

# Portuguese Manners of Speaking

**Valeria de Paiva**

University of Birmingham

valeria.depaiva@gmail.com

**Alexandre Rademaker**

IBM Research and FGV/EMAp

alexrad@br.ibm.com

## Abstract

Lexical resources need to be as complete as possible. Very little work seems to have been done on adverbs, the smallest part of speech class in Princeton WordNet counting the number of synsets. Amongst adverbs, manner adverbs ending in ‘-ly’ seem the easiest to work with, as their meaning is almost the same as the one of the associated adjective. This phenomenon seems to be parallel in English and Portuguese, where these manner adverbs finish in the suffix ‘-mente’. We use this correspondence to improve the coverage of adverbs in the lexical resource OpenWordNet-PT, a wordnet for Portuguese.

## 1 Introduction

Adverbs get a short shrift in Lexical Semantics. They have the smallest number of synsets in Princeton WordNet (PWN) (Fellbaum, 1998), perhaps rightly so, as most of their meaning, at least as far as *manner adverbs* are concerned, can be gleaned from the associated adjective. Manner adverbs tell us about the way something happens or is done. Many English adverbs are formed by adding the suffix -ly to an adjective (e.g. calm + ly = calmly). Many Portuguese adverbs are formed in a similar manner by adding the suffix ‘-mente’ to the singular feminine form of the adjective. For example, *sincera* (sincere, feminine) + ‘mente’ = *sinceramente* (sincerely). (Note the ‘a’ in *sinceramente*, marking the feminine version of the adjective.) This phenomenon of pairs adjective-adverb, where morphologically only the suffix ‘-ly’ is added to the adjective to form the adverb is very widespread both in English and in Portuguese. Many manner adverbs in Portuguese consist of an adjective plus the suffix ‘-mente’. These modify a predicate or phrase, specifying how the action unfolds. They can be paraphrased by “one way of Xing” or “a form of X”, where

X corresponds to the base adjective for the adverb’s formation. For example, the adverb *rapidamente/rapidly* lets you qualify that a given action was performed “in a rapid way”.

The generic pattern here seems to be that, since most of these adverbs are a kind of “manner”, and manner in Portuguese is a feminine word, the adverb tends to be the feminine form of the adjective, plus the suffix ‘-mente’. This can be used by NLP systems, such as FreeLing (Padro and Stanilovsky, 2012), to avoid keeping duplicate meanings that are almost the same. For example, for *stertorously*, we have the feminine adjective *estertorosa* plus the suffix *-mente*, *estertorosamente*. We take the view that all the reasonably used adjectives and adverbs should be in the lexicon and strive to complete our online, open source Portuguese wordnet (de Paiva et al., 2012). This goal of completing the a wordnet adverbial lexicon, using corpora, we share with (Côrtes et al., 2018). Our work can be seen as taken a subcase of their approach (only “manner” adverbs are considered here), but looking at varied-sized corpora in Portuguese.

Princeton WordNet v3.0 contains 3621 adverb synsets. Out of the total number of adverb synsets, some 2646 synsets have word forms terminating in ‘-ly’, which usually indicates a manner adverb, for example *attributively*: *in an attributive manner*. Also most of the manner adverbs are, in first approximation, monosemous, i.e. mostly they are in a single synset. Princeton WordNet statistics on polysemy can be found in <https://wordnet.princeton.edu/documentation/wNSTATS7wn>. Out of the 3621 adverb synsets, 2554 have no translation in Portuguese in our OpenwordNet-PT lexical resource (de Paiva et al., 2012), at the moment. Out of these 2554 synsets, 1863 have words finishing in ‘-ly’, so we presume that these are manner adverbs. These synsets follow the usual distribution,

mostly (1375) have a single word in English. Then 348 synsets have two words and 109 synsets have three words. Few synsets (31) have 4 or more words.

|                                 |      |
|---------------------------------|------|
| Total English adverb synsets    | 3621 |
| EN synsets no-translation       | 2554 |
| EN synsets with -ly             | 1863 |
| Single word EN-adverb-ly        | 1375 |
| Two words EN-adverb-ly          | 348  |
| Three words EN-adverb-ly        | 109  |
| Four or more words EN-adverb-ly | 31   |

Table 1: PWN Coverage of manner adverbs

The adverbs with only a word should be the ones that are easy to translate and, in first inspection, seem to behave similarly both in English and in Portuguese. Since the phenomena are so similar in English and Portuguese, an approach based in translations and checking should work well for our goal of completing the lexical resource in Portuguese. Our main goal here is to reduce the number of untranslated synsets as much as we can, as far as manner adverbs finishing in ‘-ly’ (English) and ‘-mente’ (Portuguese) are concerned.

## 2 PWN, OMW and OWN-PT

Our work is based on using Princeton WordNet (PWN) and the Open Multilingual WordNet (OMW) (Bond and Foster, 2013) to improve the OpenWordNet-PT (de Paiva et al., 2012). PWN is one of the most used lexical resources in Computational Linguistics. It is the most used resource for Word Sense Disambiguation, according to Wikipedia <https://en.wikipedia.org/wiki/WordNet>, but it has many other applications like Natural Language inference, text summarizing, question answering, etc.

The Open Multilingual WordNet (OMW) (Bond and Foster, 2013) provides access to open licensed wordnets in a variety of languages, all linked to the Princeton Wordnet of English (PWN). Their goal is to make it easy to use wordnets in multiple languages. Other wordnets with reasonably large number of adverbial synsets are associated to the OMW, but only the French wordnet WOLF (Wordnet Libre du Français) (Sagot et al., 2009) seems to have had a similar problem to the one we face, of extending the adverbial coverage of synsets from English. For French, as well as for Portuguese,

the synsets in previously existing wordnets were not available, because of license issues. However to complete WOLF, Segot et al had the lexicon Leff and the synonyms database DicoSyn to help them and we have not found similar resources for Portuguese.

OpenWordNet-PT (OWN-PT) is an ongoing project to build up a large wordnet for Portuguese. OWN-PT is still half the size of PWN, and we have done so far only partial evaluations of its coverage. For instance, punctual evaluations of verbs was done in (de Paiva et al., 2016, 2014a) and of (gentilic) adjectives in (Real et al., 2016). Several discussions on nominalizations and their arguments are presented in (Real et al., 2014; Freitas et al., 2014; de Paiva et al., 2014b). As suggested in (Real et al., 2015) the use of a visual interface for OWN-PT helps to discover some interesting subsets of synsets to work on, like the ones ending in “ly” that we investigate here. OWN-PT is the Portuguese wordnet in the OMW project, it has currently 47,932 synsets, 56,928 word forms and 81,374 senses.

## 3 Kinds of Adverbs

Translation seems to work fairly well for this tightly restricted subset of manner adverb synsets, but there are still some generic issues and problems.

WordNet and other lexical resources have to cope with affixes (prefixes and suffixes) and these seem to be used differently in different languages. For instance Portuguese has no suffix “less” as in *bloodlessly*, *mindlessly*, and *painlessly*. One needs to use a separate preposition *sem/without* to express the same meaning *sem sangue*, *sem pensar*, *sem dor*, respectively. Thus corresponding to the suffix “less” in English we need adverbial phrases instead of adverbs for adjectives and adverbs in Portuguese. PWN lists 76 synsets that are manner adverbs ending in “lessly”. Of course these can have single word translations in Portuguese, but these are not as direct as we would like them to be. For instance for *bloodlessly* (in the appropriate abstract synset 00418712-r) we can use the similar *pacificamente/pacifically* but the direct translation would be an adverbial locution *sem sangue/without blood*.

English also has several ways of negating an adjective. The prefix “un” appears in several of the manner adverbs, for example in *unofficially*, *unde-*

servedly, and *unmelodiously*. This prefix has no exact corresponding version in Portuguese. Thus while the positive versions e.g. *officially*, *deservedly*, and *melodiously* have exact correspondents in Portuguese (namely *oficialmente*, *merecidamente*, and *melodiosamente*), the negative versions do not exist as single words in Portuguese. Again we need either adverbial expressions (e.g. *não oficialmente* or less direct translations, like *dissonantemente* for *unmelodiously*). For a few words we have the similar prefix “i/in” in *imerecidamente*. Princeton WordNet has 202 synsets for manner adverbs of the form ‘un\*ly’, out of which, 153 have no Portuguese words at all, to begin with.

English has an extremely useful way of creating adjectives from nouns, by adding a “y” or “ly” to the noun. For example *sandy*, *hilly*, and *pearly* mean “like sand”, “with many hills” and “of the off-white color of pearls”. Some of these are considered informal language and do not appear in PWN, for example *yellowy*, which is in the Collins online dictionary but not in PWN. Some of these adjectives coming from nouns do not seem to have similars in Portuguese, for example *girly*, *womanly*, and *manly*. We can say *femininamente* and *masculinamente* in Portuguese and we can say *womanly* and *manly* in English, but we do not have the adverbs *femininely* and *masculinely* in English, neither do we have *meninamente*, *mulhermente*, or *homenmente* in Portuguese.

For some cases, where the adverb/adjective pair comes from a noun, e.g. *thirstily*, *thirsty*, and *thirst*, in Portuguese we can have the noun and the adjective, but the adverb is much less used. So *sede* (noun) and *sedento* (adjective) are common enough, but *sedentemente* (adverb) is less so. The meaning in this case is obvious and if someone uses the word, they will be understood, but it is not widespread.

Other suffixes that do not translate well are the very Anglo-Saxon ones like “ward” (in the direction of) as in e.g. *heavenward*.

Sometimes an almost exact translation exists, but the words are not from the same root. For example *regally*: *in a regal manner* corresponds to *regiamente*: *uma maneira de ser rei*.

We have to be careful with the so called *faux amis*. For instance the work *cruelly* in English has two fairly distinct meanings or synsets: 00232425-r *cruelly* (excessively; “a cruelly bitter winter”) and 00232499-r *cruelly* (with cruelty; “he treated

his students cruelly”) The first one is not what *cruelmente* means in Portuguese, it means something like *excessivamente/excessively* only, while the second is exactly the same meaning as the Portuguese word, some kind of evilness. To make it clear, we do say in Portuguese *um inverno cruel*, (a cruel winter), but the adverb is not used in this sense. All of these issues have to be manually checked and we are in the process of doing so.

## 4 Tools and Evaluation

Evaluation of semi-automatically constructed resources is a thorny subject. There are no canonical ways of evaluating these resources, the usual mechanical turk style evaluation seems both pointless and not objective enough. However, there are plenty of lists of most used words on the web, usually collected by teachers of English as a second language and these may help us guarantee that, at least the most used manner concepts are in place. We used a collection of lists to evaluate this work.

We checked lists of both English and Portuguese adverbs. The first list of English adverbs, comes from the English Club<sup>1</sup>, and consists of 130 popular manner adverbs in English, translated automatically and manually verified. We check how many of these 130 adverbs, are already present in OpenWordNet-PT. This is a qualitative measure, as the English adverbs (automatically translated) are sometimes polysemous and translations, even trying to keep roots, whenever possible, can lead to different results. With this list we were hoping to check our coverage for popular manner adverbs, as a list produced for learners of English as a second language tries to cover as many popular meanings as possible. The results were not very inspiring. Out of 130 English verbs translated, 5 were not of the form ending in -mente in Portuguese. Then of the 125 left, 45 were not present in OWN-PT, so we were missing 36% of this set. We had expected to have a number closer to the whole set already covered.

However, many of the adverbs coming out from this list are not very popular in Portuguese. So we decided to check the manner adverbs in the Parallel Universal Dependencies corpora<sup>2</sup>, since these exist both in Portuguese and English and

<sup>1</sup>list at <https://www.englishclub.com/vocabulary/adverbs-manner.htm>

<sup>2</sup>see [https://github.com/UniversalDependencies/UD\\_Portuguese-PUD](https://github.com/UniversalDependencies/UD_Portuguese-PUD)

they are already aligned. Parallel Universal Dependencies (PUD) treebanks were created for the CoNLL 2017 shared task on Multilingual Parsing from Raw Text to Universal Dependencies<sup>3</sup>. There are 1000 sentences in each language, always in the same order. The sentences are taken from the news domain and from Wikipedia. There are usually only a few sentences from each document, selected randomly, not necessarily adjacent, so sentences usually make sense by themselves.

There are 127 manner adverbs ending in *-ly* in the English Parallel Treebank PUD-EN. Perhaps surprisingly, PWN misses three of these adverbs *proactively*, *definitively*, *logistically*. These can arguably be considered newish adverbs, if one looks at the usage distribution provided by Google, e.g. Figure 1.

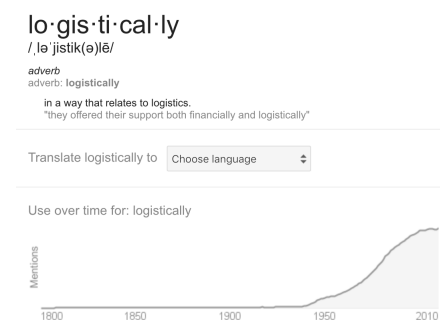


Figure 1: The Universe

We checked whether all 127 English manner adverbs had translations in Portuguese and found that only 12 synsets had not, more in line with what we were expecting. Two of these 12 were not in PWN to begin with, so we had to add word forms to our resource in Portuguese for ten synsets. These synsets correspond to adverbs *reportedly*, *unusually*, *thematically*, *technologically*, *posthumously*, *persistently*, *mindlessly*, *inimitably*, *culturally*, *catastrophically*. Taking advantage of the parallelism of the resources, we then looked at all the “-mente” adverbs in the Parallel Universal Dependencies corpus in Portuguese (PUD-PT). There we found 118 adverbs ending in *-mente*, of which only 10 were not described in the OpenwordNet-PT. (Of the three adverbs missing in PWN two would need to be added to Portuguese too (*logistically*, *proactively*, but *definitivamente*, corresponding to *definitively* was already in the base.) The Portuguese manner adverbs added

<sup>3</sup>check <http://universaldependencies.org/conll17/>

are *tecnologicamente*, *presumivelmente*, *postumamente*, *proativamente*, *incomumente*, *escrupulosamente*, *culturalmente*, *catastroficamente*, *arduaamente*, *tematicamente*

Encouraged by these results, which looked more the way we expected, we also looked at the list of the manner adverbs from the largest Universal Dependencies (UD) English corpus, UD-EWT<sup>4</sup>. First we looked at only the hundred most used adverbs in UD-EWT and realized that not many manner *ly*-adverbs were present. Then we looked at the whole list of *ly*-adverbs in UD-EWT, which counted 417 unique terms and checked how many we have already in Portuguese, in our resource. Here we come up against one of the problems of corpus work, in that there are 28 words that are typos, informal language (like “proolly” for ‘probably’) or misspellings. But 334 were correctly translated, so we needed to add some 60 synset translations.

Finally we turned the tables and looked at Portuguese “mente”-adverbs in our favorite application corpus, the Brazilian Historical Dictionary of Biographies (the acronym in Portuguese is DHBB)(Paiva et al., 2014). Again not so many *mente*-adverbs in the top 100 most used adverbs, only 20, out of which two we cannot find an easy PWN synset for. (PWN has *parallel*, but not the adverbial form, in Portuguese *paralelamente*. PWN has *provisionally*, which is similar to *interinamente*, but not exactly the same). Then we tried to get all the manner adverbs in this corpus of 323K sentences, but ran out of time to complete the checking. In summary, by looking at all these lists and also by simply checking the glosses for similar adverbs in Portuguese we managed to complete 151 synsets in OWN-PT. While this means that there are still 1727 empty manner adverbs synsets, our error reduction rate was respectable for a manually checked resource. Given the total number of adverb synsets in PWN empty before (2554), completing another 151 synsets gives us 6% more information than we had before.

## 5 Conclusion

We worked to semi-automatically complete the OpenWordNet-PT, as far as adverbs of manner, finishing in “ly” (in English) or “-mente” (in Portuguese) are concerned. These adverbs are called “deadjectival”, as they are usually a “way of X”

<sup>4</sup>see <http://universaldependencies.org/>

where X is an adjective. The phenomenon seems completely parallel in English and Portuguese and hence it seemed to us that it should be an easy task to accomplish.

The task proved somewhat more complicated than envisaged. Out of the 1863 synsets with no Portuguese words we have managed to manually check some 600 and this uncovered some similarities and issues that seem worth pointing out. We also realized that an open dictionary of deadjectival adverbs that we had intended to use as a superset of our synsets (Portal da Língua Portuguesa <http://www.portaldalinguaportuguesa.org/>) has too many entries that we do not feel comfortable listing, as they do not seem very much used in the Portuguese we speak.

On the positive side, the way we found of comparatively checking the Parallel Universal Dependencies corpus points out a new application of these parallel corpora, to help evaluate coverage of lexical resources. We plan to use it for checking adjectives and verbs in PWN and OWN-PT in the near future. Finally, we need to finish checking all the other thousand adverbial synsets in PWN, which are empty in Portuguese. We also plan to develop our own dictionary of deadjectival pairs that work both in English and Portuguese.

## References

- Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1352–1362.
- Priscila Côrtes, Mateus Riva, and Livy Real. 2018. Extending the adverbial coverage of OpenWordnet-PT. In *Proceedings of the PROPOR Conference (PROPOR2018)*, Canela, Brazil. Springer.
- Christiane Fellbaum. 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 *Computational Processing of the Portuguese Language, 11th International Conference, PROPOR 2014*, São Carlos, Brazil. Springer.
- Lluís Padro and Evgeny Stanilovsky. 2012. Freeling 3.0: Towards Wider Multilinguality. In *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey. ELRA.
- Valeria de Paiva, Fabricio Chalub, Livy Real, and Alexandre Rademaker. 2016. Making virtue of necessity: a verb lexicon. In *PROPOR – International Conference on the Computational Processing of Portuguese*, Tomar, Portugal.
- Valeria de Paiva, Cláudia Freitas, Livy Real, and Alexandre Rademaker. 2014a. [Improving the verb lexicon of openwordnet-pt](#). In *Proceedings of Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish (ToRPorEsp)*, São Carlos, Brazil. Biblioteca Digital Brasileira de Computação, UFMG, Brazil.
- Valeria de Paiva, Alexandre Rademaker, and Gerard de Melo. 2012. [OpenWordNet-PT: An open Brazilian Wordnet for reasoning](#). In *Proceedings of COLING 2012: Demonstration Papers*, pages 353–360, Mumbai, India. The COLING 2012 Organizing Committee. Published also as Techreport, <http://hdl.handle.net/10438/10274>.
- Valeria de Paiva, Livy Real, Alexandre Rademaker, and Gerard de Melo. 2014b. NomLex-PT: A lexicon of Portuguese Nominalizations. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Valeria De Paiva, Dário Oliveira, Suemi Higuchi, Alexandre Rademaker, and Gerard De Melo. 2014. [Exploratory information extraction from a historical dictionary](#). In *IEEE 10th International Conference on e-Science (e-Science)*, volume 2, pages 11–18. IEEE.
- Livy Real, Fabricio Chalub, Claudia Freitas, Alexandre Rademaker, and Valeria de Paiva. 2015. Seeing is correcting: curating lexical resources using social interfaces. In *Proceedings of the 4th Workshop on Linked Data in Linguistics: Resources and Applications*, pages 20–29.
- Livy Real, Valeria de Paiva, Fabricio Chalub, and Alexandre Rademaker. 2016. Gentle with gentils. In *Joint Second Workshop on Language and Ontologies (LangOnto2) and Terminology and Knowledge Structures (TermiKS) (co-located with LREC 2016)*, Slovenia.
- Livy Real, Alexandre Rademaker, Valeria de Paiva, and Gerard de Melo. 2014. [Embedding NomLex-BR nominalizations into OpenWordnet-PT](#). In *Proceedings of the 7th Global WordNet Conference*, pages 378–382, Tartu, Estonia.
- Benoît Sagot, Karën Fort, and Fabienne Venant. 2009. Extending the adverbial coverage of a french wordnet. In *NODALIDA 2009 workshop on WordNets and other Lexical Semantic Resources*, page 0.