

# OpenWordNet-PT: A Project Report

**Alexandre Rademaker**

IBM Research Brazil / FGV/EMAp  
Rio de Janeiro, Brazil  
alexrad@br.ibm.com

**Valeria de Paiva**

Nunance Communications  
Sunnyvale, CA, USA  
valeria.depaiva@gmail.com

**Gerard de Melo**

Tsinghua University  
Beijing, China  
gdm@demelo.org

**Livy Maria Real Coelho**

Univ. Federal do Paraná  
Curitiba, Brazil  
livyreal@gmail.com

**Maira Gatti**

IBM Research Brazil  
Rio de Janeiro, Brazil  
mairacg@br.ibm.com

## Abstract

This paper presents OpenWordNet-PT, a freely available open-source wordnet for Portuguese, with its latest developments and practical uses. We provide a detailed description of the RDF representation developed for OpenWordnet-PT. We highlight our efforts to extend the coverage of our resource and add nominalization relations connecting nouns and verbs. Finally, we present several real-world applications where OpenWordnet-PT was put to use, including a large-scale high-throughput sentiment analysis system.

## 1 Introduction

Semantic relationships between words are crucial in many forms of natural language processing. Computational systems are not aware of the fact that *carro* and *automóvel* both refer to cars, or that *caminhão* (truck) is related to these words as well in that they all share a common more general hyponym.

OpenWordnet-PT (or OpenWN-PT for short) is a lexical-semantic resource describing (Brazilian) Portuguese words and their relationships. It is modelled after and fully interoperable with the original Princeton WordNet for English (Fellbaum, 1998), relying on the same identifiers as WordNet 3.0. This means that one can easily find Portuguese equivalents for specific English word senses and vice versa. This also means that OpenWN-PT is part of a large ecosystem of compatible resources, including domain identifiers (Magnini and Cavaglia, 2000) and mappings to Wikipedia (de Melo and Weikum, 2010).

In this paper, we specify the RDF-based representation chosen for OpenWN-PT (Section 2) and describe our recent efforts to extend this resource (Sections 3 to 5), most notably with nominalization relations connecting nouns and verbs (Section 4). We also highlight several important applications of OpenWN-PT (Section 6).

## 2 RDF Representation

Wordnets have been distributed in a wide range of different incompatible data formats. An increasingly popular way of addressing the issue of interoperability is to rely on Linked Data and Semantic Web standards such as RDF (Cyganiak and Wood, 2003) and OWL (Hitzler et al., 2012), which have led to the emergence of a number of Linked Data projects for lexical resources (de Melo and Weikum, 2008; Chiarcos et al., 2012).

We believe that OpenWN-PT should best be encoded and distributed in RDF/OWL. Not only do these standards allow us to publish both the data model and the actual data in the same format. They also provide for instant compatibility with a vast range of existing data processing tools, including databases (so-called “triple stores”) providing SQL-like query interfaces based on the SPARQL standard (Harris and Seaborne, 2013).

Some years ago, a task force of the Semantic Web Best Practices Working Group proposed a standard encoding of WordNet in RDF (van Assem et al., 2006). This effort made WordNet directly accessible to Semantic Web applications. The proposed conversion aimed to be as complete as possible. The suggested representation also stayed as close to the original source as possible, that is, it reflects the original WordNet data model

without interpretation. Comparing with previous RDF translations of WordNet, the main features of this version are: (1) It does not model the hyponym hierarchy as a subclass hierarchy. (2) It represents words and word senses as separate entities with their own URI which makes it possible to refer to them directly. (3) It contains all relations that are present in Princeton WordNet. (4) It provides OWL semantics in the form of inverse properties, definition of property characteristics and property restrictions on classes that can be used by both the RDFS and OWL infrastructures.

The schema of the conversion has three main classes: `Synset`, `Word` and `WordSense`. There are three kinds of properties in the schema. A first set of properties connects instances of the main classes together. The class `Synset` is linked to its `WordSenses` with the property `containsWordSense`, and `WordSense` to its `Word` with the property `word`. A second set of properties represents the WordNet relations such as hyponymy and meronymy, including those that relate two `Synsets` to each other (e.g. `hyponymOf`), those that relate two `WordSenses` to each other (e.g. `antonymOf`), and a miscellaneous set containing `gloss` and `frame`. Finally, a third set of properties provides additional information about entities using literals. Examples are `synsetId`, which records the original ID given in Princeton WordNet to a `synset`, and the `tagCount` of a `WordSense`. The actual lexical form of a `Word` is recorded with the property `lexicalForm`. Each `synset` has an `rdfs:label` that is filled with the lexical form of the first word sense in the `synset`.

OpenWN-PT is completely aligned to Princeton WordNet. This means that each OpenWN-PT `synset` is a translation of an original Princeton WordNet `synset`, with no additional `synsets` or relations so far. Given this direct relation, we decided that our RDF representation does not require a full redundant modeling of all relations and information in Princeton WordNet. Instead, we chose to model our RDF as an add-on to WordNet 3.0 that extends it with information about the Portuguese language. For this, we simply add new `Synset` and `WordSense` instances that are linked to the English WordNet.

OpenWN-PT's RDF will thus only be useful together with an RDF version of Princeton WordNet.

While there is a previous RDF version of WordNet 3.0 online,<sup>1</sup> we wanted to ensure that all information in the WordNet 3.0 distribution was transformed to RDF. To this end, we wrote our own Common Lisp code to translate the WordNet 3.0 data files to RDF, following the W3C model (van Assem et al., 2006) with a few modifications as follows.

1. We add two more classes named `BaseConcept` and `CoreConcept` to identify the `synsets` that are base concepts (Vossen, 2002) or core concepts (Boyd-Graber et al., 2006), respectively.
2. We have added properties to capture information from WordNet 3.0 not available in the Prolog distribution nor in the “database files only” distribution. To this end, we have parsed and read the files `sents.vrb`, `sentidx.vrb` and `lexnames`. A `WordSense` can have a `lexFile`, `lexId`, `senseKey`, and an example sentence (for a `WordSense` of a `VerbSynset`). A `synset` can have a `lexicographerFile` and a `frame` (in the case of a `VerbSynset`).
3. We omitted redundant subclasses of `WordSense` like `NounWordSense`, as the part-of-speech can be derived from the corresponding `synset`. A subclass of `Word` called `Collocation` is also omitted, as the lexical form of `Word` instances can easily be examined to check for collocations.
4. We have adopted a different schema for naming the resources identifiers (URIs).

In Figure 1, we show the `synset` 00001740-n encoded in RDF in its more readable N3 notation variant (Berners-Lee and Connolly, 2011). `Word` instances are blank resources, that is, resources without a URI or unnamed resources. In Figure 2, we present the same `synset` in a graphical way, additionally showing its connection with the corresponding `synset` in the Princeton WordNet, including relevant semantic relations. Our code for this RDF version of WordNet 3.0 is freely available.<sup>2</sup>

<sup>1</sup>See <http://bit.ly/1cVExvj>.

<sup>2</sup>See <http://bit.ly/1ctbGSL>. The code requires AllegroGraph and Allegro Common Lisp. Both are commercial tools but free editions can be obtained on the Franz Inc. website at <http://www.franz.com>.

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix wn30: <http://arademaker.github.com/wn30/schema/> .
@prefix wn30en: <http://arademaker.github.com/wn30-en/instances/> .
@prefix wn30br: <http://arademaker.github.com/wn30-br/instances/> .

wn30br:synset-00001740-n wn30:synsetId "00001740" ;
  rdf:type wn30:NounSynset ; rdf:type wn30:BaseConcept ;
  owl:sameAs wn30en:synset-00001740-n ;
  wn30:containsWordSense wn30br:wordsense-00001740-n-1 ;
  wn30:containsWordSense wn30br:wordsense-00001740-n-3 ;
  wn30:gloss "o que é percebido, conhecido ou inferido como
  tendo existência própria (vivente ou não vivente)" .

wn30br:wordsense-00001740-n-1 wn30:wordNumber "1" ;
  rdfs:label "ser" ;
  rdf:type wn30:WordSense ;
  wn30:word _:anon642 .

wn30br:wordsense-00001740-n-3 rdfs:label "entidade" ;
  wn30:wordNumber "3" ;
  rdf:type wn30:WordSense ;
  wn30:word _:anon24777 .

_:anon24777 wn30:lexicalForm "entidade" ; rdf:type wn30:Word .
_:anon642 wn30:lexicalForm "ser" ; rdf:type wn30:Word .

```

Figure 1: The synset 00001740-n in N3 notation

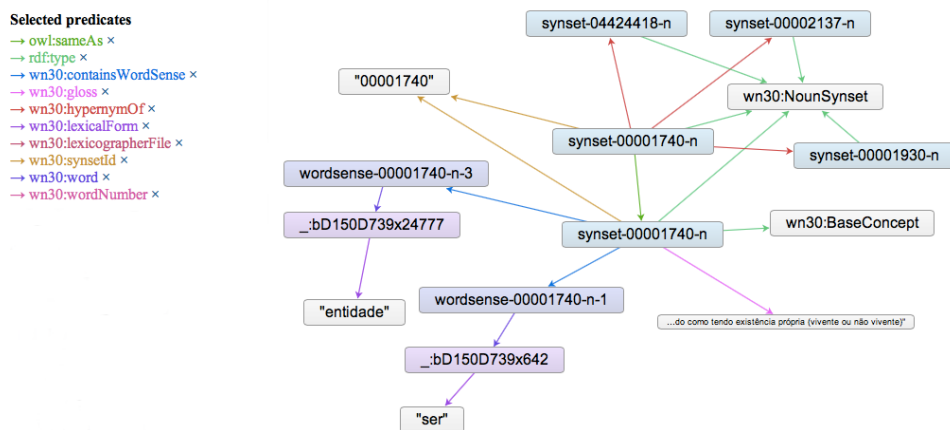


Figure 2: Synset 00001740-n and its neighbors in Princeton WordNet and OpenWordNet-PT

### 3 Extending the Coverage

The first version of OpenWN-PT was created using a semi-automated process drawing on UWN (de Melo and Weikum, 2009) and on manual revisions and gloss translations (Rademaker et al., 2012). Table 1 summarizes how OpenWN-PT has increased over the last two years. The number of synsets should be understood as the number of synsets with at least one Portuguese word. The sources of the new data were (Bond and Foster, 2013) and some manual addition of entries while working on projects that make use of the resource. These use cases are described later in Section 6.

	2011	2013	increase
synsets	41,810	43,895	5%
words	52,220	54,125	3%
senses	68,285	74,054	8%

Table 1: OpenWN-PT’s coverage development

Among resources that we can use to expand OpenWN-PT, we are considering (Dias-Da-Silva and de Moraes, 2003) and (Gonçalo Oliveira, 2013). Both projects are also concerned with the construction of a WordNet-like lexical resource for Portuguese. The former is more limited, offering around 19,888 synsets without any links to the Princeton WordNet and no relations between synsets, other than synonymy. The latter has already incorporated OpenWN-PT and is also encoded in RDF following the same vocabulary of (van Assem et al., 2006). This means that it should be straightforward to obtain data from Onto.PT.

Besides the continuous work on increasing the number of translated synsets, we have also invested some time to expand the relations. All semantic relations in Princeton WordNet 3.0 are included in our RDF export. Figure 2 shows how one can navigate from a OpenWN-PT synset in the graph to the Princeton WordNet synset. Most semantic relations also apply to the Portuguese words. However, since the first version of OpenWN-PT came from the UWN, which does not have word sense-specific relations, we do not have any generic way to map the lexical relations (relations between word senses) from Princeton WordNet to specific words in OpenWN-PT.

Mainly because of the sentiment analysis project described later in Section 6, we focused in particular on antonymy relationships. Studying the plot in Figure 3, which shows the dis-

tribution of the number of senses per synset in both wordnets, it is clear that we could take advantage of the fact that the majority of synsets in both wordnets have only one sense to propagate the antonym pairs in Princeton WordNet to the senses in OpenWN-PT with also only one sense. We search for synsets  $A$  in Princeton WordNet with only one sense, where this specific sense is related to another sense that is also unique in its synset in Princeton WordNet, say  $B$ . We can propagate this antonymy relation to OpenWN-PT if synset  $A$  and  $B$  in OpenWN-PT also have only one sense each. Using this idea, we were able to add 707 antonymy relation instances to OpenWN-PT (only about 10% of the number of pairs in the antonym relation of Princeton WordNet 3.0). In the future, we plan to additionally use common prefixes like “des”, “in” to match senses.

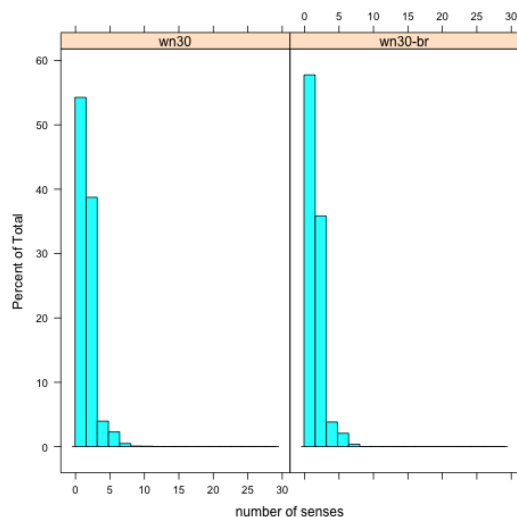


Figure 3: distribution of senses per synset

### 4 Nominalizations and NomLex-BR

Another extension of OpenWN-PT aims at incorporating links to connect deverbal nouns with their corresponding verbs. A sentence like “Alexander destroyed the city in 332 BC” can easily be parsed to obtain its semantic arguments, such as the agent (Alexander), the object destroyed (the city), and the time of the destruction (332 BC). In contrast, a sentence like “Alexander’s destruction of the city happened in 332 BC” is typically much harder to interpret correctly. The latter sentence describes the same event with the same semantic arguments, but these arguments are usually much harder to obtain automatically from a syntactic parser, given

that the event is described in terms of its nominalization *destruction* instead of its verbal form *destroy*. A proper handling of nominalizations (we are especially interested here in nominalizations of verbs, also called deverbal nouns) is important in numerous natural language understanding and inference tasks (Gurevich et al., 2008).

For English, NOMLEX (Macleod et al., 1998) has provided extensive descriptions of nominalizations. The original NOMLEX was constructed starting out with nominalizations with the suffixes *-ion*, *-ment* and *-er*, relying on frequent words in a corpus. NOMLEX sought not only to describe the possible complements for a nominalization, but also to relate the nominal complements to the arguments of the corresponding verb.

Our NomLex-BR project (Coelho et al., 2014) started with a manual translation of NOMLEX to Brazilian Portuguese, as NOMLEX is relatively small but still covers the most salient vocabulary. Many cases were very straightforward, due to the morphology of the words with similar nominalizer morphemes in both languages, e.g. pairs like *adjournment/adiamento*, *beneficiary/beneficiário*, *corrosion/corrosão*.

Overall, we have created over 1,000 entries. These have been integrated into OpenWN-PT, which we hope will facilitate their use for linguistic research of the traditional kind. For now, most of the words from NomLex-BR are linked to `Word` instances of OpenWN-PT. Eventually, we would like to have entries of NomLex-BR linked to specific `WordSense` instances of OpenWN-PT to the extent possible. We are currently also devising strategies to create entries and model phenomena specific to Portuguese.

Incorporating NomLex-BR data into OpenWN-PT has shown itself useful in pinpointing some issues with the coherence and richness of OpenWN-PT. In particular, it seems that 20% of words in NomLex-BR (which were manually chosen) are missing in OpenWN-PT. For instance, the word *abatement* corresponds in NOMLEX to the verb *abase*, and thus we would like a similar correspondence between the Portuguese noun *aviltamento* and the verb *aviltar* (our suggested translations). However, while *abatement* in English is present in two synsets with Portuguese equivalents, the synsets for the verb *abase* have a repetition in Portuguese. OpenWN-PT simply has two synsets *humilhar*, *abaixar* and *humilhar*, *rebaixar*. The

more common verb *humilhar* is repeated, while the uncommon *aviltar* was left out. Thus by verifying that verb-noun pairs in English are mapped to verb-noun pairs in Portuguese, we help ensure that the richness of synonyms in Portuguese is not lost in OpenWN-PT, which, being automatically derived from connectivity graphs, often gives preference to more commonly used words.

Other useful kinds of relationships between parts of speech (say the connections between adjectives and adverbs) are likely to also help to improve the accuracy and richness of our automatically derived resource. Altogether we reckon that by examining at random relationships that we know hold in the English WordNet in its translated Portuguese version, we should be able to both check the accuracy of OpenWordNet-PT and simultaneously investigate the parallelism between the two languages. From this perspective, one of the more interesting relationships, as far as knowledge representation is concerned, is the relationship of *entailment* between synsets. We have a goal of checking some 200 random English relationships in their translated forms as a way of measuring accuracy of the OpenWN-PT in the very immediate future.

## 5 Accuracy

Following the ideas of (Cruse, 1986), both (Vossen, 2002) and (Marrafa, 2002) used diagnostic templates of sentences to verify relations between synsets. We started a similar exercise. We choose six relations: `hypernymOf`, `memberHolonymOf`, `instanceOf`, `substanceHolonymOf`, `entails` and `causes`. For each of these relations, we randomly chose 30 pairs of synsets and then random words from each synset. Note that we had to keep drawing random synset relationships until both synsets included at least one Portuguese word. We ended up with 180 random sentences that we submitted to a linguist for manual verification (a single linguist to begin with). The linguist had to mark each sentence as being “correct”, “wrong” or “dubious”. As a result, we obtained 150 sentences marked as correct (83% of the sentences), 17 marked as wrong (one of the two words used to fill the template is probably placed in a wrong synset), and 13 marked as dubious (the linguist was not sure about the semantics of the sentence). In some

cases, the linguist was able to give detailed feedback like indicating misspelt words or providing a more specific reason for why the sentence was considered wrong. There were also trivial pairs in which the same word was chosen from both synsets. We hope to improve our tests in these cases.

Finally, some data mining could also help us to improve the accuracy of OpenWN-PT. For instance, synsets with an uncommonly high number of senses or words with an unexpected number of senses should be reviewed.

## 6 Usage Reports

### 6.1 Word Sense Disambiguation

OpenWN-PT has been incorporated into Freeling (Padró and Stanilovsky, 2012), a well-known suite of NLP tools. With OpenWN-PT's data and Freeling's word sense disambiguation framework, a given Portuguese text can automatically be annotated with word senses, and we can use these annotations in the projects below.

### 6.2 Sentiment Analysis

We have been investigating the OpenWN-PT usage in one of our projects at IBM Research-Brazil. In this project the main concern is to gather the sentiment of microblogging posts about football matches in Portuguese in *real-time*. The most famous microblogging online social network is Twitter<sup>3</sup>. As of 2013, there are more than 550 million active registered users and 58 million tweets are posted per day on average. These tweets are short messages that people send to provide updates on their activities, observations, or other interesting content, directly or indirectly to others. In sports, for instance, a lot of sentiment is expressed during a game match. Recently there have been several approaches that tackle the problem of classifying tweet sentiments using supervised or semi-supervised machine learning approaches (Celikyilmaz et al., 2010; Bakliwal et al., 2012) or lexicon-based methods, which are mostly unsupervised approaches (Li et al., 2011; Hogenboom et al., 2013).

As people react to events and generate a large Twitter stream of data, it is impossible to manually process and analyze all these data during the event's lifespan. There are several challenges related to analyzing all this data as quickly as possi-

ble. First, the system must be reliable: no information should be lost. This means that a highly available system is called for, with redundancy and active fault tolerance mechanisms. Second, it must have a high throughput, which leads us to an infrastructure that allows parallelism. Thirdly, sentiment classifiers should be able to work with limited resources in both time and space. The training phase should handle an unbalanced distribution of sentiments and in real time, it should be adaptive.

OpenWN-PT, Princeton WordNet, and SentiWordNet (Baccianella et al., 2010) were used with the goal of assessing a Machine Learning-based sentiment analysis component integrated into the IBM InfoSphere Streams (ISS) platform. ISS was used to address the problem of handling large streaming Twitter data with availability and scalability in real-time. One main advantage of using OpenWN-PT and SentiWordNet during the development of the Machine Learning-based classifier was that we could start experimenting without training data. The experiment was possible because OpenWN-PT synsets are linked to Princeton WordNet synsets which, in turn, have their sentiment scores in SentiWordNet. In order to train the classifiers for sentiment analysis, we have built a training corpus comprising data posted on Twitter during four friendly matches of the Brazilian team in 2013. About 1 million tweets have been gathered from these games. We built an online interface for a collaborative labeling of the tweets with respect to seven different classes: Certainly Negative (CN), Negative (N), Maybe Negative (MN), Neutral (N), Maybe Positive (MP), Positive (P), and Certainly Positive (CP). Here, we divided both negative and positive sentiment into three more specific classes in order to capture the degree of confidence for which the user is able to associate that tweet with one of these two main sentiment classes. Another class, Don't Know (D), represents tweets for which the sentiment could not be identified by the user. We used this annotated corpus to train a Naïve Bayes classifier. OpenWN-PT and SentiWordNet were used to check the consistency of the annotations and to provide insights during the entire course of the project. Unfortunately, given the real-time characteristic of project we were not able to run both classifiers on all collected data. As future work, we plan to use OpenWN-PT to expand the training corpus. For instance, from the manually annotated tweets we

<sup>3</sup>See <http://twitter.com>

can produce others tweets with synonym words which are likely to retain the same semantics and thus also the same sentiment with high probability.

### 6.3 Historical Biographic Dictionary

The second project we used OpenWN-PT for is related to the digitalization of a dictionary of historical biographies. The Getulio Vargas Foundation (FGV) maintains the Brazilian Dictionary of Historical Biographies (DHBB), a resource with 7,000 entries, the majority of which are biographical entries about politicians in Brazil’s recent history. The FGV would like to transform this static collection of entries into new methods of learning, teaching, acquiring, storing, and using information. Thus we decided that this knowledge would be more actionable if we could operate on it with semantic tools.

We used Freeling to automatically process this data by performing tokenization, sentence splitting, part-of-speech tagging, and finally word sense disambiguation with respect to OpenWN-PT. The DHBB corpus has a vocabulary size of 247,063 words. Table 2 shows the most frequent synsets in the corpus. The first line refers to the number of tokens without any associated synset. For instance, synset 00024720-n refers to “the way something is with respect to its main attributes” and synset 08050678-n is about “the organization that is the governing authority of a political unit” which, in OpenWN-PT, contains the word “governo”. Note that synset 10467395-n is about “the person who holds the office of head of state of the United States government”, for which OpenWN-PT contains the word “presidente” (president) as one of its words.

Var1	Freq
-	7343611
01835496-v	76589
02604760-v	47854
00047534-r	30351
00065639-v	28571
15202634-n	27294
08050678-n	26527
10467395-n	23238
00024720-n	23026
00118531-r	20434

Table 2: Frequent synsets in DHBB entries

Table 3 presents some of the most frequent to-

kens (after lemmatization) that could not be found in OpenWN-PT. The entries are not very surprising, as our efforts have not focused on domain-specific vocabulary from the political/historical domain.<sup>4</sup> The exercise of running FreeLing on the DHBB entries gave us good insights about the coverage of OpenWN-PT and we plan to extend the translations of synsets from the most frequent missing words in OpenWN-PT to the less frequent ones.

Token	Freq
[?:?/?/1/??:??.??:?]	15294
partido	15292
durante	10962
contra	10180
tornar	9991
militar	9906
outro	9739
segundo	9577
são_paulo	8803
estadual	8727
voto	8287
pmdb	6615
câm	6271
direito	6170
câmara_dos_deputados	5849

Table 3: Frequent tokens without synset

## 7 Conclusion

We have discussed the implementation and some applications of OpenWordNet-PT, an open WordNet for Brazilian Portuguese. Recent improvements include better coverage and nominalization links connecting nouns and verbs. The resource has been used in developing a high-throughput commercial system as well as in a cultural heritage project, and we anticipate that numerous further applications will follow. The data is freely available from <http://github.com/arademaker/wordnet-br/> and via a SPARQL endpoint<sup>5</sup>. We are also grateful to Francis Bond for providing an online interface via the Open Multilingual Wordnet website<sup>6</sup>.

<sup>4</sup>The first entry is the January encoded as a date template.

<sup>5</sup>See <http://logics.emap.fgv.br:10035>

<sup>6</sup>See <http://bit.ly/1aNOXxd>

## References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.
- Akshat Bakliwal, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, and Vasudeva Varma. 2012. Mining sentiments from tweets. In *Proc. of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 11–18.
- Tim Berners-Lee and Dan Connolly. 2011. Notation3 (N3): A readable RDF syntax. <http://bit.ly/1bYAs8y>.
- Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *Proc. of ACL 2013*, pages 1352–1362, Bulgaria. ACL.
- Jordan Boyd-Graber, Christiane Fellbaum, Daniel Oserson, and Robert Schapire. 2006. Adding dense, weighted connections to wordnet. In *Proc. of the 3th Global WordNet Conf.*, pages 29–36, Jeju Island, Korea, January.
- Asli Celikyilmaz, Dilek Hakkani-Tur, and Junlan Feng. 2010. Probabilistic model-based sentiment analysis of twitter messages. In *IEEE Spoken Language Technology Workshop (SLT)*, pages 79–84.
- Christian Chiarcos, Sebastian Nordhoff, and Sebastian Hellmann. 2012. *Linked data in linguistics: Representing and connecting language data and language metadata*. Springer.
- Livy Maria Real Coelho, Alexandre Rademaker, Valeria De Paiva, and Gerard de Melo. 2014. Embedding NomLex-BR nominalizations into OpenWordnet-PT. In *Proc. of Global WordNet Conf. 2014*. to appear.
- Alan Cruse. 1986. *Lexical semantics*. Cambridge University Press.
- Richard Cyganiak and David Wood. 2003. RDF 1.1 concepts and abstract syntax. Technical Report Draft 23 July 2013, W3C.
- Gerard de Melo and Gerhard Weikum. 2008. Language as a foundation of the Semantic Web. In *Proc. of ISWC 2008*, volume 401.
- Gerard de Melo and Gerhard Weikum. 2009. Towards a universal wordnet by learning from combined evidence. In *Proc. of CIKM 2009*, pages 513–522, New York, USA. ACM.
- Gerard de Melo and Gerhard Weikum. 2010. MENTA: inducing multilingual taxonomies from Wikipedia. In *Proc. of CIKM 2010*, pages 1099–1108. ACM.
- Bento Carlos Dias-Da-Silva and Helio Roberto de Moraes. 2003. A construçao de um thesaurus eletrônico para o português do brasil. *ALFA: Revista de Linguística*, 47(2).
- Christiane Fellbaum. 1998. *WordNet: An electronic lexical database*. The MIT press.
- Hugo Gonçalo Oliveira. 2013. *Onto.PT: Towards the Automatic Construction of a Lexical Ontology for Portuguese*. Ph.D. thesis, University of Coimbra.
- Olga Gurevich, Richard Crouch, Tracy Holloway King, and Valeria de Paiva. 2008. Deverbal nouns in knowledge representation. *J. Log. and Comput.*, 18(3):385–404.
- Steve Harris and Andy Seaborne. 2013. SPARQL 1.1 query language. Technical Report W3C Recommendation 21 March 2013, W3C.
- Pascal Hitzler, Markus Krotzsch, Bijan Parsia, Peter F. Patel-Schneider, and Sebastian Rudolph. 2012. OWL 2 web ontology language primer. Technical Report W3C Rec 11 Dec 2012, W3C.
- Alexander Hogenboom, Daniella Bal, Flavius Frasin-car, Malissa Bal, Franciska de Jong, and Uzay Kaymak. 2013. Exploiting emoticons in sentiment analysis. In *Proc. of the ACM Symposium on Applied Computing*, pages 703–710, New York, NY, USA. ACM.
- Lin Li, Yunqing Xia, and Pengzhou Zhang. 2011. An unsupervised approach to sentiment word extraction in complex sentiment analysis. *International Journal of Knowledge and Language Processing*, 2(1).
- Catherine Macleod, Ralph Grishman, Adam Meyers, Leslie Barret, and Ruth Reeves. 1998. Nomlex: A lexicon of nominalizations. In *Proc. of Euralex 1998*, pages 187–193, Liege, Belgium.
- Bernardo Magnini and Gabriela Cavaglia. 2000. Integrating subject field codes into WordNet. In *Proc. of LREC*, pages 1413–1418.
- Palmira Marrafa. 2002. Portuguese wordnet: general architecture and internal semantic relations. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 18(SPE):131–146.
- Lluís Padró and Evgeny Stanilovsky. 2012. Freeing 3.0: Towards wider multilinguality. In *Proc. of the 8th Intern. Conf. on Language Resources and Evaluation (LREC'12)*, pages 23–25, Istanbul, Turkey, may.
- Alexandre Rademaker, Valeria De Paiva, and Gerard de Melo. 2012. Openwordnet-pt: An open brazilian wordnet for reasoning. In *Proc. of COLING 2012*, pages 353–360, Mumbai.
- Mark van Assem, Aldo Gangemi, and Guus Schreiber. 2006. RDF/OWL representation of WordNet. Technical Report W3C Working Draft 19 June 2006, W3C. <http://bit.ly/1jtGsA8>.
- Piek Vossen. 2002. EuroWordNet general document. Technical Report Version 3 Final July 1, University of Amsterdam.