marrafa, P., Ribeiro, C., Santos R., Correia, J.: Gathering Information from a Relational Lexical-Conceptual Database: A Natural Language Question-Answering System. 8th World Multi-Conference on Systemics, Cybernetics and Informatics. Orlando (2004)

RAQUEL AMARO
CENTRO DE LINGUÍSTICA,
UNIVERSIDADE DE LISBOA,
AVENIDA PROFESSOR GAMA PINTO, 2, 1649-003 LISBOA, PORTUGAL
E-MAIL: < RAMARO@CLUL.UL.PT >

SARA MENDES
CENTRO DE LINGUÍSTICA,
UNIVERSIDADE DE LISBOA,
AVENIDA PROFESSOR GAMA PINTO, 2, 1649-003 LISBOA, PORTUGAL
AND
UNIVERSITAT POMPEU FABRA
ROC BORONAT 138, BARCELONA, SPAIN
E-MAIL: < SARA.MENDES @CLUL.UL.PT >

PALMIRA MARRAFA
CENTRO DE LINGUÍSTICA,
UNIVERSIDADE DE LISBOA,
AVENIDA PROFESSOR GAMA PINTO, 2, 1649-003 LISBOA, PORTUGAL
E-MAIL: < PALMIRA.MARRAFA@NETCABO.PT >