On the Automatic Enrichment of a Portuguese Wordnet with Dictionary Definitions

Hugo Gonçalo Oliveira and Paulo Gomes

CISUC, University of Coimbra, Portugal
{hroliv,pgomes}@dei.uc.pt

Abstract. Besides synsets and semantic relations, synset glosses are an important feature of wordnets. However, due to the required effort, their creation is sometimes left undone. This happens in Onto.PT, a Portuguese wordnet created automatically, which does not have glosses. In our work, we exploited Portuguese dictionaries to automatically assign definitions to the synsets of Onto.PT. For this purpose, definitions are selected according to their overlap with the context of the synsets. Using three Portuguese dictionaries, more than one third of the Onto.PT synsets have at least one definition, with assignment accuracy close to 80%, which we believe to be interesting results.

Keywords: wordnets, synsets, definitions, dictionaries, senses, relations

1 Introduction

Wordnets are lexical knowledge bases (LKBs) structured on the so called synsets and on semantic relations between them. Each synset groups synonymous word senses, in what can be seen as the possible lexicalisations of a natural language concept. Moreover, some wordnets contain synset glosses, which are textual definitions that describe the meaning transmitted by each synset. Glosses may add specific information that does not fit in the semantic relations and enable to use a wordnet as a dictionary, besides their common utilisation as a thesaurus with relations. Glosses have been exploited to extend WordNet automatically [1], and they are sometimes used as an additional source of contextual information in algorithms for word sense disambiguation (WSD) [2].

Since Princeton WordNet [3], several wordnets were created for non-English languages, including those involved in the EuroWordNet [4] or MultiWordNet [5] projects, which shows the wide acceptance of this model. But wordnets are typically large, and their creation requires much human effort. To minimise this, there have been automatic attempts for the enrichment of wordnets [1], or for the creation of new wordnets, either by the translation of Princeton WordNet [6], or from scratch [7]. There are also semi-automatic approaches where synsets are created manually, for the target language, and then linked to their English correspondent, from which relations are inherited [8].

In a translation approach, obtaining a gloss for a synset can be just a matter of translating the gloss in the corresponding English synset. But when it comes to the automatic creation of wordnets from scratch, acquiring definitions for the synsets is not always trivial. Even in the manual creation, adding definitions is another time-consuming step. This is why the creation of wordnets can be optimised if definitions are selected or suggested automatically for each synset.

Onto.PT [7] is a freely available wordnet-like resource for Portuguese, created automatically by exploiting Portuguese dictionaries and thesauri. But Onto.PT is focused on the establishment of synsets and semantic relations and does not contain synset glosses. So, the meaning transmitted by a synset needs to be inferred both from the words it groups and from words in related synsets. This can be problematic, especially for synsets with only one word and few relations. Thus, we were inspired by the mapping between GermaNet senses to Wiktionary definitions [9], and exploited Portuguese public dictionaries in the automatic selection of definitions for the Onto.PT synsets. This resulted in more than a third of the synsets of Onto.PT with a definition, with an estimated accuracy of almost 80% on the definitions of synsets with more than one word.

This paper starts by referring some related work, covering the creation of wordnets, with some focus on the glosses, and also on the integration of lexical resources. Then, we describe the generic procedure for assigning definitions to the synsets. After that, we address the specificities of doing this for Portuguese, while introducing the used resources and some implementation details. Before concluding, we present the obtained results and their evaluation.

2 Related Work

The wide acceptance of Princeton WordNet [3] is evidenced by its utilisation in various natural language processing tasks, such as question answering (QA) [10] or WSD [2], and by the adoption of its model by LKBs in other languages [4, 5]. A wordnet is structured in synsets and semantic relations, which often include hypernymy, and possibly part-of, antonymy and others. Moreover, in most wordnets, synsets have a gloss, which is a natural language description of their meaning and works as a dictionary definition. In fact, wordnets represent the same knowledge as dictionaries, but are structured for computer applications.

The manual creation of a wordnet requires too much human effort and, when time is an issue, the creation of glosses tends not to be a priority. In non-English wordnets aligned to Princeton WordNet, including those generated by translation [6], the original English glosses are often used. As for wordnets created from scratch, whether manually, as GermaNet [11], or automatically, as Onto.PT [7], the glosses are sometimes left undone. This leaves the interpretation of the synset to be made only by the words it groups and by the relations where it is involved.

This is different from MindNet [12], a LKB created automatically. MindNet is not structured as a wordnet and is closer to dictionaries, the primary resources exploited in its extraction. Therefore, the sense entries of MindNet have an associated definition. On the other hand, PAPEL [13], a LKB for Por-

tuguese, also extracted from a dictionary, does not have glosses. This is due to copyrights (proprietary dictionary) and structural issues (just words, no senses).

But there are other Portuguese LKBs that follow the wordnet model. WordNet.PT [14] has complete glosses but is not available to use by a third party. From the available wordnets, TeP [15] and OpenWN-PT [16] contain just synsets, but the latter is aligned to Princeton WordNet, from which it can inherit both relations and glosses (in English). Onto.PT [7] covers several types of relations, is large, and created automatically, but does not have synset glosses. That is why it is the target of our work. Moreover, if the addition of glosses is automated, the creation pipeline can remain fully automatic.

A great inspiration for this work is the mapping of GermaNet to Wiktionary [9] that enabled the automatic suggestion of definitions to the GermanNet synsets. A related task is the automatic extraction of definitions from text, addressed in the creation of domain glossaries from the Web [17], definitional QA [18], and others. For WSD, however, it might be enough to extract, for each word sense, just a set of highly correlated words, regarding that they are not related with other senses of the same word [19].

Also, to some extent, our work integrates dictionaries and Onto.PT. Therefore, it is also related to those that integrated or aligned lexical resources. Besides the multilingual wordnets, there are resources such as UBY [20], a unified resource that integrates Princeton WordNet, GermaNet, Wiktionary, Wikipedia and others, or BabelNet [21], a multilingual knowledge base that integrates WordNet with the Wikipedias in several languages. In BabelNet, in addition to the WordNet glosses, the first sentences of the Wikipedia articles are used as a definition for the Wikipedia entries.

3 Generic procedure

The proposed procedure assigns one or more sense definitions, in a dictionary, to a synset, in a wordnet. It is thus suitable to wordnets whose synsets do not have a gloss. For this, we were inspired by WSD-VAL [9], an algorithm used to map the lexical units of GermanNet [11] to the German Wiktionary. But we made some adaptations regarding the available resources for Portuguese, which included using more than one dictionary and a wordnet created automatically.

The result is close to WSD, as words are matched with corresponding senses. In fact, WSD-VAL is a variation of the classic Lesk [22] algorithm. However, while in typical WSD words in a context are matched to word senses in an inventory, we are matching word senses of two potential inventories, because both wordnets and dictionaries describe word senses.

A wordnet can be seen as a graph $G = \{S, R\}$, with $|S|$ nodes (the synsets) and $|R|$ edges (the relations), $R \subset S^2$. A synset has a part-of-speech (POS) and groups word senses, denoted by words $s_j \in S$. A dictionary, $D = \{K, T\}$, has $|K|$ unique keys (definienda) and $|T|$ textual definitons, that describe a possible

meaning of the key. Each key is a word and may be used to retrieve its possible definitions, $defs(k_i \in K) = T'$, $T' \subset T$.

In order to select matches between the synsets of a wordnet and the word senses in a dictionary, we exploit their contexts. In a dictionary, the context of a word sense consists of the words in its definition. As for a synset, the context is typically made by the words in its gloss or in example sentences. However, given that this procedure applies for wordnets without glosses, we cannot rely on the previous. Alternatively, the context of our synsets is defined by all the words the synsets group and words in related synsets.

The generic algorithm for our task encompasses the following steps. For each synset S_i in the wordnet:

1. Create a bag with its words $s_j \in S$ and in related synsets, $s_r \in S_r : \forall(S_r \in S) \rightarrow \{S_i, S_r\} \in R$. This will represent the context of S_i, Σ_{Si}.
2. Search in the dictionary for all the definitions of all the words in S_i with its POS, and create a set of candidate definitions for this synset, $foreach(s_j \in S_i), C_{Si} = C_{Si} \cup defs(s_j)$
3. For each candidate definition $T_{Sik} \in C_{Si}$, perform POS-tagging, in order to obtain the POS and the lemma of each of its words.
4. Use the lemmas of the content words in T_{Sik} as a bag-of-words representing the definiendum context, $\Delta_{T_{Sik}}$.
5. Similarly to the Lesk [22] algorithm, give a score ω to T_{Sik} according to the intersection between the contexts, $\Sigma_{Si} \cap \Delta_{T_{Sik}}$
6. If there are definitions in C_{Si} with $\omega_{S_i,T_{Sik}} > \theta$, a predefined acceptance threshold, select those with higher score. Otherwise, select none.

Ideally, this algorithm would lead to a definition per synset. However, besides having different coverages and a different organisation, different lexical resources tend to have different sense granularities that do not always match [23]. Therefore, after applying this algorithm, a synset can have one definition, no definitions at all, or several definitions.

It might be fruitful to compile definitions from more than one dictionary in a virtual dictionary used by the algorithm, thus increasing the probability of having synsets with more than one definition. But, while some definitions can be simple paraphrases of each other, sometimes, they can also be complementary, which is welcome. In fact, we have made experiments with more dictionaries, in an attempt to cover more synsets.

4 Selecting definitions for Portuguese synsets

The generic procedure, described earlier, was applied for Portuguese. We exploited freely available dictionaries to select definitions automatically for the synsets of Onto.PT, a wordnet whose synsets do not contain glosses. After presenting Onto.PT and the dictionaries, we describe some additional processing and implementation details, including the scoring of the candidate definitions.

4.1 Onto.PT

Onto.PT [7] is a LKB for Portuguese, structured similarly to Princeton WordNet [3]. As typical wordnets, Onto.PT covers the whole language and not a specific domain, and it is structured on synsets and semantic relations. However, in opposition to most wordnets, Onto.PT is not handcrafted, but created automatically, by exploiting Portuguese dictionaries and thesauri. It intends to provide a free large LKB that merges lexical knowledge from several resources, while minimising the effort needed to create a wordnet manually. The creation of Onto.PT is briefly described in the following steps:

1. Exploit regularities in dictionary definitions to extract instances of semantic relations, connecting words.
2. If possible, attach each synonymy relation to a synset in an existing Portuguese handcrafted thesaurus (currently, TeP [15]). Then, discover clusters in the unattached synonymy relations and add them as new synsets.
3. Use similarity-based heuristics to integrate the rest of the relations, by assigning each relation argument to the most suitable synset.

Using this procedure, it is possible to obtain a larger resource with more relation types, in a trade-off for lower reliability. In Onto.PT, relations go from well-known hypernymy and part-of, to relations established between words of different POS, including, for instance, purpose-of, manner-of, or has-quality. But, as different resources have different sense granularities, synset glosses are not part of Onto.PT. This makes it a suitable target for our work.

Onto.PT was released in 2012 and, as the result of an automatic approach, is always under development[1]. We have used version 0.4.1 which contains ≈163,000 unique words, organised in ≈110,000 synsets, connected by ≈176,000 relation instances. In version 0.3.5, the attachment of synonymy to synsets was between 76-81% accurate and clustering between 83-89% (verbs) and 95% (adjectives). The accuracy of hypernymy relations was only 65% and the combined accuracy of the other relations is between 78%-82%. These numbers are expected to be higher in Onto.PT 0.4.1.

4.2 Dictionaries

Although several Portuguese dictionaries are available online, most of them are proprietary and thus not available for download. The dictionaries used in this work are exceptions to this situation.

Dicionário Aberto (DA) [24] is the electronic version of a Portuguese dictionary from 1913. Given its age, the orthography of the DA is currently in the process of being modernised. However, in the project's website[2], where the resource can be downloaded in several formats, available statistics show that the majority of words is still to be revised towards modernisation.

[1] See http://ontopt.dei.uc.pt for download and additional information.
[2] See http://www.dicionario-aberto.net/

Wiktionary is a collaborative initiative, maintained by the Wikimedia Foundation, for providing multilingual electronic dictionaries of free content. As Wiktionaries are built manually by non-professional volunteers on the Web, the provided information can be incomplete and inconsistent. On the other hand, they are constantly growing and freely available as database dumps[3]. Wiktionary.PT[4] is the Portuguese Wiktionary. As compared to other languages, it is small. To have an idea, on March 2013, the English Wiktionary had more than 3M entries, while Wiktionary.PT had just about 187,000. Still, as Wiktionaries are multilingual, not all of those entries correspond necessarily to words in those languages. We have used the 28th February 2013 dump of this resource.

We should add that both DA and Wiktionary.PT are among the resources exploited in the creation of Onto.PT. Therefore, most of their words and senses are expected to be covered by this wordnet.

We have also made some experiments with Wikipedia.PT[5], the most popular initiative by the Wikimedia Foundation, that provides encyclopedias in several languages. Wikipedia is not a dictionary nor structured on lexical knowledge. It covers encyclopedic knowledge, so most of its articles are about proper nouns, such as individuals, locations, organisations, or events. Also, although the Wikipedia articles are typically larger than dictionary definitions and provide additional information, their first sentence often works as a definition. In fact, using these sentences as glosses is a common practice (see e.g. [21]). We have used the 3rd March 2013 dump of this resource.

4.3 Dictionary processing

Before using the dictionaries, they were all converted to a common textual format, with a definition per line, after the definiendum and its POS. While converting the DA, we performed some automatic replacements towards modernisation of disused sequences. Changes were only accepted if the new word existed in another lexical resource (e.g. PAPEL [13]). As for Wiktionary.PT, we used an existing parser [25] to make some cleaning, including the removal of non-Portuguese words, and to convert the dump to the common format. For Wikipedia.PT, we only kept the abstracts matching the most typical definition pattern for nouns: `X é <det> D` (*X is <determiner> D*). These were converted to a dictionary entry, using `X` as the definiendum, and `D` as the definition.

Moreover, we knew beforehand that the dictionaries contained several definitions that would not add additional information to the meaning of synsets. So, the following kinds of definitions were discarded:

- Those with only one word or an enumeration of words, which tends to include words already in a synset, as in:
  - `negociar verbo comerciar` (*negotiate, v: trade*)
  - `caos nome confusão, desordem` (*chaos, n: confusion, disorder*)

[3] See http://dumps.wikimedia.org/
[4] See http://pt.wiktionary.org/
[5] See http://pt.wikipedia.org

– Those starting with `o mesmo que` (*same as*), always followed by a synonym of the definiendum, as in:
  - `aberração nome o mesmo que anomalia` (*aberration, n: same as anomaly*)
– Those belonging to closed category words (all but nouns, verbs in the infinitive, adjectives and adverbs), because Onto.PT only covers open categories.

Table 1 shows the number of definienda and definitions of each resource, including unique definitions in the common format (Total), and those actually used, after discarding uninformative definitions (Used). Wiktionary.PT is the smallest resource and Wikipedia.PT is the largest but, as we will show in section 5, most Wikipedia definitions are not suitable for our work.

Resource(s)	**Definienda**	**Definitions**	
		Total	Used
DA	130,501	229,296	174,235
Wiktionary.PT	36,371	81,759	57,750
Wikipedia.PT	306,857	347,171	345,947
DA+Wiktionary.PT	145,965	311,055	228,804
All 3	440,843	658,226	574,576

Table 1. Data of the dictionaries used.

4.4 Implementation details

The procedure in section 3 was applied to Portuguese by using Onto.PT as a wordnet and the dictionaries as a large virtual dictionary. In a first experiment, we merged just DA and Wiktionary.PT, and later we added Wikipedia.PT.

To obtain the content words in each definition, we POS-tagged them and applied some lemmatisation rules. For this purpose, we used the tagger provided by OpenNLP[6], using the available models trained for Portuguese. Only the lemmas of the nouns, verbs, adjectives and adverbs in the definition were used in the bag-of-words of the definienda context, Δ. As referred in section 3, the bag-of-words with the synset context, Σ, contained all the words in the synset and in related synsets. For each synset S_i, definitions T_j were scored as follows:

$$\omega_{S_i,T_{Sik}} = \#(\Sigma_{Si} \cap \Delta_{T_{Sik}}) + \#(H_{Si} \cap \Delta_{T_{Sik}})$$

There, H_i is a set with the words in the synsets directly connected to S_i by a hypernymy relation, $H_{Si} \subset \Sigma_{Si}$. It means that the hypernyms in the definition count twice as much as other related words. This option relies on the fact that noun definitions typically start with a *genus*, which identifies the superordinate concept of the definiendum, meaning that the definiendum is a "type of" the *genus* and there is a hyponymy relation between the former and the latter. The following definition, where 'plant' is the *genus*, illustrates this fact:

[6] See http://incubator.apache.org/opennlp/

– `amendoim nome planta herbácea de fruto comestível, cuja semente é rica proteínas e óleo`
(peanut noun herbaceous plant of edible fruit, whose seed is rich protein and oil)

Back to the implementation details, the acceptance threshold θ was set to 1. Therefore, only definitions with at least one hypernym or two other related words were considered for each synset S_i as they lead to $\omega_{S_i,T_{Sik}} > 1$.

5 Results and Evaluation

This section presents the quantitative results of this work by the number of synsets with an assigned definition. Then, it provides some examples followed by qualitative results, obtained after manual evaluation.

5.1 Covered synsets

The procedure for assigning definitions to the synsets of Onto.PT was run several times, using different dictionaries. First, DA, then Wiktionary.PT, and then both of the previous. After taking some conclusions, we ran it again, this time using Wikipedia.PT, and then all the three resources. Table 2 summarises the results.

Resource(s)	Synsets w/ definition	Avg(definitions/synset)
DA	32,040 (≈29.1%)	1.24
Wiktionary.PT	13,834 (≈12.6%)	1.40
Wikipedia.PT	6,377 (≈5.8%)	1.19
DA+Wiktionary.PT	37,325 (≈33.9%)	1.34
All 3	37,981 (≈34.5%)	1.36

Table 2. Results of the assignment of definitions to Portuguese synsets.

If we had to select only dictionary for this task, it would be DA. It is larger, has higher coverage, and enables to assign definitions to more than twice the number of synsets, as compared to Wiktionary.PT. However, given that the dictionaries used are free, it is worth using more than one. Using only DA, about 29% of the 110,000 Onto.PT synsets have at least one definition. This number grows to more than 34% if DA and Wiktionary.PT are used. The previous numbers are not surprising, because both DA and Wiktionary.PT are two of the resources exploited in the creation of Onto.PT.

On the other hand, in spite of being the largest resource, Wikipedia.PT provides definitions to much less synsets. In fact, when using the three resources instead of just the two dictionaries, only 656 additional synsets have a definition. This happens not only because Wikipedia.PT is not integrated in Onto.PT, but also because it is not limited to lexical knowledge. Most Wikipedia entries are actually named entities, which are not the focus of Onto.PT.

Table 2 also shows that, on average, each synset has between 1.19 definitions (Wikipedia.PT) and 1.40 (Wiktionary.PT). When there is a tie on ranking the definitions, a synset gets more than one definition. Sometimes, they are mere paraphrases but they might also complement each other.

5.2 Examples

Our results are illustrated in table 3 by three Onto.PT synsets, their candidate definitions with the original definienda and ranking, as well as relevant relations for ranking. Definitions that ended up being select are in bold.

Synset	{*borralheira, borralheiro, chanfana*}
Relevant relations	...hyponym-of {*lugar, local, sítio, loco*}
Definitions	*borralheira:* ***lugar, onde se junta a borralha da cozinha ou do forno*** *(2)* *chanfana: comida mal feita* (0) *chanfana: carne magra de carneiro, pelanga* (0)
Synset	{*alugar, arrendar, locar, alquilar*}
Relevant relations	...causes {*aluguer, arrendamento, renda, locação, alugamento*}
Definitions	*locar:* ***dar de aluguer ou de arrendamento*** (2) *arrendar: dar de renda* (1) *alugar: dar ou tomar de aluguer* (1) *locar: meter em loca* (0)
Synset	{*agraciação, bendição, benção, graça*}
Relevant relations	...hyponym-of {*ação, atitude*} ...result-of {benzer, abençoar, bençoar, abendiçoar} ...hyponym-of {*palavra, expressão, frase, vocábulo, dito, dicção, dizer*} ...hyponym-of {*sentimento, sentir, opinião, voz*}
Definitions	*bênção:* ***ação de benzer ou abençoar*** *(4)* *bênção:* ***palavras e sentimentos de gratidão*** *(4)* *graça: dicção espirituosa* (2) *bênção: chute frontal, que atinge o oponente no abdómen, no peito ou rosto* (0)

Table 3. Examples of synsets, relevant relations and scored definitions.

Most of the definitions not selected apply for only one or two words of the synset. Some are not informative enough, but this problem also concerns dictionaries. Also, when more than one definition is selected, they can be complementary, as in the last example.

5.3 Evaluation

In order to have an idea on the quality of the results, we conducted a manual evaluation of a subset of the definitions assigned using DA and Wiktionary.PT. For this purpose, we selected a random sample with 364 definitions and the synset they were assigned to, as long as the synset had two or more words, because single-word synsets were expected to have always a valid definition. Then, we asked two human judges to independently score the validity of each definition for its synset. We consider that a definition is valid for a word if it adds correct information about one of its meanings (e.g. its type, constituents,

antonyms, or other properties). Therefore, we defined the following scale for scoring synset-definition pairs:

0: Definition is not valid for any word of the synset;
1: Definition is valid for only one word of the synset;
2: Definition is valid for only two words of the synset, but the synset has more words;
3: Definition is valid for all or at least three words of the synset.

We believe that suiting at least three words of the synset is a strong hint that the definition is valid for the whole synset. This way, we minimise both the required effort for manual evaluation and the possible impact of Onto.PT errors. Table 4 shows the evaluation results according to the judge. Almost 80% of the scored synset-definition pairs were considered to be valid by both judges. As expected, definitions were always valid for at least one word, because the words in the synset were used as keys to get the definitions from the dictionaries. The inter-annotator agreement was 87.1% for an expected agreement of 66.3%. Consequently, Cohen's Kappa is 0.62, which corresponds to strong agreement between judges.

Judge	**Synset valid words**			
	0	**1**	**2**	**≥3**
J1	0.0%	16.8%	3.8%	79.4%
J2	0.0%	16.2%	3.8%	79.9%

Table 4. Number of synset words valid for the gloss, depending on the judge.

We recall that these results correspond just to the synsets with more than one word, which are 14,639, about 40% of the synsets with definition. This proportion is higher than in Onto.PT, where only about 28% of the synsets have size greater than 1. If we consider that, at least using these evaluation criteria, the single-word synsets have always a suitable definition, accuracy rises to more than 90%. So, despite some room for improvement, we believe these are promising results.

Although there are differences in the experimentation, used resources and in the evaluation strategies, our results are not very far from those reported for the mapping between GermaNet and Wiktionary [9]. Their definitions were 93.8%, 93.3% and 93.2% accurate, respectively using only synonyms, hypernymy, and hyponymy relations. To increase recall though, they used all relations, which lowered accuracy to 91.9%. Anyway, even though the reliability issues of Onto.PT might have lowered our results, they did not have a big impact. This shows that its wordnet-like structure makes sense and, at least in this kind of task, there is enough correct information to enable some tolerance to existing errors.

6 Conclusions and Future Work

We have described the first steps towards the automatic enrichment of a Portuguese wordnet with dictionary definitions, which thus become synset glosses. For this purpose, we were inspired by previous work for German [9] and used

only freely available resources for Portuguese. These include Onto.PT, a Portuguese wordnet created automatically, and two dictionaries, namely DA and Wiktionary.PT. We have also made some experiments using Wikipedia.PT. In the end, with the three definition sources, more than 35% of the synsets of Onto.PT had, at least, one definition. The correspondence between definitions and synsets was estimated to be close to 80% accurate.

The results are promising and it is in our plans to make the selected definitions and their synset correspondences publicly available. We are positive that they will be important to improve Onto.PT and broaden the range of tasks where it can be used, and also that the results will improve for newer versions of Onto.PT. But there are experiments left to do. First, we should analyse the impact of using different synset relations when ranking definitions, or giving different weights to different relations. So far, we only increased hypernymy's weight. Second, we will devise using alternative definition sources, in order to improve the coverage. We could perhaps exploit web search engine snippets to collect more definitions, similarly to what others have done [17]. It would also be interesting to exploit the web interfaces of proprietary dictionaries but, due to existing copyrights, this is probably on the edge of being legal.

References

1. Harabagiu, S.M., Moldovan, D.I.: Enriching the wordnet taxonomy with contextual knowledge acquired from text. In: Natural language processing and knowledge representation: language for knowledge and knowledge for language. MIT Press (2000) 301–333
2. Navigli, R.: Word sense disambiguation: A survey. ACM Computing Surveys **41**(2) (2009) 1–69
3. Fellbaum, C., ed.: WordNet: An Electronic Lexical Database (Language, Speech, and Communication). MIT Press (1998)
4. Vossen, P.: Introduction to EuroWordnet. Computers and the Humanities **32**(2) (1998) 73–89
5. Pianta, E., Bentivogli, L., Girardi, C.: MultiWordNet: developing an aligned multilingual database. In: 1st International Conference on Global WordNet. (2002)
6. de Melo, G., Weikum, G.: On the utility of automatically generated wordnets. In: Proceedings of 4th Global WordNet Conference. GWC, University of Szeged (2008) 147–161
7. Gonçalo Oliveira, H., Antón Pérez, L., Gomes, P.: Integrating lexical-semantic knowledge to build a public lexical ontology for Portuguese. In: Natural Language Processing and Information Systems, Proceedings of 17h International Conference on Applications of Natural Language to Information Systems (NLDB). Volume 7337 of LNCS., Springer (2012) 210–215
8. Dias-da-Silva, B.C., Felippo, A.D., das Graças Volpe Nunes, M.: The automatic mapping of Princeton WordNet lexical-conceptual relations onto the Brazilian Portuguese WordNet database. In: Proceedings of the 6th International Conference on Language Resources and Evaluation. LREC, ELRA (2008)
9. Henrich, V., Hinrichs, E., Vodolazova, T.: Semi-automatic extension of GermaNet with sense definitions from Wiktionary. In: Proceedings of 5th Language & Technology Conference. LTC 2011 (2011) 126–130

10. Pasca, M., Harabagiu, S.M.: The informative role of WordNet in open-domain question answering. In: Proceedings of NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations, Pittsburgh, USA (2001) 138–143
11. Kunze, C., Lemnitzer, L.: GermaNet - representation, visualization, application. In: Proceedings of 3rd International Conference on Language Resources and Evaluation. LREC, ELRA (2002) 1485–1491
12. Vanderwende, L., Kacmarcik, G., Suzuki, H., Menezes, A.: MindNet: An automatically-created lexical resource. In: Proceedings of HLT/EMNLP 2005 Interactive Demonstrations, Vancouver, Canada, ACL Press (2005) 8–9
13. Gonçalo Oliveira, H., Santos, D., Gomes, P.: Extracção de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação. Linguamática **2**(1) (May 2010) 77–93
14. Marrafa, P., Amaro, R., Mendes, S.: WordNet.PT Global – extending WordNet.PT to Portuguese varieties. In: Proceedings of the 1st Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties, ACL Press (2011) 70–74
15. Maziero, E.G., Pardo, T.A.S., Felippo, A.D., Dias-da-Silva, B.C.: A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. In: VI Workshop em Tecnologia da Informação e da Linguagem Humana (TIL). (2008) 390–392
16. de Paiva, V., Rademaker, A., de Melo, G.: Openwordnet-pt: An open brazilian wordnet for reasoning. In: Proceedings of the 24th International Conference on Computational Linguistics. COLING (Demo Paper) (2012)
17. Velardi, P., Navigli, R., D'Amadio, P.: Mining the web to create specialized glossaries. IEEE Intelligent Systems **23**(5) (September 2008) 18–25
18. Cui, H., Kan, M.Y., Chua, T.S.: Soft pattern matching models for definitional question answering. ACM Transactions on Information Systems **25**(2) (2007)
19. Duan, W., Yates, A.: Extracting glosses to disambiguate word senses. In: Proceedings of Human Language Technologies: 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics. HLT, ACL Press (2010) 627–635
20. Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, C.M., Wirth, C.: UBY - a large-scale unified lexical-semantic resource. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. EACL, Avignon, France, ACL Press (2012) 580–590
21. Navigli, R., Ponzetto, S.P.: BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence **193** (December 2012) 217–250
22. Lesk, M.: Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In: Proceedings of the 5^{th} Annual International Conference on Systems documentation. SIGDOC, ACM (1986) 24–26
23. Dolan, W.B.: Word sense ambiguation: clustering related senses. In: Proceedings of 15th International Conference on Computational Linguistics. COLING, ACL Press (1994) 712–716
24. Simões, A., Sanromán, A.I., ao Almeida, J.J.: Dicionário-Aberto: A source of resources for the Portuguese language processing. In: Proceedings of Computational Processing of the Portuguese Language, 10th International Conference (PROPOR). Volume 7243 of LNCS., Springer (2012) 121–127
25. Anton Pérez, L., Gonçalo Oliveira, H., Gomes, P.: Extracting lexical-semantic knowledge from the portuguese wiktionary. In: Proceedings of the 15th Portuguese Conference on Artificial Intelligence. EPIA, APPIA (2011) 703–717