An overview of Portuguese WordNets

Valeria de Paiva
Nuance Communications, USA
`valeria.depaiva@nuance.com`

Livy Real
IBM Research Brazil
`livyreal@gmail.com`

Hugo Gonçalo Oliveira
CISUC, DEI, Univ. Coimbra, Portugal
`hroliv@dei.uc.pt`

Alexandre Rademaker
IBM Research and FGV/EMAp, Brazil
`alexrad@br.ibm.com`

Claudia Freitas
PUC-Rio, Brazil
`claudiafreitas@puc-rio.br`

Alberto Simões
Universidade do Minho, Portugal
`ambs@ilch.uminho.pt`

Abstract

Semantic relations between words are key to building systems that aim to understand and manipulate language. For English, the "de facto" standard for representing this kind of knowledge is Princeton's WordNet. Here, we describe the wordnet-like resources currently available for Portuguese: their origins, methods of creation, sizes, and usage restrictions. We start tackling the problem of comparing them, but only in quantitative terms. Finally, we sketch ideas for potential collaboration between some of the projects.

1 Introduction

Semantic relations are a key aspect when developing computer programs capable of handling language – they establish (labeled) associations between words and can be integrated into lexical-semantic knowledge bases. Available since the beginning of the 1990s, Princeton's WordNet (Fellbaum, 1998), henceforth PWN, is a paradigmatic lexical resource. Originally created for English, its model is now a "de facto" standard, due to its wide use in applications and its adaptation to different languages.

For Portuguese, the first resource of this kind, WordNet.PT (Marrafa, 2001), was announced in 2001 but, unlike PWN, was never free to use. This meant that, in practice, there was still no open Portuguese wordnet. In parallel, a few alternatives to the wordnet model arose, some of which were compared in (Santos et al., 2010). But if those alternatives proved themselves useful for some tasks, they were not enough to enable all of the standard uses of a wordnet in Natural Language Processing (NLP), including similarity computation or word sense disambiguation. As the need for a Portuguese wordnet was keenly felt, in the early 2010s, several projects sprung up aiming to develop free Portuguese wordnets. We describe some of those wordnets, while indicating where they were created, their construction process, their availability and, when possible, their size.

We recall the wordnet model, its adaptation to other languages, and how these adaptations may be expanded through content alignment. Then, we describe the Portuguese wordnets we are aware of, alternative lexical-semantic resources, and go on to focus on the open wordnets. After that, we briefly compare the previous along a set of relevant features for processing Portuguese. Then, we suggest work leveraging what is already planned for these wordnets, as well as some ideas for collaboration. Knowing where we are in terms of our wordnets is an essential first step in establishing lexical resources, which are vital to the computational processing of the Portuguese language. [1]

2 WordNet and Alternatives

Lexical knowledge bases are organized repositories of lexical items, usually including information about the possible meanings of words, relations

---

[1] This paper is a shorter English version of our previous article, in Portuguese (Gonçalo Oliveira et al., 2015).

between them, definitions, and phrases that exemplify their use. The Princeton WordNet model, with English as its target language, is probably the most popular representative of this type of lexical knowledge base. Its flexibility has led not only to its growing use by the NLP community, but also to the adaptation of the model to other languages.

PWN was created manually in the early 1990s and has been updated several times since then. Initially based on psycholinguistic principles, it combines traditional lexicographic information, similar to that in a dictionary, with an appropriate organization for computational use, which facilitates its application as a basis for lexical-semantic knowledge. Like a thesaurus, PWN is organized in groups of synonymous lexical items, called *synsets*, which can be seen as the possible lexicalizations for the concepts in the language. Besides synonymy, inherent to synsets, PWN covers other types of semantic relation between synsets. For example, hypernymy – a concept is a generalization of another – or meronymy – a concept is a part of another. In addition, each synset has a part-of-speech (noun, verb, adjective or adverb); a gloss, similar to a definition in a dictionary; and it may still have phrases that illustrate its use. The inclusion of a lexical item in a synset indicates a sense of that item.

Both its free availability and the flexibility of its model were crucial to the success and widespread use of PWN. This made it possible to integrate PWN into a large number of NLP or knowledge management projects, making it virtually the standard model of a lexical resource for several languages. The popularity of the PWN knowledge base model led to the creation of the Global WordNet Association (GWA), a non-commercial organization that provides a platform for discussion, sharing and linking the wordnets of the world.

2.1 Multilingual Wordnets

Many people have studied the possibility of aligning, as far as possible, wordnets of different languages, given their similarities. Thus, the unveiling of multilingual wordnets such as EuroWordNet (Vossen, 1997) or MultiWordNet (Pianta et al., 2002), which nonetheless follow very different approaches. In EuroWordNet, wordnets are created independently for each language, and only after that they are aligned, rely-

ing on similarities or, indirectly, using Princeton WordNet as a pivot, through the so-called Inter-Language Index (ILI). In MultiWordNet, the first step was to translate, as much as possible, one wordnet, usually Princeton's, into the other languages. Among the multilingual wordnets aligned with PWN, there are, for instance, BalkaNet (Stamou et al., 2002), dedicated to the languages of the Balkans, and the Multilingual Central Repository (Gonzalez-Agirre et al., 2012) (henceforth, MCR) dedicated to the languages of Spain.

Open Multilingual WordNet (Bond and Foster, 2013), henceforth OMWN, is an initiative to facilitate access to different wordnets, for different languages. To this end, wordnets, created independently, were normalized using PWN, and then connected to each other and accessed through a common interface. Another initiative that should be mentioned is the Universal WordNet (de Melo and Weikum, 2009) (henceforth, UWN), a multilingual lexical knowledge base automatically built from PWN and the alignment of multilingual versions of Wikipedia.

There are also several projects on the alignment of PWN with other lexical resources or knowledge bases. These include, for instance, YAGO (Suchanek et al., 2007), UBY (Gurevych et al., 2012), BabelNet (Navigli and Ponzetto, 2012), SUMO (Pease and Fellbaum, 2010) and DOLCE (Gangemi et al., 2010).

2.2 Closed Portuguese WordNets

There is no doubt that the open-source character of PWN was key in its wide acceptance. Still, not all resources that followed on the footsteps of PWN have chosen to make their results freely available. We describe three projects that resulted in Portuguese wordnets that are not free to use.

WordNet.PT (Marrafa, 2001), henceforth WN.PT, was the first Portuguese wordnet, in development since 1998. Its construction is essentially manual and it follows the EuroWordNet (Vossen, 1997) model, which means WN.PT is created from scratch for Portuguese. WN.PT 1.6, released in 2006, covers a wide range of semantic relations, including: hypernym, whole/part, equivalence, opposition, categorization, instrument-for, or place-of. More recently, WN.PT was expanded to *Global WordNet.PT* (Marrafa et al., 2011), which contains 10,000 concepts, including nouns, verbs and

adjectives, their lexicalizations in different variants of Portuguese and their glosses, in a network of more than 40,000 relation instances. An approach to expand the WN.PT semi-automatically with relations extracted from a corpus (Amaro, 2014) was recently presented, which shows that the project is still active.

WordNet.BR (henceforth, WN.BR) aimed to be a wordnet for Brazilian Portuguese. In its first development phase (Dias-da-Silva et al., 2002), a team of linguists analyzed five Portuguese dictionaries and two corpora to collect information on synonymy and antonymy. This resulted in the manual creation of synsets and antonymy relations between them, and writing some glosses and example sentences. In a second phase (Dias-da-Silva, 2006), its synsets were manually aligned with PWN, in a similar process to that followed in the EuroWordNet project, using bilingual dictionaries. After this alignment, the semantic relations between synsets with equivalents in Portuguese and English were inherited. It is assumed that the full version of WN.BR covers relations of hyperonymy, part-of, cause and implication (entailment). However, this version is not available online. One can view and download the results of phase one, available under the name of Electronic Thesaurus of Portuguese (TeP) (Maziero et al., 2008). TeP includes more than 44,000 lexical items, organized into 19,888 synsets, which in turn are connected through 4,276 antonymy relations.

MultiWordNet.PT, commonly referred to as MWN.PT, is the Portuguese section of the MultiWordNet project (Pianta et al., 2002), which can be purchased through the European Language Resources Association catalog. MWN.PT includes 17,200 manually validated synsets, which correspond to approximately 21,000 senses and 16,000 lemmas, covering both European and Brazilian variants of Portuguese. As a resource established under the MultiWordNet project, its synsets are derived from the translation of their PWN equivalents. Transitively, this resource turns out to be also aligned with the MultiWordNets of Italian, Spanish, Hebrew, Romanian and Latin.

The manual creation of a wordnet is a complex task, which requires much effort and time. When it was not possible to use an open Portuguese wordnet, researchers working on the processing of Portuguese felt the need to develop free alternatives which, in most cases, were also simpler. Those include OpenThesaurus.PT, typically used to suggest synonyms in word processors; PAPEL (Gonçalo Oliveira et al., 2008), a lexical-semantic network, automatically extracted from a Portuguese dictionary, with words connected through a wide range of semantic relationships; the Port4Nooj lexical resources (Barreiro, 2010), which include a set of definitions and semantic relations between words; and the Dicionário Aberto (Simões et al., 2012), an open electronic dictionary which includes also several explicit relationships between words.

3 Open Portuguese Wordnets

Open wordnets for Portuguese finaly appeared in the early 2010s. They were created by automatic or semi-automatic means and all assume that lexical-semantic resources must be open-source to be really useful to the community. We present four wordnets that fall in this category.

3.1 Onto.PT

The Onto.PT (Gonçalo Oliveira and Gomes, 2014) project begun in 2008. To create a new wordnet in a completely automatic fashion, Onto.PT used several lexical resources available at the time, with special focus on those of the project PAPEL (Gonçalo Oliveira et al., 2008), including grammars to extract relations from dictionaries. Other exploited resources include Wiktionary.PT, Dicionário Aberto (Simões et al., 2012), TeP (Maziero et al., 2008), OpenThesaurus.PT and, more recently, OpenWN-PT (de Paiva et al., 2012).

The creation of Onto.PT follows the ECO approach (Gonçalo Oliveira and Gomes, 2014), tailored to for this project, but flexible enough to integrate words and relations extracted from different sources. ECO is different from other approaches because it tries to learn the whole structure of a wordnet, including the contents and boundaries of synsets, as well as the synsets involved in semantic relations. Hence, despite exploring, automatically, handcrafted resources, the authors refer to ECO as a "fully automatic" approach. It consists of three main phases: (i) relation extraction between words; (ii) synset discovery from the clusters of the extracted synonymy network (an initial set of synsets, such as those of TeP, may be used as a starting point); (iii) mapping word arguments of remaining relations to the discovered synsets. In

Onto.PT 0.6 (Gonçalo Oliveira et al., 2014), dictionary definitions were also assigned to synsets, automatically.

Onto.PT is different from the typical wordnet, not only for its creation process, but also because it includes a wide range of semantic relations that are not in PWN. Those relations are the same as the ones in PAPEL, extracted from dictionaries, and include causation, purpose, location or manner, among others.

On the one hand, ECO allows for the creation of a large knowledge base with little effort – Onto.PT 0.6 covers $\approx$169,000 distinct lexical items, organized in $\approx$117,000 synsets, which in turn are related through $\approx$174,000 relation instances. On the other hand, there are reliability consequences. For example, in Onto.PT 0.35, 74% of synsets were correct, in 18% there was no agreement between two judges, and the remaining had at least one incorrect word. The quality of relationships also varies dramatically depending on the type. Considering that relations between incorrect synsets are also wrong, the hypernymy connections were just 65% correct and between 78%-82% in a set with other relation types. These evaluation efforts are described in (Gonçalo Oliveira and Gomes, 2014). Nevertheless, Onto.PT was used, for instance, in the expansion of synonyms for information retrieval (Rodrigues et al., 2012) or for creating lists of causal verbs (Drury et al., 2014).

Due to its design, Onto.PT is a dynamic resource and, from release to release, may have significant changes in the number and size of its synsets. Thus, it is not planned to be aligned with PWN. Onto.PT is freely available in RDF/OWL[2], following an existing PWN model (van Assem et al., 2006), expanded to cover all its relation types.

3.2 OpenWordNet-PT

OpenWordNet-PT (de Paiva et al., 2012) abbreviated to OpenWN-PT, is a wordnet originally developed as a syntactic projection of the Universal WordNet (UNW). Its long-term goal is to serve as the main lexicon for a NLP system, focused on logical reasoning, based on representation of knowledge, using an ontology such as SUMO. The process of creating OpenWN-PT uses machine learning techniques to build relations between graphs representing lexical information from versions in multiple languages of Wikipedia entries

---

and open electronic dictionaries. OpenWN-PT has constantly been improved through linguistically motivated additions, either manually or from evidence in large corpora. This is also the case for the lexicon of nominalizations, NomLex-PT, tightly integrated with the OpenWN-PT (Freitas et al., 2014).

OpenWN-PT employs three language strategies in its lexical enrichment process: (i) translation; (ii) corpus extraction; (iii) dictionaries. Regarding translations, glossaries and lists produced for other languages, such as English, French and Spanish, are used, automatically translated and manually revised. The addition of data from corpora contributes with words or phrases in common use, which may be specific to Portuguese or do not appear in other wordnets. The first corpora experiment in OpenWN-PT was the integration of NomLex-PT. The use of a corpus, while useful for specific conceptualizations in the language, brings additional challenges for the mappings alignment, since it is expected that there will be expressions for which there is no synset in the English wordnet. As for the information in dictionaries, this was used indirectly through PAPEL (Gonçalo Oliveira et al., 2008).

Like Onto.PT, OpenWN-PT is available in RDF/OWL (Real et al., 2015), following and expanding, when necessary, the mapping proposed by (van Assem et al., 2006). Both the OpenWN-PT data and schema of the RDF model are freely available for download. The philosophy of OpenWN-PT is to keep a close connection with PWN, but try to fix the biggest mistakes created by the automated methods, through language skills and tools. A consequence of this close connection is the ability to minimize the impact of lexicographical decisions on splitting/grouping the senses in a synset. While such decisions are, to a great extent, arbitrary, the practical criterion of following the multilingual alignment behaves as a pragmatic and practical guiding solution.

OpenWN-PT was chosen by the developers of Freeling (Padró and Stanilovsky, 2012), OMWW (Bond and Foster, 2013), BabelNet and Google Translate, as the representative Portuguese wordnet in those projects, respectively, due to its comprehensive coverage of the language and its accuracy. OpenWN-PT currently has 43,925 synsets, of which 32,696 correspond to nouns, 4,675 to verbs, 5,575 to adjectives and 979 to ad-

verbs. Besides being available for download, the data can be retrieved via a SPARQL endpoint[3] and can be consulted and compared with other wordnets both through the OMWN interface and its own interface[4].

3.3 PULO

PULO (Simões and Guinovart, 2014), short for Portuguese Unified Lexical Ontology, intends to incorporate resources from open publicly available wordnets into a free Portuguese wordnet, perfectly aligned and included in the MCR project (Gonzalez-Agirre et al., 2012), which already includes wordnets for Spanish, Catalan, Basque and Galician, in addition to PWN.

The beginning of this project, in late 2014, involved some experiments on the translation and alignment between the English, Spanish and Galician wordnets. Beyond those, this process used probabilistic translation dictionaries (Simões and Almeida, 2003), a dynamic Portuguese-Galician translation dictionary (Guinovart and Simões, 2013), and the official Orthographical Vocabulary of the Portuguese Language. This resulted in ≈50,000 word meanings, but only ≈17,000 were actually added to PULO. This was due to the statistical nature of the approach and the cutoff line established. The scoring value obtained for each meaning was properly stored on the database and may serve as a measure of relevance or quality of each meaning.

Currently, as the other wordnets of MCR, the ontological structure of PULO is the same as PWN. Despite this similarity, the internal structure of the database allows each individual wordnet to be easily extended to new concepts. PULO is available for download and has currently 25,711 senses, corresponding to 17,854 synsets. In a second stage of the process, a machine translation of glosses was produced using the MyMemory API[5]. Through the same interface, it is possible to consult the other languages of the MCR, as well as to browse through the base ontology.

3.4 Ufes WordNet

The Ufes WordNet (Gomes et al., 2013) (UfesWN.BR) aims at building a Brazilian Portuguese database with a similar structure to PWN,

based on automatic translation. For this, a tool based on the Google Translate API was developed to translate the contents of PWN. UfesWN.BR covers 34,979 words, grouped in 48,981 synsets, connected by 238,413 relations. However, only 31,6% of the English synsets were translated and these translations are not very reliable. In the scope of this project, the glosses of PWN were also translated. They could be useful for other projects, depending on the quality and easiness of alignment, which has not been investigated.

4 Comparing Open WordNets

Table 1 summarises the main properties of the Portuguese wordnets. The most common alternative to the creation of a wordnet for Portuguese is based on translation, manual (MWN.PT), automatic (UfesWN.BR), based on a syntactic projection (OpenWN-PT), or on triangulation between resources (PULO). Within these four approaches, PULO stands out for using as a "pivot", not only the English wordnet, but also the wordnets for Spanish and Galician. Unlike all others, the structure of Onto.PT is learned fully automatically, based on the extraction of relationships from other textual resources or wordnets, and discovering clusters of synonyms, used as synsets. Among the advantages of a completely manual approach is the creation of a resource with an accuracy of virtually 100%. On the other hand, with an automatic approach, a larger resource can be created in a shorter time, avoiding tedious and time-consuming work, prone to raise issues. A semi-automatic method where expediency can be reigned in by accuracy would seem the best approach.

Name	Creation		Update	Usage
	Synsets	**Relations**		
WN.PT	manual	manual	manual	closed
WN.BR	manual	transitivity	manual?	free synsets
MWN.PT	manual? translation	transitivity	?	paid license
Onto.PT	RE,*clustering*	RE,*clustering*	automatic	free
OpenWN-PT	UWN projection	transitivity	semi-autom	free
UfesWN.BR	machine translation	transitivity	?	free
PULO	triangulation	transitivity	semi-autom	free

Table 1: Properties of Portuguese wordnets. A '?' is shown for fields we could not fill.

We also made a superficial comparison of their lastest versions, that should not be seen as more than a purely quantitative tabling. We do not consider the consistency nor the usefulness of the con-

tents of the various Portuguese wordnets.

On the number of covered lexical items, Onto.PT stands out for including more than three times more lexical items than the second largest wordnet, OpenWN-PT. This confirms that a fully automatic construction approach leads to a larger resource. Equally important for the size of Onto.PT, is the amount (currently six) and the type of resources used, including: resources that cover different variants of Portuguese, which can lead to minor spelling variations; and dictionaries, which already have a wide coverage of the language. Either manually or automatically, it is common to exploit dictionaries in the construction of a wordnet. Still, their automatic exploitation results in many different words and meanings that exist and are valid, but a large slice are of little use in colloquial Portuguese.

On the number of word senses, synsets and relation instances, Onto.PT also stands out from the rest. But it should be noted that there is an intrinsic trade-off between the size of a wordnet and the accuracy and usefulness of the resource under scrutiny. One of the difficulties in developing a wordnet is precisely to decide, on the one hand, if two words are to be regarded as synonymous and thus placed within the same synset and, on the other hand, which words should be in different synsets. These are typical lexicography challenges to which there is probably no final unique answer. But there seems to be a consensus that a very large number of synsets is a sign of "noise" in the process of grouping words and/or in the discrimination process. Correction/accuracy is undoubtedly one of the bottlenecks of building wordnets. If, on the one hand, size and coverage are a quantitative comparison, which is relatively simple, the same cannot be said about the quality assessment. PWN, built manually, may even reflect questionable decisions, but does not contain "errors" as such, as we are using it as a baseline for comparison. As for the wordnets built automatically, or semi-automatically, for languages other than English, quality assessment will always be an issue, since there is no golden reference available – this is precisely what they want to become. From this perspective, resources that rely on human labor have an advantage, although we do not know exactly how this advantage can and should be measured. An alignment with PWN may be important for obtaining additional knowledge, mostly from other resources aligned with it In addition to relation inheritance, an alignment allows access to knowledge of other extensions of PWN, such as WordNet-domains, SentiWordNet or TempoWordNet. On the other hand, a blind alignment does not consider that different languages represent different socio-cultural realities, do not cover the same part of the lexicon and, even where they seem to be common, several concepts are lexicalized differently (Hirst, 2004).

Both WN.PT and Onto.PT cover a wide range of relation types, some not typically present in wordnets. We recall that, for Onto.PT, their extracted was possible due to the regularities in dictionary definitions.

5 Building on Open WordNets

We presented and compared various wordnets that currently exist for Portuguese. Among them, four are freely available; until recently, one synset base (TeP) was also freely available; one (MWN.PT) may be purchased; and another can be explored online (WN.PT). The creation of these wordnets followed different approaches, from completely manual labour, through translation-based approaches with more or less manual labour, to an approach in which the whole structure is populated automatically. We hope to have shown that, currently, it makes no sense to regret that there is no Portuguese wordnet. In fact, the use of a wordnet in a project targeting Portuguese is becoming less of a problem of finding a workaround solution, and increasingly more one of choosing the most suitable within the available alternatives. This selection should consider, among other things, the need to align with other wordnets, the error tolerance, the coverage needs – both with regards to the lexical items and to relationships between them – and even the available budget. Since each wordnet has distinct characteristics, one should not discard the use of more than one wordnet in the same project.

It is sensible to ask whether all these alternatives make sense or if it would be preferable to focus on a single effort to build a single Portuguese wordnet, trying to harness the strong points of each of the projects described. The authors of this article, responsible for Onto.PT, OpenWN-PT and PULO, believe that there are advantages both on converging into a single wordnet and on keeping separate projects. Thus, in the short term, the development of each wordnet will remain a respons-

ability of its original team, but there will be closer monitoring of each other's work. The idea is that each project may reuse what is done by the others, this way minimizing duplicate work, but without losing sight of specific goals.

In a near future, Onto.PT will become a fuzzy wordnet, based on the redundancy accross several Portuguese computational lexical resources, including the other open wordnets, whose further updates will be welcome by this new initiative. Following ECO, confidence degrees will be assigned to each decision taken, including the membership of words to synsets (first experiments in Santos and Gonçalo Oliveira (2015)), or the attachment of relations to synsets. This will enable the users to select between a larger but less reliable resource and a smaller one with fewer issues.

Similarly to Onto.PT, the other open wordnets will devise the integration of the contents of each other, or replicate their enrichment approaches.

6 Conclusions

We presented a collection of Portuguese wordnets. While none feels as mature as Princeton WordNet, some have already been used in applications. Joint efforts, as we started doing and hope to do more, seem the only way of making progress in this hard problem. Clearly, the envisaged applications will lead to slightly different strong points in our resources, but the common denominator remains to provide a wordnet that is open, large coverage and as reliable as possible.

References

Raquel Amaro. 2014. Extracting semantic relations from portuguese corpora using lexical-syntactic patterns. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, LREC'14, Reykjavik, Iceland, May. ELRA.

Anabela Barreiro. 2010. Port4NooJ: an open source, ontology-driven portuguese linguistic system with applications in machine translation. In *Proceedings of the 2008 International NooJ Conference (NooJ'08)*, Budapest, Hungary. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *Proceedings of 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria, August. ACL Press.

Gerard de Melo and Gerhard Weikum. 2009. Towards a universal wordnet by learning from combined evidence. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM 2009)*, pages 513–522, New York, NY, USA. ACM.

Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. OpenWordNet-PT: An Open Brazilian WordNet for Reasoning. In *Proceedings of 24th International Conference on Computational Linguistics*, COLING (Demo Paper).

Bento C. Dias-da-Silva, Mirna F. de Oliveira, and Helio R. de Moraes. 2002. Groundwork for the Development of the Brazilian Portuguese Wordnet. In *Advances in Natural Language Processing (PorTAL 2002)*, LNAI, pages 189–196, Faro, Portugal. Springer.

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

Brett Drury, Paula C.F. Cardoso, Janie M. Thomas, and Alneu de Andrade Lopes. 2014. Lexical resources for the identification of causative relations in portuguese texts. In *Proceedings of the 1st Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish (ToRPorEsp)*, ToRPorEsp, pages 56–63, São Carlos, SP, Brasil, October. BDBComp.

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

Cláudia Freitas, Valeria de Paiva, Alexandre Rademaker, Gerard de Melo, Livy Real, and Anne de Araujo Correia da Silva. 2014. Extending a lexicon of Portuguese nominalizations with data from corpora. In *Proceedings of Computational Processing of the Portuguese Language - 11th International Conference (PROPOR 2014)*, São Carlos, Brazil, oct. Springer.

Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. 2010. Interfacing WordNet with DOLCE: towards OntoWordNet. In *Ontology and the Lexicon: A Natural Language Processing Perspective*, Studies in Natural Language Processing, chapter 3, pages 36–52. Cambridge University Press.

Marcelo Machado Gomes, Walber Beltrame, and Davidson Cury. 2013. Automatic construction of brazilian portuguese WordNet. In *Proceedings of X National Meeting on Artificial and Computational Intelligence*, ENIAC 2013.

Hugo Gonçalo Oliveira and Paulo Gomes. 2014. ECO and Onto.PT: A flexible approach for creating a Portuguese wordnet automatically. *Language Resources and Evaluation*, 48(2):373–393.

Hugo Gonçalo Oliveira, Diana Santos, Paulo Gomes, and Nuno Seco. 2008. PAPEL: A dictionary-based lexical ontology for Portuguese. In *Proceedings of Computational Processing of the Portuguese Language - 8th International Conference (PROPOR 2008)*, volume 5190 of *LNCS/LNAI*, pages 31–40, Aveiro, Portugal, September. Springer.

Hugo Gonçalo Oliveira, Inês Coelho, and Paulo Gomes. 2014. Exploiting Portuguese lexical knowledge bases for answering open domain cloze questions automatically. In *Proceedings of the 9th Language Resources and Evaluation Conference*, LREC 2014, Reykjavik, Iceland, May. ELRA.

Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012. Multilingual central repository version 3.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 2525–2529. ELRA, LREC'12.

Hugo Gonçalo Oliveira, Valeria de Paiva, Cláudia Freitas, Alexandre Rademaker, Livy Real, and Alberto Simões. 2015. As wordnets do português. In Alberto Simões, Anabela Barreiro, Diana Santos, Rui Sousa-Silva, and Stella E. O. Tagnin, editors, *Linguística, Informática e Tradução: Mundos que se Cruzam*, volume 7(1) of *OSLa: Oslo Studies in Language*, pages 397–424. University of Oslo.

Xavier Gómez Guinovart and Alberto Simões. 2013. Retreading Dictionaries for the 21st Century. In José Paulo Leal, Ricardo Rocha, and Alberto Simões, editors, *2nd Symposium on Languages, Applications and Technologies*, volume 29 of *OpenAccess Series in Informatics (OASIcs)*, pages 115–126. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 2012. UBY - a large-scale unified lexical-semantic resource. In *Proceedings of 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL 2012, pages 580–590, Avignon, France. ACL Press.

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

Palmira Marrafa, Raquel Amaro, and Sara Mendes. 2011. WordNet.PT Global – extending WordNet.PT to Portuguese varieties. In *Proceedings of 1st Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 70–74, Edinburgh, Scotland. ACL Press.

Palmira Marrafa. 2001. *WordNet do Português: uma base de dados de conhecimento linguístico*. Instituto Camões.

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 Linguagem Humana*, TIL, pages 390–392.

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.

Lluís Padró and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.

Adam Pease and Christiane Fellbaum. 2010. Formal ontology as interlingua: the SUMO and WordNet linking project and global WordNet linking project. In *Ontology and the Lexicon: A Natural Language Processing Perspective*, Studies in Natural Language Processing, chapter 2, pages 25–35. Cambridge University Press.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of 1st International Conference on Global WordNet*, GWC 2002.

Livy Real, Fabricio Chalub, Valeria de Paiva, Claudia Freitas, and Alexandre Rademaker. 2015. Seeing is correcting: curating lexical resources using social interfaces. In *Proceedings of 53rd Annual Meeting of the ACL and 7th International Joint Conference on NLP of Asian Federation of NLP - 4th Workshop on Linked Data in Linguistics: Resources and Applications*, Beijing, China, jul.

Ricardo Rodrigues, Hugo Gonçalo Oliveira, and Paulo Gomes. 2012. Uma abordagem ao Págico baseada no processamento e análise de sintagmas dos tópicos. *Linguamática*, 4(1):31–39, April.

Fábio Santos and Hugo Gonçalo Oliveira. 2015. Descoberta de synsets difusos com base na redundância em vários dicionários. *Linguamática*, page accepted for publication, December.

Diana Santos, Anabela Barreiro, Cláudia Freitas, Hugo Gonçalo Oliveira, José Carlos Medeiros, Luís Costa, Paulo Gomes, and Rosário Silva. 2010. Relações semânticas em português: comparando o TeP, o MWN.PT, o Port4NooJ e o PAPEL. In *Textos seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*, APL 2009, pages 681–700. APL.

Alberto Simões and Xavier Gómez Guinovart. 2014. Bootstrapping a Portuguese wordnet from Galician, Spanish and English wordnets. In *Advances in Speech and Language Technologies for Iberian Languages, Proceedings of 2nd International Conference, IberSPEECH 2014, Las Palmas de Gran Canaria, Spain*, volume 8854 of *LNCS*, pages 239–248. Springer.

Alberto Simões, Álvaro Iriarte Sanromán, and José João 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*, pages 121–127. Springer, April.

Alberto M. Simões and J. João Almeida. 2003. NATools – a statistical word aligner workbench. *Procesamiento del Lenguaje Natural*, 31:217–224, September.

Sofia Stamou, Kemal Oflazer, Karel Pala, Dimitris Christoudoulakis, Dan Cristea, Dan Tufis, Svetla Koeva, George Totkov, Dominique Dutoit, and Maria Grigoriadou. 2002. BalkaNet: A multilingual semantic network for the balkan languages. In *Proceedings of 1st Global WordNet Conference*, GWC'02.

Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO: a core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web*, WWW 2007, pages 697–706, Alberta, Canada. ACM Press.

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

Piek Vossen. 1997. EuroWordNet: a multilingual database for information retrieval. In *Proceedings of DELOS workshop on Cross-Language Information Retrieval*, Zurich.