On the Utility of Word Embeddings for Enriching OpenWordNet-PT

Hugo Gonçalo Oliveira
CISUC, Department of Informatics Engineering, University of Coimbra, Portugal

Fredson Silva de Souza Aguiar
FGV/EMAp, Rio de Janeiro, Brazil

Alexandre Rademaker
IBM Research, Rio de Janeiro, Brazil
FGV/EMAp, Rio de Janeiro, Brazil

Abstract

The maintenance of wordnets and lexical knwoledge bases typically relies on time-consuming manual effort. In order to minimise this issue, we propose the exploitation of models of distributional semantics, namely word embeddings learned from corpora, in the automatic identification of relation instances missing in a wordnet. Analogy-solving methods are first used for learning a set of relations from analogy tests focused on each relation. Despite their low accuracy, we noted that a portion of the top-given answers are good suggestions of relation instances that could be included in the wordnet. This procedure is applied to the enrichment of OpenWordNet-PT, a public Portuguese wordnet. Relations are learned from data acquired from this resource, and illustrative examples are provided. Results are promising for accelerating the identification of missing relation instances, as we estimate that about 17% of the potential suggestions are good, a proportion that almost doubles if some are automatically invalidated.

2012 ACM Subject Classification Computing methodologies → Lexical semantics; Computing methodologies → Language resources

Keywords and phrases word embeddings, lexical resources, wordnet, analogy tests

Digital Object Identifier 10.4230/OASIcs.LDK.2021.21

Funding *Hugo Gonçalo Oliveira*: Supported by national funds through FCT, within the scope of the project CISUC (UID/CEC/00326/2020) and by European Social Fund, through the Regional Operational Program Centro 2020.

Acknowledgements The research described in this paper was partially conducted in the scope of the COST Action CA18209 Nexus Linguarum (European network for Web-centred linguistic data science).

1 Introduction

When it comes to representing lexico-semantic knowledge, there are two main approaches: lexical knowledge bases, like wordnets [13], and distributional models, like word embeddings [24] learned from raw text. Wordnets are more formalised than distributional models, and typically rely on some manual effort, often by experts, e.g., for grouping synonymous words in so-called synsets and linking them according to a small set of semantic relations with lexicographic relevance, such as hypernymy and meronymy. On the other hand, distributional models are inspired by the distributional hypothesis [19] and capture the meaning of the words of a language by analysing their neighbourhoods in large collections of text.

Even though they are not formalised at all, word embeddings can be learned automatically and do not require expert knowledge. Moreover, from the regularities in natural language text, they may capture virtually any semantic relation between words, even if not all can be acquired with simple methods, such as the vector offset [24]. This suggests that word embeddings can be of great value for minimising some of the limitations of wordnets, namely their coverage of relation instances.

3rd Conference on Language, Data and Knowledge (LDK 2021).
Editors: Dagmar Gromann, Gilles Sérasset, Thierry Declerck, John P. McCrae, Jorge Gracia, Julia Bosque-Gil, Fernando Bobillo, and Barbara Heinisch; Article No. 21; pp. 21:1–21:13
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In this paper, we explore Portuguese word embeddings having in mind the enrichment of OpenWordnet-PT (OWN-PT) [8], a public domain Portuguese wordnet in the Open Multilingual WordNet (OMW) [1] project, aligned with Princeton WordNet (PWN) [13], and with a comprehensive coverage of the language. More precisely, we: (i) create several analogy tests with data extracted from OWN-PT, each for a different relation; (ii) apply two analogy-solving methods [11] to the previous test, though with poor performance; (iii) inspect the top answers given by one of the methods and conclude that some correspond to missing relation instances in OWN-PT, which can thus be used as suggestions for its enrichment.

The remaining of the paper is organised as follows: in Section 2, we overview work on the automatic acquisition of lexico-semantic relations from text, their usage for enriching wordnets, as well as some examples of how word embeddings can be exploited for this purpose, using analogy-solving methods; in Section 3, we give a general overview of OWN-PT; in Section 4, we describe the applied methods and how we created analogy tests with OWN-PT, further used for learning how relations are represented in word embeddings; in Section 5, we report on the accuracy of analogy-solving methods; in Section 6, following an inspection of the answers given by the previous methods, we discuss on the utility of such methods for enriching OWN-PT; in Section 7, we highlight the main conclusions of this work.

2 Background and Related Work

Earlier attempts for the automatic acquisition of lexico-semantic relations and their compilation in a lexical knowledge base exploited language dictionaries, their structure, and patterns used in the definitions [5]. Once Princeton WordNet (PWN) [13] became available for English, work on the creation of such a resource from scratch was no longer a priority.

Following the success of PWN, wordnets were developed for many other languages [2]. However, PWN is the product of intensive manual labour during many years. So, the creation of wordnets varied from project to project. Roughly, two approaches have been followed for creating wordnets [35]: the *expand approach* translates the synsets in PWN to a target language, takes over the relations from PWN, and revises them; the *merge approach* defines synsets and relations in a language and then aligns them with PWN, using equivalence relations. Instead of starting from scratch with the merge approach, the expand approach is the most commonly used among wordnets in the Open Multilingual WordNet initiative.[1]

But the truth is that, no matter the approach taken, fixes will always be required in a wordnet, and having an adequate coverage will always be an issue. Not to mention that language keeps evolving and maintenance is always necessary. Therefore, it is no surprise that different automatic procedures have been proposed for enriching wordnets, most of which by exploiting raw textual corpora. Such work ranges from handcrafting useful patterns for acquiring hypernymy-hyponymy relations [21], to learning similar patterns, not only for hypernymy [32], but also other relations [28], following weakly-supervised approaches that used examples from PWN as seeds.

In the last decade, more efficient distributional representations of words became available [24, 29], with promising results regarding lexical tasks, like computing word similarity and analogies. The former aims at computing the similarity between pairs of words, e.g. *dog* should be more similar to *cat* than to *car*. Performance is typically assessed with tests where similarity was manually assigned to pairs of words.

[1] http://compling.hss.ntu.edu.sg/omw/

Computing an analogy consists of answering the question *what is to b as a^* is to a?*, e.g. *what is to Portugal as Paris is to France?*. In this case, the relation between the computed word, b^*, and b, must be as close as possible to the relation between a^* and a. But the number of possible relations between two words is huge, especially if we consider morphological and semantic, and different relations will pose different challenges. Therefore, analogy tests, used for assessing this task, typically cover different relation types. For instance, the Google Analogy Test (GAT), notably used for assessing word2vec embeddings [24], covers nine types of syntactic and five types of semantic relation. The Bigger Analogy Test Set (BATS) [15] covers a total of 40 relation types, 10 for each of four categories: grammatical inflections, word-formation, lexicographic and world-knowledge relations.

The most common method for computing an analogy in the embedding space is to compute the vector offset, also known as the 3CosAdd method [24]. Yet, alternative methods were proposed for minimising limitations of the previous method. For instance, in addition to releasing BATS, its creators propose two methods that, instead of computing an analogy from a single pair (a and a^*), consider a set of vectors between pairs of words related the same way [11]. To some extent, these methods, baptised as 3CosAvg and LRCos, can generalise the vectors that represent the target relation, and thus be used for relation discovery.

3CosAvg and LRCos have shown to perform better, not only for English [11], but also for Portuguese, where they have been used for solving a translation of GAT [33] and also a newly created dataset, TALES, focused on Portuguese lexico-semantic relations [17]. The latter work also showed that lexico-semantic analogies are significantly more challenging to solve, because there are many relation instances sharing the same argument, thus allowing for several correct answers. In fact, sometimes, correct answers are just too many to be included in a dataset or lexical resource. This further suggests that these methods can be useful for automatically suggesting potentially missing links in a lexical resource.

The aforementioned distributional representations lately became known as static word embeddings, because they have a single representation for each word, while neural language models, like BERT [10], are based on contextual embeddings, i.e., the same word is represented differently, depending on its context. There is recent work on using neural language models in related tasks, such as filling blanks in short sentences that denote specific semantic relations, and thus discovering relations of such types [30, 3]; word sense disambiguation [36], given their contextual representations; and even analogy-solving [12], despite the lack of context in analogy tests. However, exploring those models is out of the scope of this work.

Soon, researchers noted that analogy-solving methods could be assessed in the discovery of morphological and semantic relations, including lexico-semantic, from word embeddings [15]. Moreover, other researchers assumedly used word embeddings for extending wordnets, e.g. for discovering new synsets and scoring candidate hypernyms by combining distances in the wordnet graph and their distributional similarity [31]. Others worked on the automatic construction of the whole wordnet from scratch, using word embeddings, in addition to bilingual dictionaries [22].

Wordnets, focused on lexical knowledge, were also extended with world knowledge, e.g., by linking them with Wikipedia, as in the BabelNet project [25]. For Portuguese, on this scope, Onto.PT is an automatically-created wordnet [16] that combines information in existing thesauri with relations extracted from several Portuguese dictionaries [18]. On the other hand, OpenWordNet-PT [8], used in this work, is a Portuguese wordnet aligned with PWN, originally developed as a syntactic projection of the Universal WordNet [7], but, since then, manually maintained (see more in Section 3).

3 OpenWordNet-PT

OWN-PT is an ongoing project to create a large wordnet for Portuguese. It has currently 52,559 synsets, 52,210 word forms and 83,841 senses.[2] It is the Portuguese wordnet in the Open Multilingual WordNet (OMW) [1] project, Freeling [26], BabelNet [25] and Google Translate.[3] OWN-PT synsets are aligned with the corresponding PWN synset and relations among the PWN synsets are projected to the OWN-PT synsets. OWN-PT is distributed in RDF following the vocabulary first described by de Paiva et al. [9].

In PWN, the main relation among words is synonymy. Synonyms – words that denote the same concept and are interchangeable in many contexts – are grouped into synsets. Each PWN synset is linked to other synsets by means of a small number of conceptual relations. Word forms[4] with several distinct meanings are represented in as many distinct synsets. Thus, each form-meaning pair (i.e., a word sense, the occurrence of a word in a synset) in PWN is unique. Synsets and word senses are interlinked by means of conceptual-semantic and lexical relations. The latter hold between word senses, whereas semantic relations hold between synsets; and there is also a small set of relations between synsets and word senses. Examples of semantic relations in PWN are: hyperonym, hyponym, meronym/holonym (part, substance and member), troponyms. Examples of lexical relations are: antonym and derivationally related.

The majority of the PWN relations connect words of the same part-of-speech (POS). Thus, PWN really consists of four sub-networks, respectively for nouns, verbs, adjectives and adverbs, with few cross-POS pointers. Cross-POS relations include the "morphosemantic" links that hold among semantically similar words sharing a stem with the same meaning, e.g., observe (verb), observant (adjective), observation and observatory (nouns). In many of the noun-verb pairs (i.e., nominalizations) the semantic role of the noun with respect to the verb has been specified, e.g., "painter" is the agent of "paint" (verb) while "painting" and "picture" is its result.

OWN-PT synsets are also classified into two additional classes: Core and Base. "Core" synsets are obtained from a semi-automatically compiled list with the 5,000 most frequently used word senses, followed by some manual filtering and adjustment by the PWN team [4]. The notion of base concepts was introduced in the EuroWordNet project [35] to reach maximum overlap and compatibility across wordnets in different languages. At the same time, this allows for the distributive development of wordnets in the world, each wordnet being a language specific structure and lexicalization pattern. "Base" Concepts are selected to be those that play an important role in the various wordnets of different languages.

4 Analogy Tests from OpenWordNet-PT Contents

Our main goal was to explore static word embeddings in the discovery of relation instances that could be useful for enriching OWN-PT. We thus needed an implementation of useful methods for this purpose, as well as data for training and assessing them.

The most common method for computing an analogy in the embedding space is to compute the vector offset [24], also known as the 3CosAdd method (Equation 1).

$$b^* = \underset{w \in V}{\operatorname{argmax}} \cos(w, a^* - a + b) \tag{1}$$

[2] Numbers can be compared to other open wordnets listed in the OMW at `http://compling.hss.ntu.edu.sg/omw/`.

[3] `https://translate.google.com/intl/en/about/license/`

[4] The term "word form" refers to single words or multi-word expressions.

Yet, as referred in Section 2, analogy-solving methods like 3CosAvg and LRCos suit our purpose better, because they exploit such embeddings for learning the relation between several pairs of words. 3CosAvg (Equation 2) computes the average offset between words in position a and respective words in position a^*, in a set of relation instances of the target type. The answer, b^*, must maximise the cosine with the vector resulting from summing the average offset to b.

$$b^* = \underset{w \in V}{\operatorname{argmax}} \cos(w, b + avg\_offset) \tag{2}$$

LRCos (Equation 3) considers the probability that a word w belongs to the same class as other words in position a^*, as well as the similarity between w and b, measured with the cosine. Although any classification algorithm could be used for this, the default implementation of LRCos relies on logistic regression for computing the likelihood of a word belonging to the class of words a^*.

$$b^* = \underset{w \in V}{\operatorname{argmax}} P(w \in target\_class) * cos(w, b) \tag{3}$$

In order to analyse how well the previous methods could learn a selection of relations in OWN-PT, we adopted Vecto,[5] a package for loading static word embeddings that includes implementations of 3CosAvg and LRCos, and supports analogy tests in the format of the BATS test [15]. For this purpose, analogy tests were created from OWN-PT. Table 1 presents the twelve relations considered in their production. This choice was guided by the number of instances available (see below), but also by the kind of relations that we believe could be learned from word embeddings. Therefore, relations like "see also", "classified by" and "same verb group" were discarded.

Analogy tests are organized in two-column tabular text files. Each test has several lines with a question word, in the first column; and a list of possible answers, in the second. All the words in the answer have to be related to the question word, according to OWN-PT. A different test was created for each relation, meaning that, in the same test, the relation between the question words and those in the answer was always the same. Figure 1 illustrates the format of the analogy test files with examples for three relations. For better understanding, rough translations were added for each line, but they are not part of the test.

For the creation of the tests, each conceptual-semantic relation instance between synsets was first expanded into a cartesian product of their word senses, using the SPARQL query in Listing 1. This query can be submitted to the OWN-PT SPARQL endpoint at `http://openwordnet-pt.org`.

Then, we group the instance pairs for each relation by their first projection (source, first column) and list all the related words in the second column (target, second column). Several experiments were made, for further improving the quality of the tests, given our goal. For instance, it is expected that the analogy-solving methods will learn better representations from single-sense words that are frequent enough in corpora. Specifically, in this work, we decided to consider only lines where the question word is in a "Core" synset. Moreover, the words in the answer were ordered so that words in "Core" synsets, if there were any, and words with fewer senses were listed first. This became relevant once we noticed that, in the training phase, Vecto considers only the first word in the list of possible answers. After this, we decided not to use tests with fewer than 30 questions (lines), or with more than 1,000,

[5] `https://github.com/vecto-ai`

atividade (activity)	ativo/agencioso/inativo (active / inactive)
bem-estar (wellness)	doentio/adoentado/insalubre/salubre/doente/são/saudável (sick/diseased/salubrious/unhealthy/healthy)
bondade (goodness)	boa/bom (good)
aberto (open)	fechado (closed)
abstrato (abstract)	concreto (concrete)
alto (tall)	baixo (low)
pálpebra (eyelid)	celha/conjuntiva/pestana/cílio (lash/conjunctiva/eyelash/cilium)
pássaro (bird)	plumagem/ala/bico/pluma/asa/pena/culatra/fúrcula/garupa/... (plumage/wing/beak/feather/wing/rump/croup)
pé (foot)	calcanhar/alfândega/artelho/polegada/dedo/calcâneo/hálux/sola (heel/calf/ankle/insole/finger/sole)

Figure 1 Excerpts of the generated datasets and rough translations, for the relations: attribute, antonymOf, partMeronymOf.

Listing 1 SPARQL query to produce the input data for Vecto.

```
select ?au ?t1 ?rel ?t2 (group_concat(?bu; separator = "/") AS ?values)
{
 ?s1 wn30:containsWordSense ?ss1 ; a ?t1 ; a wn30:CoreSynset .
 ?s1 skos:inScheme <http://logics.emap.fgv.br/wn/> .
 ?ss1 wn30:word/wn30:lexicalForm ?a .
 ?s2 wn30:containsWordSense ?ss2 ; a ?t2 .
 ?ss2 wn30:word/wn30:lexicalForm ?b .
 ?s1 skos:inScheme <http://logics.emap.fgv.br/wn/> .
 BIND(replace(lcase(str(?b))," ","_") AS ?bu)
 BIND(replace(lcase(str(?a))," ","_") AS ?au)

 ?s1en ?rel ?s2en .
 ?s1en owl:sameAs ?s1 .
 ?s2en owl:sameAs ?s2 .
}
group by ?au ?rel ?t1 ?t2
```

which included, for instance, hypernymOf and hyponymOf. If few questions would not be enough for generalizing the relations, the option for not considering larger tests was mostly practical, having in mind the manual validation and analysis of the results. This does not mean that, in the future, these relations cannot be considered as well.

5 Accuracy in Relation Learning

In order to run 3CosAvg and LRCos in the OWN-PT analogy tests, we used Vecto on the 300-sized Portuguese GloVe embeddings from the NILC repository [20]. This choice was supported by previous works, for English [11] and for Portuguese [33, 17], where GloVe embeddings have shown to perform better when it comes to solving semantic analogies.

At a lower level, each analogy-solving method is trained with every line of the test – corresponding to the question word (first column) and the answer (first word in the second column) – except one, and then tested on the remaining line, i.e., given the word in the target question (b), the model learned from all other questions and their answers tries to predict one of the words in its answer (b^*). In the end, Vecto computes the average accuracy of repeating the previous process for every question in the test.

For every considered relation, Table 1 shows the number of questions in its test and the accuracies achieved with the analogy solving methods – 3CosAvg and LRCos – and also with simple similarity (SimToB), here used as a baseline. For each question word, the latter consists of answering with its most similar word in the embeddings, i.e., the one maximising the cosine similarity. This also helps to take conclusions on whether the analogy-solving methods are improving upon this simple computation.

In fact, for eight out of 12 relations, analogy-solving methods lead to improvements, and there is only one (memberMeronymOf) for which both of them perform below the baseline. For the former eight relations, the best performance is achieved with LRCos, whereas for three of the remaining four 3CosAvg matches the performance of the baseline. Out of them, the accuracy of LRCos is 0 for the relation for which available data is less (substanceHolonym).

Table 1 Accuracy of the 3CosAvg and LRCos methods in different types of relation.

Relation	Questions	Accuracy SimToB	Accuracy 3CosAvg	Accuracy LRCos
agent	75	1.3%	1.3%	**29.3%**
antonymOf	68	**19.1%**	**19.1%**	7.4%
attribute	88	2.3%	9.1%	**21.6%**
byMeansOf	41	14.6%	26.8%	**46.3%**
causes	60	5.0%	6.7%	**8.3%**
entails	123	5.7%	5.7%	**6.5%**
memberHolonymOf	157	**5.1%**	**5.1%**	4.5%
memberMeronymOf	77	**10.4%**	7.8%	3.9%
partHolonymOf	417	2.2%	3.1%	**7.2%**
partMeronymOf	569	1.2%	1.2%	**3.0%**
substanceHolonymOf	33	**6.1%**	**6.1%**	0.0%
substanceMeronymOf	84	1.2%	2.4%	**7.1%**

Still, despite the noted improvements over the baselines, accuracies are still poor – for LRCos, only three (agent, attribute, byMeansOf) are above 20% and none is above 50%. The more homogeneous the first arguments of a relation are, the better LRCos seems to perform, which makes sense, because it makes the task of the classifier easier. For instance, in the byMeansOf and agent relations, first arguments are of a specific kind of verb, which favors the underlying classification, considered by LRCos. Despite some improvements, the performance of 3CosAvg is more in line with the baseline, with higher accuracy for relations with more semantically-similar arguments, starting with antonymOf.

On the one hand, figures show that generalising lexico-semantic relations in word embeddings is a challenging task, even if much more challenging for some relations (e.g., part) than for others (e.g., attribute, byMeansOf). On the other hand, accuracy is computed in OWN-PT, a resource that tries to cover the whole Portuguese language but is in constant development and, as it happens for all wordnets, has its gaps.

Moreover, accuracy is far from telling the whole story, because it only considers the first answer. Despite this fact, the report generated by Vecto also provides the top-n answers for each question. And if, for some relations typically found in an analogy test (e.g., morphological relations, country-capital, country-currency) questions tend to have a single answer, this does not happen for many lexico-semantic relations, e.g., an object will generally have several parts, and an attribute will have several possible values. Following the aforementioned reasons, we saw the list of top answers given as a useful source of suggestions for new relation instances in OWN-PT, i.e., the approach taken could be seen as an automatic way of providing such suggestions.

To better illustrate this, we show the top-8 answers for "*aperfeiçoar* (ameliorate) causes b^*", after automatic lemmatization (see Section 6) and removal of resulting duplicates: *aprender* (learn), *rever* (review), *trabalhar* (work), *repensar* (rethink), *evoluir* (evolve), *precisar* (need), *progredir* (progress), *melhorar* (improve). Out of them, only one is in OWN-PT (*melhorar*), in the eighth position, while six others could be considered as correct, but are just not in OWN-PT. Of course that there are also questions with no useful answers like, for instance, "sagrado (sacred) antonymOf b^*". Out of the answers for this question, only three matched the adjective POS: *eterno* (eternal), *religioso* (religious) and *obscuro* (obscure). Even if a different relation could possibly be established between some of them, none is an antonym of the question word. Next section tries to better quantify the proportion of potentially useful relations that could be suggested by this approach, with a manual validation.

6 Utility Analysis

Following the experiment reported in the previous section and the considerations regarding the potential utility of the given answers, we aimed at better quantifying that utility. This was necessary for better ascertaining the applicability of the analogy-solving methods for enriching wordnets, specifically OWN-PT.

For this purpose, we sampled a list of relation instances for manual inspection and human validation. As a preliminary validation, the criteria for selecting the relations to sample were pragmatic: we tried to cover four significantly different relation types, with varying performances in the first experiment (Section 5), also having in mind how easy it would be for a human to judge on their quality. Such a selection would mean a conservative estimation of the benefits of the proposed approach for enriching OWN-PT. It would also confirm the limited conclusions one can take from the accuracy values achieved and the preliminary inspection of the given answers.

For each selected relation, the sample included ten questions and their answers by LRCos, the method with the highest accuracy for more relations. Validation consisted of judging whether a relation of the given type actually holds between the question and each of the answers (e.g., *dente* parHolonymOf *cabeça*?).

Evaluating semantic relations between out-of-context words is always a challenging task. Despite this fact, as a preliminary evaluation, we decided to keep it simple and our main focus was on judging whether a relation of the target type can actually hold between the question and each of the answers (e.g., *largura* (width) attribute *transversal* (transversal)?). The sample was to be annotated in a spreadsheet, with relations meaning clarified by canonical examples (e.g., attribute *altura*-NOUN $\rightarrow$ *alto*-ADJ, in English, height-NOUN, high-ADJ). A human annotator had to label the suggested relation as Correct (i.e., the relation may hold for the question-answer pair) or Incorrect (i.e., the relation does not hold for the pair).

Yet, in order to accelerate human validation, some answers in the sample were automatically validated before presented to the annotators. This was performed with the help of MorphoBR [6], a large-coverage full-form lexicon for the morphological analysis of Portuguese, and included the following checks:

- If the POS of the answer did not match the POS of the range of the target relation, it was automatically labelled as invalid. For instance, if a relation is defined to hold between nouns and adjectives (e.g., attribute), answers that were not found in the lexicon's adjectives would fail this test;
- If the POS of the answer matched the POS of the range of the target relation but was not in the lemma form, the answer was lemmatized. In a minority of cases, this could lead to duplicate answers.

Moreover, if the answer was already in OWN-PT, it was automatically labeled as correct.

Table 2 summarises the results of this manual validation when made by one of the authors of this paper, who is part of the team that maintains OWN-PT. It organises the answers into those: corresponding to relation instances already in OWN-PT; invalid due to incompatible POS; or not in OWN-PT, but labelled as Correct. For some instances, the annotator provided an additional comment that the relation is incorrect, but a relation of a different type indeed holds between question and answer (e.g., synonymy instead of antonymy).

In a later stage, the sample was also validated by another author of the paper, which enabled us to measure the Cohen's Kappa κ. When the automatically labeled entries are not considered, κ was 0.63, which corresponds to substantial agreement [23].

Table 2 Summary of manual validation of 376 pairs of words, covering four relations, by one human annotator (OWN-PT maintainer). Numbers in parenthesis are percentages for each relation, considering all the entries in the sample.

Relation	Total	In OWN-PT	Invalid	Correct	Other Relation
antonymOf	90	0	39 (43%)	3 (3%)	26 (29%)
attribute	94	3 (3%)	35 (37%)	27 (29%)	23 (24%)
causes	95	2 (2%)	39 (41%)	10 (11%)	4 (4%)
partHolonym	97	6 (6%)	22 (23%)	26 (27%)	17 (18%)

We see that, depending on the relation, useful suggestions vary significantly. For instance, for antonymyOf only three were labeled as correct, whereas for attribute and partHolonym more than a quarter of the suggestions were good, respectively 29% and 27%. It is also clear that these figures are not proportional to the accuracies achieved for each relation in Section 5, confirming that those results are of limited application. For instance, the accuracy for the aforementioned relations with LRCos is as different as 21% (attribute) and 7% (partHolonymOf).

Despite the simplicity of the task, some non-trivial examples were not hard to find. For instance, *ponta* (lead, end, point, or tip) is indeed a part holonym of many objects (i.e., many objects do have a tip), but the challenge is to identify those objects where it is important to have this relation explicit. Among the good findings, some could be added to OWN-PT right away, including the following examples:

- *integrado* (integrated) antonym of *separado* (separate);
- *ideologia* (ideology) attribute *marxista* (Marxist);
- *aperfeiçoar* (ameliorate) causes *evoluir* (evolve);
- *dente* (tooth) part holonym of *elefante* (elephant);

Considering all four relations in the sample, the proportion of useful suggestions is about 17%. Yet, we should note that a significant proportion (39%) of the suggestions was automatically labeled, most of which for being invalid. This made it possible to decrease the amount of suggestions that required human validation. If such suggestions are ignored, the proportion of useful suggestions is close to 29%. This shows the potential of the proposed approach for accelerating the process of enriching wordnets, by suggesting the inclusion of relation instances that are missing from the resource. At the same time, this proportion confirms that the process needs human intervention, i.e., we cannot simply add all suggestions automatically. In fact, a second step is still required for selecting the attachment points in OWN-PT, i.e., the synsets corresponding to the arguments of the suggested relations. Furthermore, the example of *ponta* suggests that better inclusion criteria are needed to improve human judgment.[6]

7 Conclusion

This paper described how methods for automatic analogy-solving with word embeddings were applied to the discovery of lexico-semantic relations in Portuguese. It further analysed the utility of discovered relation instances for enriching OWN-PT, a Portuguese wordnet. Even if the accuracy of such methods is poor, among other challenges, it is harmed by the gaps in the wordnet, resulting in the consideration of some answers that would be correct, as incorrect. Yet, as we have shown, some of the given answers are good suggestions for manual inclusion in the wordnet. In a small validated sample of answers, we found about 17% good suggestions. We also noted that some suggestions can be automatically labeled as invalid, leading to about 29% suggestions out of all that required human validation. We thus see the described approach as a promising avenue for finding gaps and enriching wordnets. Although applied to Portuguese, a similar procedure could be adopted for other languages for which a wordnet and a model of word embeddings are available. Still, this was just a preliminary validation. An evaluation considering more answers and all relation types should be performed in the future. Such an exercise may also enable an analysis of the confusion between relations, and possibly identify actual errors in OWN-PT.

Despite accelerating the process, human intervention is always required for discriminating correct suggestions. Moreover, since this approach is based on word representations and not word senses, a human will also be necessary to find the suitable attachment points (i.e., word senses) for the suggested relation instance in the wordnet. So far, when a relation instance involved a lemma not covered by the wordnet, this lemma was added to a proper synset, if there was one. If not, nothing was done. In the future, this might lead to the creation of new synsets.

The process of enriching and maintaining a wordnet is never over, and so is not this work. In the near future, we aim to make the process of relation suggestion from word embeddings more flexible. In addition to lemmatization and exclusion criteria (i.e., valid POS) already applied to the obtained suggestions, we will work on isolating the analogy-solving methods from Vecto, which will enable to select only a controlled subset of relations for training, and then apply the learned models to a broader test set. A controlled training set could consider only core concepts or single-sense words, and possibly also features like word frequency, concreteness / imageability [27], experiential familiarity [14], among others. At the same time, a different test set will enable the discovery of relations for any word.

It is also our intention to explore neural language models for this process. As others have shown [30, 3], BERT's masked language model can be used as source of relational knowledge. We could probably adopt their approaches for Portuguese, using a BERT model pretrained for our language [34]. Finally, it would be interesting to consider word senses in the process. This could be explored in the discovery step and include the exploitation of contextual embeddings, e.g., from BERT; or in the validation step, where looking at the discovered relations in context, ideally with disambiguated words, should help the human judgement.

[6] In `https://globalwordnet.github.io/gwadoc/` there is an initial attempt at consistent documentation and examples for semantic/lexical relations used by different wordnets.

References

1 Francis Bond and Ryan Foster. 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, 2013.

2 Francis Bond and Kyonghee Paik. A survey of wordnets and their licenses. *Small*, 8(4):5, 2012.

3 Zied Bouraoui, Jose Camacho-Collados, and Steven Schockaert. Inducing relational knowledge from bert. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7456–7463. AAAI, 2020.

4 Jordan Boyd-Graber, Christiane Fellbaum, Daniel Osherson, and Robert Schapire. Adding dense, weighted connections to wordnet. In *Proceedings of the third international WordNet conference*, pages 29–36. Citeseer, 2006.

5 Nicoletta Calzolari, Laura Pecchia, and Antonio Zampolli. Working on the italian machine dictionary: a semantic approach. In *COLING 1973 Volume 2: Computational And Mathematical Linguistics: Proceedings of the International Conference on Computational Linguistics*. Association for Computational Linguistics, 1973.

6 Leonel Figueiredo de Alencar, Bruno Cuconato, and Alexandre Rademaker. Morphobr: An open source large-coverage full-form lexicon for morphological analysis of portuguese. *Texto Livre: Linguagem e Tecnologia*, 11(3):1–25, 2018.

7 Gerard De Melo and Gerhard Weikum. Towards a universal wordnet by learning from combined evidence. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 513–522, 2009.

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

9 Valeria de Paiva, Livy Real, Alexandre Rademaker, and Gerard de Melo. Nomlex-pt: A lexicon of portuguese nominalizations. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, 2014. European Language Resources Association (ELRA).

10 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, NAACL-HLT 2019, pages 4171–4186. Association for Computational Linguistics, 2019.

11 Aleksandr Drozd, Anna Gladkova, and Satoshi Matsuoka. Word embeddings, analogies, and machine learning: Beyond king - man + woman = queen. In *Proceedings the 26th International Conference on Computational Linguistics: Technical papers COLING 2016*, COLING 2016, pages 3519–3530, 2016.

12 Kawin Ethayarajh. How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, 2019.

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

14 Morton A Gernsbacher. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of experimental psychology: General*, 113(2):256, 1984.

15 Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka. Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't. In *Proceedings of the NAACL 2016 Student Research Workshop*, pages 8–15. ACL, 2016.

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

17 Hugo Gonçalo Oliveira, Tiago Sousa, and Ana Alves. TALES: Test set of Portuguese lexical-semantic relations for assessing word embeddings. In *Proceedings of the ECAI 2020 Workshop on Hybrid Intelligence for Natural Language Processing Tasks (HI4NLP 2020)*, volume 2693 of *CEUR Workshop Proceedings*, pages 41–47. CEUR-WS.org, 2020.

18 Hugo Gonçalo Oliveira, Diana Santos, Paulo Gomes, and Nuno Seco. 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 2008. Springer.

19 Zelig Harris. Distributional structure. *Word*, 10(2-3):1456–1162, 1954.

20 Nathan S. Hartmann, Erick R. Fonseca, Christopher D. Shulby, Marcos V. Treviso, Jéssica S. Rodrigues, and Sandra M. Aluísio. Portuguese word embeddings: Evaluating on word analogies and natural language tasks. In *Proceedings 11th Brazilian Symposium in Information and Human Language Technology (STIL 2017)*, 2017.

21 Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of 14th Conference on Computational Linguistics*, COLING 92, pages 539–545, Morristown, NJ, USA, 1992. Association for Computational Linguistics.

22 Mikhail Khodak, Andrej Risteski, Christiane Fellbaum, and Sanjeev Arora. Automated wordnet construction using word embeddings. In *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*, pages 12–23, 2017.

23 J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.

24 Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *Proceedings of the Workshop track of ICLR*, 2013.

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

26 Lluís Padró and Evgeny Stanilovsky. FreeLing 3.0: Towards wider multilinguality. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2473–2479, Istanbul, Turkey, May 2012. European Language Resources Association (ELRA).

27 Allan Paivio, John C Yuille, and Stephen A Madigan. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, 76(1p2):1, 1968.

28 Patrick Pantel and Marco Pennacchiotti. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Procs of 21st International Conference on Computational Linguistics and 44th annual meeting of the Association for Computational Linguistics*, pages 113–120, Sydney, Australia, 2006. ACL Press.

29 Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2014, pages 1532–1543. ACL, 2014.

30 Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. Language models as knowledge bases? In *Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473. Association for Computational Linguistics, 2019.

31 Heidi Sand, Erik Velldal, and Lilja Øvrelid. Wordnet extension via word embeddings: Experiments on the norwegian wordnet. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*, pages 298–302, 2017.

32 Rion Snow, Daniel Jurafsky, and Andrew Ng. Learning syntactic patterns for automatic hypernym discovery. *Advances in neural information processing systems*, 17:1297–1304, 2005.

33 Tiago Sousa, Hugo Gonçalo Oliveira, and Ana Alves. Exploring different methods for solving analogies with Portuguese word embeddings. In *Proceedings 9th Symposium on Languages, Applications and Technologies, SLATE 2020, July 13-14, 2020, School of Technology, Polytechnic Institute of Cávado and Ave, Portugal*, volume 83 of *OASIcs*, pages 9:1–9:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

34 Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. Bertimbau: Pretrained bert models for brazilian portuguese. In *Proceedings of the Brazilian Conference on Intelligent Systems (BRACIS 2020)*, volume 12319 of *LNCS*, pages 403–417, Cham, 2020. Springer.

35 P. Vossen. *EuroWordNet: A multilingual database with lexical semantic networks*. Computers and the humanities. Springer Netherlands, 1998.

36 Gregor Wiedemann, Steffen Remus, Avi Chawla, and Chris Biemann. Does bert make any sense? interpretable word sense disambiguation with contextualized embeddings. In *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*, pages 161–170, Erlangen, Germany, 2019. German Society for Computational Linguistics & Language Technology.