Onto.PT: recent developments of a large public domain Portuguese wordnet

Hugo Gonçalo Oliveira
CISUC, University of Coimbra
Portugal
hroliv@dei.uc.pt

Paulo Gomes
CISUC, University of Coimbra
Portugal
pgomes@dei.uc.pt

Abstract

This document describes the current state of Onto.PT, a new large wordnet for Portuguese, freely available, and created automatically after exploiting and integrating existing lexical resources in a wordnet structure. Besides an overview on Onto.PT, its creation and evaluation, we enumerate the developments of version 0.6. Moreover, we provide a quantitative view on this version, its comparison to other Portuguese wordnets, in terms of contents and size, as well as some details about its global coverage and availability.

1 Introduction

Onto.PT is a new wordnet-like resource for Portuguese. It is under development since 2009 in the Center for Informatics and Systems of the University of Coimbra, after we realised the limitations of existing Portuguese wordnets and related resources. Onto.PT was one of the main contributions of Hugo Gonçalo Oliveira's PhD (Gonçalo Oliveira, 2013), concluded on May 2013, under the supervision of Paulo Gomes. Since then, several developments were made and a new version (v0.6) was released.

Likewise Princeton WordNet (PWN, Fellbaum (1998)), Onto.PT is freely available but, in opposition to the previous resource and most wordnets, it is created automatically, after the exploitation of existing public lexical resources. While the latter fact led to a resource which may not be 100% reliable, it also enabled the development of a larger resource and with a wider coverage, as compared to other Portuguese wordnets. This makes Onto.PT a viable alternative for several natural language processing tasks. Having this in mind, in order to ease the integration of Onto.PT with other applications, this resource is available as a standard model for knowledge representation, namely the Resource Description Framework (RDF, Miller and Manola (2004)).

In the rest of this document, we give a brief overview on the creation of Onto.PT, where several lexical resources for Portuguese are exploited and integrated in a wordnet-like structure, across four automatic steps that combine different information extraction techniques. We then highlight the developments that lead to version 0.6. After that, we describe the contents of Onto.PT, compare it with other wordnets for Portuguese, and provide some details on its availability and global coverage. The latter reports the results of finding suitable matches between Onto.PT synsets and the so-called "core" wordnet concepts. We conclude with additional information on the utility of Onto.PT and leave ideas for future work.

2 Creation

The creation of Onto.PT follows ECO, an automatic approach for creating wordnets, described briefly in this section, and more extensively elsewhere (Gonçalo Oliveira and Gomes, 2013a). Also in this section, we enumerate the resources integrated in the current version of Onto.PT and how they were exploited. The section ends with a brief reference to the evaluation of Onto.PT.

2.1 The ECO approach

Originally, ECO consisted of three main steps, that combine different information extraction techniques, namely:

1. **Extraction:** exploitation of regularities in textual sources to extract instances of semantic relations, connecting plain words – e.g. [*vírus* causation-of *doença*] ([*virus* causation-of *disease*])

2. **Synset discovery**:

(a) Computation of graph-based similarities between the extracted synonymy instances and available synsets, as those in existing thesauri, if available. When there is enough confidence, the synonymy instances are added to suitable synsets – e.g. [*comutar* synonym-of *mutuar*] + {*trocar, permutar, mutuar*} → {*trocar, permutar, mutuar, comutar*} ([*interchange* synonym-of *exchange*] + {*change, swap, exchange*} → {*change, swap, exchange, interchange*})

(b) Cluster discovery on the remaining synonymy instances and inclusion of the identified clusters as new synsets – e.g. [*tiritante* synonym-of *trémulo*] ∧ [*trémulo* synonym-of *convulso*] ∧ [*convulso* synonym-of *tiritante*] → {*tiritante,trémulo,convulso*} ([*shivering* synonym-of *trembling*] ∧ [*trembling* synonym-of *shaking*] ∧ [*shaking* synonym-of *shivering*] → {*shivering,trembling,shaking*})

3. **Ontologisation**: Computation of graph-based similarity measures to integrate the rest of the relations, by assigning each argument to a suitable synset – e.g. [*iluminar* purpose-of *vela*] → {*iluminar, candear*} purpose-of {*vela, tocha, lume*} ([*illuminate* purpose-of *candle*] → {*illuminate, light_up*} purpose-of {*candle, torch, fire*})

Recently, a fourth step was added to ECO:

4. **Definition assignment**: selection of suitable dictionary definitions for the discovered synsets. Definitions might work as glosses, also common in wordnets – e.g. {*multidão, massa*}: *grande quantidade de pessoas* ({*crowd, mass*}: *great amount of people*)

2.2 Integrated resources

The current version of Onto.PT includes lexical-semantic information acquired from six public domain lexical resources of Portuguese, namely:

- The relation instances of PAPEL (Gonçalo Oliveira et al., 2009), a lexical-semantic network extracted automatically from a proprietary Portuguese dictionary. Those are represented by {<*arg1*> RELATION-TYPE <*arg2*>} with words as arguments, and a rich set of relation types that include, for instance, synonymy, hypernymy, several types of meronymy, causation, purpose-of and property-of.
- The definitions and relations instances, extracted from Dicionário Aberto (DA, Simões et al. (2012)) and from the Portuguese Wiktionary (Wikt.PT)[1], both open dictionaries;
- The antonymy instances and synsets of TeP (Maziero et al., 2008), an electronic thesaurus, created manually by experts;
- The synsets of OpenThesaurus.PT (OT.PT)[2], another electronic thesaurus, smaller than TeP, and created collaboratively;
- More recently, the synsets of OpenWordNet.PT (OWN.PT, de Paiva et al. (2012)), a Portuguese wordnet obtained after the translation of part of PWN.

In the first step of ECO, DA and Wikt.PT are exploited using the grammars developed during the creation of PAPEL, which are distributed freely[3]. The extracted relation instances are merged with those from PAPEL's network thus originating a larger lexical-semantic network where words are connected by semantic relations.

Then, the synonymy instances extracted from the dictionaries, as well as those of OT.PT, are assigned to suitable synsets, according to their similarity. Clusters are discovered in a synonymy network established by the unassigned synonymy instances, and added as new synsets.

After that, the arguments of the non-synonymy relations are assigned to the discovered synsets, thus becoming synset relations. Antonymy relations from TeP are also added in this step. Finally, when possible, the synsets have assigned suitable definitions from DA and Wikt.PT (see more in Gonçalo Oliveira and Gomes (2013b)).

2.3 Evaluation

Besides occasional evaluations of each step of ECO, which guided us in the selection of the appropriate parameters, a previous version of Onto.PT (v0.3.5) was the target of an extensive manual evaluation where synsets and synset relations were evaluated by two human judges[4]. We estimated that about 81% to 85% of the synsets were correct. More precisely, for the synsets with

[1] See `http://pt.wiktionary.org`

[2] See `http://openthesaurus.caixamagica.pt/`

[3] See `http://www.linguateca.pt/PAPEL`

[4] See additional details in section 8.3 of Hugo Gonçalo Oliveira's PhD thesis (Gonçalo Oliveira, 2013)

more than one word, 73.9% were classified as correct and 7.5% as incorrect by both judges. For the remaining 18.6% synsets, there was no agreement. As for the relations, considering only correct synsets, hypernymy relations were estimated to be about 79% accurate, with a κ agreement of 0.47. A set containing relations of the other types got between 88% and 92% accuracy, depending on the judge, with a κ agreement of 0.48.

The accuracy of the definition assignment step was estimated to be between 79-80% for Onto.PT v0.4.1, with 0.62 κ agreement. This number should be similar in Onto.PT v0.6, because no big changes were made.

3 Developments of Onto.PT v.0.6

The most recent version of Onto.PT was released after some progress regarding, namely: improvements in the creation process, integration of the OWN.PT synsets, removal of redundant hypernymy instances, and the availability of synset definitions. This also lead to improvements on the resource evaluation.

3.1 Procedural improvements

Onto.PT v0.6 was created after several improvements on the previous versions, including:

- The refinement of some extraction patterns, after exploring the results of previous evaluations;
- Increasing the synonymy attachment threshold to improve synset accuracy.

3.2 Integration of OpenWordNet.PT

For the first time, in the creation of Onto.PT, we took advantage of OWN.PT and integrated part of its contents. More precisely, TeP and OWN.PT were merged before synset discovery, in order to create a single synset resource. For this purpose, synsets with high word intersections are clustered – e.g. {*praia, beira-mar, borda, litoral, ribamar*} + {*praia, beira-mar, litoral, costa*} → {*praia, beira-mar, borda, litoral, ribamar, costa*} ({*beach, seaside, seaboard, seashore*} + {*beach, seaside, coast*} → {*beach, seaside, seaboard, seashore, coast*})

3.3 Removal of redundant hypernymy

In order to move towards a better-formed taxonomic tree, redundant hypernymy relation instances in Onto.PT were removed. These instances are those that may be inferred by transitivity – e.g. ~~{*animal*} hypernym-of {*porco, suíno*}~~ ∧ {*animal*} hypernym-of {*mamífero, mastozoário*} ∧ {*mamífero, mastozoário*} hypernym-of {*porco, suíno*} (~~{*animal*} hypernym-of {*pig, swine*}~~ ∧ {*animal*} hypernym-of {*mammal, mammalian*} ∧ {*mamífero, mammalian*} hypernym-of {*pig, swine*})

3.4 Synset definitions

Although the first experiments on assigning definitions to the synsets of Onto.PT were done with version 0.4.1 of the resource, definitions were only made available together with the resource in version 0.6. We recall that these definitions might work as glosses.

3.5 New evaluation results

Given that a similar extensive evaluation effort required much time, we reused the classified synsets and synset relation instances from Onto.PT v0.3.5[5] to speculate on the current quality of Onto.PT. Depending on the judge, the new evaluation led respectively to synset accuracy between 89-93%, hypernymy accuracy between 79-100%, and accuracy of other relations between 93-96%.

These results should, nevertheless, be analysed more carefully in the future. While a substantial amount of incorrect contents were removed or corrected, a lower, but still substantial, number of contents that were previously classified as correct were also removed.

4 Contents and Availability

This section presents a quantitative view on the contents of Onto.PT v0.6, including the covered relations types, a comparison to other Portuguese wordnets, and its global coverage. Details about the availability of Onto.PT are provided in the end of this section.

4.1 Quantitative view

Onto.PT v0.6 contains almost 169k unique lexical items, organised in about 117k synsets, which are connected by almost 174k relation instances. Table 1 shows the distribution of covered lexical items, according to their part-of-speech (POS), and included synsets according both to their POS and number of words (size).

Table 2 shows the set of covered semantic relations, richer than in typical wordnets, as well as their quantities. In fact, these are relation types

[5]Datasets available at `http://ontopt.dei.uc.pt`

POS	Lexical Items	Synsets		
		size = 1	size > 1	Total
Nouns	97,531	44,495	23,378	67,873
Verbs	32,958	20,723	5,728	26,451
Adjectives	34,392	10,909	9,851	20,760
Adverbs	3,995	1,283	1,083	2,366
Total	168,876	77,410	40,040	**117,450**

Table 1: Onto.PT v0.6 synsets.

originally defined during the creation of PAPEL, after the analysis of frequent patterns in dictionary definitions. In this set, for each relation type, there are different subtypes, depending on the POS of the accepted arguments. For instance, [x purpose-of y] has the following subtypes:

- $noun(x)$ *fazSeCom* $noun(y)$
  $\rightarrow$ x is performed or obtained with y
- $noun(x)$ *fazSeComAlgoComPropriedade* $adj(y)$
  $\rightarrow$ x is performed or obtained with something that is y
- $verb(x)$ *finalidadeDe* $noun(y)$
  $\rightarrow$ x is an action performed with y
- $verb(x)$ *finalidadeDeAlgoComPropriedade* $adj(y)$
  $\rightarrow$ x is an action performed with something that is y

Different types of meronymy are also covered, namely part-of, member-of, contained-in and material-of. Moreover, for each relation subtype, there is an inverse type (e.g. [x causadorDe y] $\rightarrow$ [y resultadoDe x]), except for antonymy, which is a symmetric relation. If we consider the inverse subtypes, Onto.PT has about 341k relation instances.

4.2 Comparison with Portuguese wordnets

Though it is commonly referred that there is not a wordnet for Portuguese, this is not completely true. The problem is that all wordnet projects targeting Portuguese have strong limitations. To our knowledge, besides Onto.PT, there are other four resources – Wordnet.PT (WN.PT, Marrafa et al. (2011)), Wordnet.Br (WN.Br, Dias-da-Silva (2006)), MultiWordNet.PT (MWN.PT)[6] and OpenWordnet.PT (OWN.PT, de Paiva et al. (2012)) – listed in Table 3, together with some information on their creation and availability.

¿From those, besides Onto.PT, only OWN.PT is freely available[7]. The synsets of WN.Br are free,

Resource	Availability	Creation
WN.PT	web interface no download	manual
WN.Br	free synsets	man. (synsets) from PWN (relations)
MWN.PT	paid license	man. translation (synsets) from PWN (relations)
OWN-PT	free	man. translation (synsets) from PWN (relations)
Onto.PT	free	automatic

Table 3: Portuguese WNs: availability & creation

with the name of TeP (Maziero et al., 2008), but the relations, inherited from PWN given manual synset correspondences, are not. MWN.PT is not free but it is available upon a paid license. However, this resource only covers nouns, while all the others cover verbs, adjectives and adverbs as well.

All but WN.PT and Onto.PT follow a translation approach, one of the most popular alternatives to the creation of non-English wordnets, where PWN is translated to a target language (de Melo and Weikum, 2008). This approach is followed at different levels by WN.Br, MWN.PT and OWN.PT. In WN.Br, the synsets were created specifically for Portuguese and manual correspondences to PWN were defined afterwards. On the other hand, the synsets of MWN.PT and OWN.PT are, as far as possible, the direct translation of a set of key PWN synsets. But a problem arises for this kind of approaches. Different languages represent different socio-cultural realities, they do not cover exactly the same part of the lexicon and, even where they seem to be common, several concepts are lexicalised differently (Hirst, 2004). This explains the existence of "lexical gaps" in some MWN.PT synsets. We thus believe that, whether created manually, semi-automatically or automatically, a wordnet should be developed from scratch for its target language. Only after that, it should be devised to align part of the synsets to wordnets of other languages, but having in mind that some rich meanings might be lost in the translation process.

Table 4 presents the same wordnets regarding their size, in terms of covered lexical items, included synsets, semantic relations and the presence of glosses written in Portuguese. Regarding the last property, the wordnets relying on translation do not contain glosses in Portuguese, even though the English glosses can potentially be inherited from PWN and translated. WN.PT has contained Portuguese glosses for a long time. And since the last version of Onto.PT, part of its synsets

[6] See `http://mwnpt.di.fc.ul.pt/`

[7] OWN.PT is available from `https://github.com/arademaker/wordnet-br` and distributed in two main RDF files, one with the synsets and their PWN match, and another with PWN, including relations, glosses and other inheritable properties.

Relation	Args	Given name	Instances
Hypernymy	n, n	*hiperonimoDe*	79,425
Part	n, n	*parteDe*	3,782
	n, adj	*parteDeAlgoComPropriedade*	4,922
	adj, n	*propriedadeDeAlgoParteDe*	101
Member	n, n	*membroDe*	5,957
	n, adj	*membroDeAlgoComPropriedade*	111
	adj, n	*propriedadeDeAlgoMembroDe*	922
Contained	n, n	*contidoEm*	365
	n, adj	*contidoEmAlgoComPropriedade*	272
Material	n, n	*materialDe*	879
Causation	n, n	*causadorDe*	1,396
	n, adj	*causadorDeAlgoComPropriedade*	30
	adj, n	*propriedadeDeAlgoQueCausa*	667
	v, n	*accaoQueCausa*	8,168
	n, v	*causadorDaAccao*	84
Producer	n, n	*produtorDe*	1,662
	n, adj	*produtorDeAlgoComPropriedade*	80
	adj, n	*propriedadeDeAlgoProdutorDe*	463
Purpose	n, n	*fazSeCom*	6,787
	n, adj	*fazSeComAlgoComPropriedade*	77
	v, n	*finalidadeDe*	8,507
	v, adj	*finalidadeDeAlgoComPropriedade*	338
Location	n, n	*localOrigemDe*	1,458
Quality	n, n	*temQualidade*	977
	adj, n	*devidoAQualidade*	1,118
State	n, n	*temEstado*	334
	adj, n	*devidoAEstado*	197
Property	adj, n	*dizSeSobre*	9,769
	adj, v	*dizSeDoQue*	24,131
Manner	adv, n	*maneiraPorMeioDe*	1,976
	adv, adj	*maneiraComPropriedade*	1,675
Manner without	adv, n	*maneiraSem*	231
	adv, v	*maneiraSemAccao*	20
Antonymy	n, n	*antonimoNDe*	2,300
	adv, adv	*antonimoAdvDe*	127
	adj, adj	*antonimoAdjDe*	2,475
	v, v	*antonimoVDe*	1,844
		Total	**173,627**

Table 2: Onto.PT v0.6 relations and their quantities

also contain glosses, automatically selected from dictionaries (see section 2).

The numbers on the size of the Portuguese wordnets are put side-by-side to those of PWN, to show that they are substantially smaller, except for Onto.PT. Despite being the second youngest Portuguese wordnet (OWN.PT is the youngest), Onto.PT has a size comparable to PWN, and it covers a richer set of semantic relations. We should recall that Onto.PT integrates several public resources for Portuguese, including the synsets of WN.Br (TeP) and of OWN.PT, so it was expected to be larger than those two.

Although size is probably not the most important property of a wordnet, these numbers show the benefits of an automatic creation. Besides typically larger resources, automatic approaches provide a faster creation, an easier maintenance, and a higher growth potential, in a trade-off on the virtual 100% reliability. Therefore, in the case of Portuguese, selecting the most adequate(s) wordnet(s) to use in some project should consider, among others, the language coverage needs, the tolerance to errors and the available budget.

Resource	Lexical items	Synsets	Relations	Glosses (in PT)
WN.PT	11k	13k	40k	Yes
WN.Br	44k	20k	N/A	No
MWN.PT	16k	17k	69k	No
OWN.PT	48k	39k	83k	No
Onto.PT	169k	117k	341k	Yes (40%)
PWN 3.0	155k	118k	285k	Yes (EN)

Table 4: Portuguese WNs: contents

4.3 Global coverage

The Global WordNet Association provides several lists of key concepts that should be present in wordnets. One of them, contains a reduced set of

164 Core Base Concepts which can be seen as the most important in the wordnets of four languages[8]. They are divided into 98 abstract and 66 concrete concepts, and are represented as PWN 1.5 synsets.

We used this set to speculate on the global coverage of Onto.PT v0.6. For this purpose, we made manual rough matches between the 164 base concepts and suitable Onto.PT synsets. We concluded that Onto.PT roughly covers most of the concepts in the list, more precisely 95 abstract and 66 concrete synsets (98%). The three uncovered concepts are the following: {*change magnitude, change size*}, {*spacing, spatial arrangement*} and {*visual property*}. As one can see, they denote abstract generic classes which are sometimes created artificially, in order to work as the hypernym of a set of more specific concepts. We should add that the global coverage increased since Onto.PT v0.3.5, where 93% base concepts were covered. The integration of OWN.PT had a positive impact on this improvement.

Looking at the other Portuguese wordnets, we can say that, given that WN.PT was created in EuroWordNet's framework, it should cover all the 164 concepts. Moreover, the website of MWN.PT refers that it covers all these concepts. However, MWN.PT only contains nouns, while 35 of the abstract concepts are verbs. So, this information is probably incorrect.

4.4 Availability

Onto.PT and related resources are freely available from `http://ontopt.dei.uc.pt`. There, the resource can be downloaded as a RDF model, and in two different notations, RDF/XML and the more compact N3. This model is based on the WordNet RDF/OWL basic representation (van Assem et al., 2006) that was adapted for Portuguese and to include our broader relation set. Moreover, Onto.PT may be browsed through an online interface, OntoBusca, very similar to the PWN search interface and available from the previous website.

5 Concluding remarks

We believe that Onto.PT is a valuable add to the Portuguese wordnets and an important contribution to Portuguese NLP, that may be useful in a broad range of tasks. So far, previous versions of Onto.PT were used in query expansion and we have shown that it can be used for word sense disambiguation[9]. And we have some preliminary results of exploiting Onto.PT and OWN.PT for answering open domain cloze question automatically – the results show that, due to its larger size, more questions are answered correctly using Onto.PT.

We should add that Portuguese was recently added to range of languages covered by the multilingual knowledge base BabelNet (Navigli and Ponzetto, 2012). This resource integrates PWN with Wikipedia and some open wordnets, in a very large ontology. Therefore, from this moment, BabelNet should also be seen as one more alternative to Portuguese wordnets. Or, perhaps, as a complement, because, despite its large size (9M synsets in all languages), BabelNet integrates both lexical and world knowledge and the Portuguese Wikipedia (about 800k articles) is still small when compared, for instance, to the English (about 4.3M) and the German (about 1.63M).

We recall that Onto.PT is created automatically and is not a static resource, but an ongoing project. Therefore, improvements are expected in the future. Among other ideas, we are devising the conversion of Onto.PT to specific representations for lexical ontologies (e.g. Lemon, Buitelaar et al. (2009)), we are considering to assign confidence values to its contents and to exploit the World Wide Web for more synset definitions, and we are studying approaches for aligning it to PWN, given that the Onto.PT synsets are not static. We are also devising the integration of the relations of OWN.PT. In fact, with ECO, we can likewise integrate knowledge from additional sources including, for instance, Wikipedia, but keeping in mind that most information in Wikipedia is out of the scope of classic wordnets.

For more information on ECO and on Onto.PT, please refer to Hugo's PhD thesis (Gonçalo Oliveira, 2013) or to our article in the Language and Resources Evaluation Journal (Gonçalo Oliveira and Gomes, 2013a).

Acknowledgements

The development of Onto.PT v0.6 has been supported by the iCIS project (CENTRO-07-ST24-FEDER-002003), co-financed by QREN, in the scope of the Mais Centro Program and European Union's FEDER.

[8] Available from `http://w.globalwordnet.org/gwa/ewn_to_bc/corebcs.html`

[9] See section 8.4 of Hugo Gonçalo Oliveira's PhD thesis (Gonçalo Oliveira, 2013)

References

Paul Buitelaar, Philipp Cimiano, Peter Haase, and Michael Sintek. 2009. Towards linguistically grounded ontologies. In *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*, ESWC 2009, Heraklion, Crete, Greece. Springer. Pages 111–125.

Gerard de Melo and Gerhard Weikum. 2008. On the utility of automatically generated wordnets. In *Proceedings of 4th Global WordNet Conference*, GWC 2008, Szeged, Hungary. University of Szeged. Pages 147–161.

Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. OpenWordNet-PT: An open brazilian wordnet for reasoning. In *Proceedings of the 24th International Conference on Computational Linguistics*, COLING (Demo Paper).

Bento C. Dias-da-Silva. 2006. Wordnet.Br: An exercise of human language technology research. In *Proceedings of the 3rd International WordNet Conference*, GWC 2006, South Jeju Island, Korea, January. Pages 301–303.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.

Hugo Gonçalo Oliveira and Paulo Gomes. 2013a. ECO and Onto.PT: A flexible approach for creating a Portuguese wordnet automatically. *Language Resources and Evaluation*, to be published.

Hugo Gonçalo Oliveira and Paulo Gomes. 2013b. On the automatic enrichment of a Portuguese wordnet with dictionary definitions. In *Advances in Artificial Intelligence, Local Proceedings of the 16th Portuguese Conference on Artificial Intelligence*, EPIA 2013, Angra do Heroísmo, Azores, Portugal. APPIA. Pages 486–497.

Hugo Gonçalo Oliveira, Diana Santos, and Paulo Gomes. 2009. Relations extracted from a portuguese dictionary: results and first evaluation. In *Proceedings of 14th Portuguese Conference on Artificial Intelligence*, EPIA 2009. APPIA, October. Pages 541–552.

Hugo Gonçalo Oliveira. 2013. *Onto.PT: Towards the Automatic Construction of a Lexical Ontology for Portuguese*. Ph.D. thesis, University of Coimbra. `http://eden.dei.uc.pt/~hroliv/pubs/GoncaloOliveira_PhdThesis2012.pdf`.

Graeme Hirst. 2004. Ontology and the lexicon. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, International Handbooks on Information Systems. Springer. Pages 209–230.

Palmira Marrafa, Raquel Amaro, and Sara Mendes. 2011. 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*, Edinburgh, Scotland, July. ACL Press. Pages 70–74.

Erick G. Maziero, Thiago A. S. Pardo, Ariani Di Felippo, and Bento C. Dias-da-Silva. 2008. 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)*, pages 390–392.

Eric Miller and Frank Manola. 2004. RDF primer. Published: W3C Recommendation.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, December.

Alberto Simões, Álvaro Iriarte Sanromán, and José Jo ao Almeida. 2012. Dicionário-Aberto: A source of resources for the Portuguese language processing. In *Proceedings of Computational Processing of the Portuguese Language, 10th International Conference (PROPOR 2012), Coimbra Portugal*, volume 7243 of *LNCS*. Springer, April. Pages 121–127.

Mark van Assem, Aldo Gangemi, and Guus Schreiber. 2006. RDF/OWL representation of WordNet. W3c working draft, World Wide Web Consortium, June.