# Assigning Polarity Automatically to the Synsets of a Wordnet-like Resource

**Hugo Gonçalo Oliveira[1], António Paulo Santos[2], and Paulo Gomes[3]**

1 **CISUC, Department of Informatics Engineering**
**University of Coimbra, Portugal**
hroliv@dei.uc.pt
2 **GECAD, Institute of Engineering**
**Polytechnic of Porto, Portugal**
pgsa@isep.ipp.pt
3 **CISUC, Department of Informatics Engineering**
**University of Coimbra, Portugal**
pgomes@dei.uc.pt

## Abstract

This article describes work towards the automatic creation of a conceptual polarity lexicon for Portuguese. For this purpose, we take advantage of a polarity lexicon based on single lemmas to assign polarities to the synsets of a wordnet-like resource. We assume that each synset has the polarity of the majority of its lemmas, given by the initial lexicon. After that, polarity is propagated to other synsets, through different types of semantic relations. The relation types used were selected after manual evaluation. The main result of this work is a lexicon with more than 10,000 synsets with an assigned polarity, with accuracy of 70% or 79%, depending on the human evaluator. For Portuguese, this is the first synset-based polarity lexicon we are aware of. In addition to this contribution, the presented approach can be applied to create similar resources for other languages.

**1998 ACM Subject Classification** I.2.7 Natural Language Processing

**Keywords and phrases** sentiment analysis, polarity, lexicon, wordnet, Portuguese

**Digital Object Identifier** 10.4230/OASIcs.SLATE.2014.169

## 1 Introduction

The World Wide Web has become an important resource for supporting decision making. People often turn to this infrastructure to search for pieces of information that, hopefully, will help them choose their next smartphone, a country to visit during the holidays or, sometimes, even to decide on what party to vote in the next election. But it is a well-known fact that a rational decision is highly limited both by the available information and by the available time. So, as the Web keeps growing, the aforementioned procedure is losing efficiency.

Sentiment analysis (SA), or opinion mining (see [20] for an extensive survey), deals with the computational treatment of opinions and sentiments in order to provide more efficient decision making. Opinions given by an author towards a subject are typically classified according to their polarity, which might be positive, negative and, sometimes, just neutral. Many approaches to this topic rely on existing linguistic resources to predict the polarity of a given piece of information. One of the main research lines in SA is thus the creation of adequate sentiment resources, including annotated corpora and polarity lexicons.

Most polarity lexicons are structured in word lemmas, identified by their orthographical form, and the typical polarity they express. However, natural language is ambiguous and

3rd Symposium on Languages, Applications and Technologies (SLATE'14).
Editors: Maria João Varanda Pereira, José Paulo Leal, and Alberto Simões; pp. 169–184

OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

words have different meanings which, depending on the context, might lead to different polarities for the same word. So, the previous representation fails to capture words with senses with different polarities. This problem has been recognised and lead to the creation of polarity lexicons based on concepts, materialised, for instance, by the attribution of polarities to the synsets of a wordnet. For English, SentiWordNet [3] and Q-WordNet [1] include a subset of Princeton WordNet [5] synsets and an automatically assigned polarity. Similarly to most language resources, the manual creation of polarity lexicons involves time-consuming human effort, which lead to the development of automatic tools for this task, either by exploiting corpora or other resources, such as wordnets (see section 2).

A polarity resource based on synsets enables the combination of SA with word sense disambiguation (WSD) techniques [19] to identify the polarity of words in context – WSD algorithms based on the structure of a wordnet can be used to identify the synset corresponding to a word meaning in context and then access a synset polarity lexicon to obtain the transmitted polarity. In fact, the benefits of WSD to SA were emphasised by the acknowledgement that a supervised sentiment classifier modeled on word senses performs better than one based on word-based features, when classifying the polarity transmitted by textual documents [2].

Given the importance of having such a resource, we set our goal to the automatic creation of a synset-oriented polarity lexicon for Portuguese. Our effort towards this goal involved the automatic assignment of polarities to the synsets of a Portuguese wordnet-like lexicon, Onto.PT [7] – a resource currently in development, extracted automatically from textual resources. Having our goal in mind, we describe a procedure with two steps: (i) one for initial polarity assignment; (ii) and another for polarity propagation. Both steps are relatively straightforward and can be applied to other languages, provided that there is a polarity lexicon on the words of that language. Moreover, during this work, there was nothing like a concept-oriented/synset-oriented polarity resource for Portuguese. Therefore, this work can be viewed as an important contribution to the development of Portuguese SA. For the same reason, we had to perform several manual evaluations, which we also describe.

In the proposed procedure, polarity is first assigned to synsets according to the polarity of the words they contain. The main idea is that the synset polarity will be the same as the typical polarity of the majority of its lemmas, when uncontextualised. Then, through several iterations, polarity is propagated to synsets connected, by semantic relations, to the previously polarised synsets.

After evaluating the results of this procedure manually, we concluded that the initial polarity assignment is an effective way of moving from existing polarity information based on lemmas, to polarity based on meanings, as the accuracy of this step is between 73% or 86%, depending on the human evaluator. As for propagation, regarding that polarities are not transmitted in the same way by all relation types, we conducted an experiment to investigate which relations could be exploited for this task. We concluded that relations between adjectives and qualities or states, as well as those between adjectives and adverbs, tend to preserve the polarity. Furthermore, antonymy relations invert the propagated polarity. Using only the best performing relations, polarity propagation is about 63% accurate. In the end, a polarity lexicon with about 10,300 synsets is obtained, with an estimated global accuracy between 70% and 79%.

This introduction is followed by section 2, where some related work is introduced, with special focus on the automatic creation of polarity lexicons. Section 3 gives a general presentation of our approach for assigning polarities to the synsets of a wordnet in two automatic steps: initial assignment and polarity propagation. Section 4 describes the resources

used when following the proposed approach for Portuguese, Onto.PT and SentiLex-PT [24]. Section 5 reports on the results of applying the initial assignment on Onto.PT synsets, including their evaluation. Following, section 6 reports on the results of applying polarity propagation on Onto.PT, including the evaluation performed towards the selection of the relation types to exploit. Before concluding, section 7 presents and discusses the overall evaluation of the polarity lexicon generated with the proposed approach and using the resources described earlier.

## 2 Related Work

In order to treat opinions computationally, in SA, sentiments are commonly converted to a simpler "formal language", such as the representation of polarity as numeric values. Polarity lexicons, or sentiment lexicons, are lists of words (or conceptual representations) classified according to their polarity, which may be positive, negative and, possibly, neutral. They are typically used as external sources of knowledge in the automatic classification of textual information, according to its polarity. Polarity lexicons are thus valuable resources for SA, and it is no surprise that their creation is one of the main research lines in this area. There are two main approaches for the automatic construction of polarity lexicons: (i) corpus-based; and (ii) wordnet-based.

Corpus-based approaches (e.g. [9, 26, 14, 12]) explore the co-occurrence of words in large collections of texts. Co-occurrence is explored using linguistic and statistical heuristics. For instance, conjunctions (e.g. *and*, *or*, *but*) between adjectives provide indirect information about their orientation [9] – while sequences like *fair and legitimate* and *corrupt and brutal* have the same orientation and may co-occur in a corpus, the pairs *fair and brutal* and *corrupt and legitimate* would be semantically anomalous. Other approaches compute the orientation of sentences based on their association with positive (e.g. *excellent*, *good*) and negative (e.g. *poor*, *bad*) references [26].

WordNet-based approaches (e.g. [16, 13, 23]) explore information provided by Princeton WordNet [5], or similar resources, to generate polarity lexicons. Exploited information goes from the semantic relations (e.g hypernymy, antonymy) connecting synsets, to textual glosses, which are additional sources of related words.

As words have different meanings, the same word might have senses with different polarities. Polarity lexicons structured in simple lemmas are thus impractical for most applications. This problem has been recognised and lead to the creation of polarity lexicons structured on concepts. For English, SentiWordNet [3] is an example of such a resource. Each relevant WordNet synset is assigned three numerical scores indicative of how positive, negative, and objective (neutral) is the sentiment it transmits. Synsets are classified after combining the results of eight ternary classifiers. The classification score is proportional to the number of classifiers that have assigned one of the three polarities.

Instead of relying on supervised classifiers, the creation of Q-WordNet [1] is based on an unsupervised binary classifier that tries to link each synset to a positive or negative quality. For this purpose, they start with the WordNet synsets that are in an attribute relation with a sense of the word *quality*, which include the adjective synsets with senses of *good*, *bad*, *positive*, *negative*, *superior* and *inferior*. Synsets that are accessible from the previous are then polarised, according to their connection.

Synset-based polarity lexicons for other languages have also been created. Some are based on the automatic translation of existing English resources, including SentiWordNet (see [15, 17]). Of those, some start with a set of manually labelled synset seeds and propagate

polarities to other synsets, through some of the semantic relations [17]. In addition to the ideas developed in Q-WordNet, the authors of SentiWordNet have shown that a random-walk model as the PageRank algorithm may be used for assigning polarities automatically to synsets, and thus expand polarity lexicons [4]. For this purpose, WordNet is seen as a graph, where synsets are nodes connected by relations $<synset_1$ referred-by $synset_2>$. PageRank is ran twice: first, to obtain the positivity strength, only the positive synsets of SentiWordNet have inital weights; then, the same is done for negative synsets.

For Portuguese, the field is growing and there have been a few attempts to create or enrich polarity lexicons automatically. But so far, existing resources are structured in words and not concepts. Exploratory work on the automatic construction of a word-based polarity lexicon for Portuguese includes combining information from different sources [25], propagating polarity through dictionary entries [21], and exploiting synonymy resources for expanding a handcrafted polarity lexicon [24]. On the first [25], a polarity lexicon was obtained after combining information from: (i) the application of Turney's method [26] to a Portuguese corpus of movie reviews; (ii) the application of Kamps' method [13] to the Portuguese thesaurus TeP [18]; (iii) the translation of Liu's English Opinion Lexicon [11] to Portuguese. On the second, starting with a small set of positive and negative seeds (about 10), textual patterns in the definitions of an electronic dictionary were exploited for polarity propagation [21]. The third is JALC [24], an algorithm for the automatic expansion of a handcrafted polarity lexicon of Portuguese, SentiLex-PT. Still, about three quarters of the entries of SentiLex-PT are the result of manual labour. Another work [6] describes on the manual creation of a polarity lexicon for Portuguese, though much smaller than SentiLex-PT, based on the analysis of a corpus of book descriptions.

Our work combines several ideas from the aforementioned works. More precisely, it assumes that all the words of a synset contribute to its overall polarity (as in [3]) and that polarity is propagated through several semantic relations (as in [4, 1, 17, 21]).

## 3 Assigning Polarity to Wordnet Synsets

This section gives a general overview of our approach for assigning polarity to part of the synsets of a wordnet. This can be achieved by following a straightforward automatic procedure with two sequential steps, namely:

1. Initial polarity assignment, described in section 3.2;
2. Polarity propagation through semantic relations, described in section 3.3.

The initial assignment step can be used, for instance, to assign polarities to the synsets of a simple thesaurus, containing just synsets and not synset links. On the other hand, polarity propagation is made through semantic relations, and thus requires that the initially polarised synsets are explicitly connected to other synsets, according to their meaning. Before describing each step, in section 3.1 we refer the kind of pre-existing resources that are required for applying this procedure and, at the same time, we introduce the notation used to describe each step.

### 3.1 Set-up

The starting point of this procedure in an existing polarity reference $R$, and a wordnet with synsets $W$. $R$ is a list of pairs $(l_i,\ pol(l_i))$ that assigns polarities to lemmas, $R = \{< l_1, pol(l_1) >, < l_2, pol(l_2) >, ..., < l_n, pol(l_n) >\}$, $n = |R|$. The polarity of a lemma, $pol(l_i)$, denotes the sentiment typically expressed by the lemma, when alone, which can be

positive (+1), negative (-1) or neutral (0). This is usually the polarity of the most frequent sense of the lemma, and the one that first comes to mind when in its presence.

A wordnet contains synsets and semantic relations. A synset $S$ is a set of synonymous lemmas $l_i$, $S = \{l_1, l_2, ..., l_m\}$, $m = |S|$. This means there is a context where all the lemmas of a synset have the same meaning and they can be seen as the possible lexicalisations of the same natural language concept. Semantic relations have a defined type, $RT$, and connect two synsets according to their meaning $<S_i\ RT\ S_j>, S_i \in W, S_j \in W$.

## 3.2 Initial Polarity Assignment

The initial assignment is based on the assumption that all the lemmas in a synset contribute to its overall polarity, as in SentiWordNet. Even though some lemmas might have other senses, we believe that the majority of lemmas in a positive synset will be labelled as such in $R$. Likewise, negative synsets will have a majority of negative lemmas and neutral synsets will contain mostly neutral lemmas.

For the initial assignment, each synset has a counter for each polarity value, respectively $c_+$, $c_-$ and $c_0$, all initially set to 0. Then, for each lemma in a synset that is also in the reference, $l_i \in S \wedge l_i \in R$, the polarity of the lemma according to $R$, $pol(l_i)$, is summed to the corresponding counter:

$$\begin{aligned} pol(l_i) > 0, &\quad c_+ = c_+ + 1 \\ pol(l_i) < 0, &\quad c_- = c_- + 1 \\ pol(l_i) = 0, &\quad c_0 = c_0 + 1 \end{aligned}$$

The overall synset polarity, Pol(S), will be positive (+), negative (-) or neutral (0), if the counter with the highest value is $c_+$, $c_-$ or $c_0$ respectively. If there is a tie, the synset is considered to transmit an ambiguous sentiment and will not have an assigned polarity. To illustrate this step, take the following example:

- **Reference** ($R$):
  - $nice(+)$
  - $overnice(-)$
  - $squeamish(-)$
  - $prissy(-)$
- **Synset** ($S$):
  - $\{squeamish, prissy, overnice, nice, dainty\}$
- **Counters**: $c_+ = 1, c_- = 3, c_0 = 0$
- $max(c_x) = c_- \implies Pol(S) < 0$

## 3.3 Polarity Propagation

The propagation step is based on the assumption that the polarity of related synsets tends to related. Therefore, after the initial assignment, all synsets with an assigned polarity can propagate it to their adjacent synsets. In other words, polarised synsets transmit their polarity directly to related synsets, which are those directly connected by a semantic relation.

This might not be true for all types of relations. For instance, *joy* and *sadness* are both *emotions*, meaning that <*emotion* hypernym-of *joy*> and <*emotion* hypernym-of *sadness*>.

But *emotion* can have neutral polarity, while *joy* is definitely positive and *sadness* is negative. Similar situations could happen for meronymy relations, as an object can have both "good" and "bad" parts. In order to identify the relations that propagate polarity more consistently, and could thus be exploited for automatic polarity propagation, we conducted the experimentation described in section 6.1.

Propagation can go through several iterations and occur for both ways. Therefore, in a relation between synsets $S_a$ and $S_b$, $<S_a\ RT\ S_b>$, if $S_a$ is polarised and $S_b$ is not, the polarity of $S_a$ is propagated to $S_b$, as well as, if $S_b$ is polarised and $S_a$ is not, $S_b$ propagates its polarity to $S_a$, such that $Pol(S_a) = Pol(S_b)$ or, depending on the relation, possibly $Pol(S_a) = -Pol(S_b)$. Take the following illustrative example, considering that *similar-to* is a semantic relation that propagates the same polarity and *antonym-of* is a semantic relation that propagates an inverted polarity.

- **Synsets**:
  - $S_0 = \{squeamish, prissy, overnice, nice, dainty\}$
  - $S_1 = \{fastidious\}$
  - $S_2 = \{unfastidious\}$
- $Pol(S_0) < 0$
- $<S_0$ similar-to $S_1> \implies Pol(S_1) < 0$
- $<S_0$ antonym-of $S_2> \implies Pol(S_2) > 0$

This process can occur for several iterations and synsets already polarised can be reached again. But, as the accuracy in polarity attribution decreases for higher iterations, we decided to keep only the polarity transmitted in the first iteration the synset is reached. Its polarity does not change in further iterations. Still, if a synset is reached more than once in the same iteration, we have used the three counters, in a similar fashion to the initial assignment. This way, in the end of each iteration, the synset gets the polarity corresponding to the counter with the highest value.

## 4 Used Resources

The automatic creation of a synset-based polarity lexicon for Portuguese relied on two existing resources, both freely available from the Web. In this section, we present both of them, namely: SentiLex-PT [24], a lemma-based polarity lexicon, and Onto.PT [7], a wordnet-like lexical knowledge base.

### 4.1 SentiLex-PT

SentiLex-PT is a polarity lexicon for Portuguese, compiled from several publicly available Portuguese resources. The sentiment entries of this lexicon are words, associated with their morphological properties and predicted sentiment towards human subjects. SentiLex-PT is especially suitable for opinion mining in Portuguese, particularly for detecting and classifying sentiments and opinions targeting human entities.

We used SentiLex-PT02[1], the most recent version of this resource. It is available in two files: (i) one where the word entries are inflected; (ii) and a "compressed" file where

[1] Available from `http://dmir.inesc-id.pt/project/SentiLex-PT_02`

all entries are lemmatised. In this work, we used the second one, because Onto.PT also contain lemmatised words. This lemma-oriented file covers 7,014 lemmas, of which 4,779 are adjectives, 1,081 are nouns, 489 are verbs and 666 are idiomatic expressions.

Besides other properties, each entry of SentiLex-PT02 contains the polarity of the word, which can be positive, neutral or negative, and also its kind of annotation, which is either manual (5,473 lemmas) or automatic (1,541 lemmas). Manual polarity labels were given by a linguist and automatic labels were assigned by the JALC algorithm, which its authors claim to be 87% accurate [24]. Out of curiosity, the lemma-oriented file of SentiLex-PT02 contains about three times more lemmas with negative polarity (4,596) than positive (1,548), and just 860 lemmas have neutral polarity.

## 4.2 Onto.PT

Onto.PT is a lexical-semantic knowledge base for Portuguese, structured similarly to Princeton WordNet [5]. As typical wordnets, Onto.PT tries to cover the whole language and not just specific domains. It is also structured on synsets, which group synonymous word senses, represented by lemmas, that may be seen as natural language concepts. For each of their senses, polysemous words are included in a different synset. Synsets may be connected to other synsets by means of semantic relations, which help describing possible interactions between their meanings.

However, we refer to Onto.PT as a wordnet-like resource because, in opposition to typical wordnets, it is not handcrafted, but created automatically, by exploiting Portuguese dictionaries and thesauri. The construction of Onto.PT is briefly described in three steps, that comprise the ECO approach [7]:

1. Regularities in the definitions of dictionaries are exploited for the extraction of instances of semantic relations, connecting words, identified by their lemma.
2. If possible, each synonymy relation is attached to a synset in an existing Portuguese thesaurus. TeP [18], an electronic thesaurus for Brazilian Portuguese, is used in this step. Clusters are then identified in the set of unattached synonymy relations, and added as new synsets.
3. Graph-based similarities are used to integrate the rest of the semantic relations automatically. Each lemma argument of a relation is assigned to the most suitable synset. If there are no synsets with the lemma, a new synset is created with that lemma.

Following this procedure, it is possible to have a larger resource without having to rely on time-consuming manual work. This lead to other differences towards typical wordnets, including more relation types covered. In Onto.PT, relations go from well-known hypernymy and part-of, to relations established between words of different parts-of-speech (POS), including, for instance, purpose-of, manner-of, or has-quality.

Onto.PT was released in 2012 and, as the result of an automatic approach, it is always under development[2]. In this work, we have used version 0.3 of Onto.PT, which contains 160,791 unique lemmas, organised in 105,500 synsets – 60,197 nouns, 25,346 verbs, 17,961 adjectives and 1,996 adverb synsets – connected by 184,521 instances of semantic relations.

Other Portuguese wordnets, as MultiWordNet.PT[3] or OpenWordNet.PT [22], would not apply for this work because they are both smaller and cover mostly hypernymy and

[2] Check `http://ontopt.dei.uc.pt` for updates and additional information on Onto.PT.
[3] Check `http://mwnpt.di.fc.ul.pt/` for additional information on MultiWordNet.PT.

**Table 1** Examples of synsets and their polarity in the initial assignment.

| Synset | Polarity |
|---|---|
| *contente* (+), *alegre* (+), *satisfeito* (+), *radiante* (+), *feliz* (+), *jubiloso* (+)<br>(*content*, *cheerful*, *satisfied*, *radiant*, *happy*, *joyant*) | + |
| *verdadeiro* (+), *veraz* (+), *verídico* (+), *fidedigno* (+), *fiel* (+), *exacto*<br>(*true*, *truthful*, *veridical*, *reliable*, *faithful*, *exact*) | + |
| *esmorecido* (-), *débil* (-), *sumidiço*, *mortiço* (-), *fraco* (-), *apagado* (-)<br>(*faltering*, *feeble*, *dull*, *weak*, *out*) | - |
| *médio* (0), *mediano* (0), *medíocre* (-), *moderado* (+)<br>(*average*, *median*, *mediocre*, *moderate*) | 0 |
| *severo* (-), *implacável*, *justiceiro* (+), *estrito*, *rigoroso* (+), *incomplacente*, *inflexível* (-)<br>(*severe*, *implacable*, *justicer*, *strict*, *rigorous*, *austere*, *inflexible*) | *null* |

part-of relations. As discussed earlier, those are not the best suited for polarity propagation. Moreover, MultiWordNet.PT is not free for research purposes and only covers noun synsets.

## 5 Results of Initial Assignment

The polarity assignment procedure, presented in section 3 was followed in the creation of a synset-based polarity lexicon for Portuguese, using the resources described in section 4. We recall that this approach encompasses two steps: initial polarity assignment and polarity propagation. This section reports on the results of the first step, which consisted of assigning a polarity to the synsets of Onto.PT, using SentiLex-PT as the polarity reference. Section 6 presents the results of the polarity propagation in Onto.PT.

### 5.1 Quantities and Examples

The initial polarity assignment was applied to Onto.PT, using the full SentiLex-PT as a polarity reference. This resulted in 7,556 Onto.PT synsets with an assigned polarity, more precisely: 1,875 positive, 4,792 negative and 889 neutral. Of those, 1,374 synsets included both lemmas with positive and lemmas with negative polarities, but one of the three polarity counters was higher than the others. An additional 424 synsets were considered to be sentiment ambiguous. Table 1 shows examples of synsets, the polarity of their lemmas according to SentiLex-PT, and their consequently assigned polarity. The few lemmas without polarity are not covered by SentiLex-PT. Examples include synsets where all lemmas have the same polarity, even though some are not in SentiLex-PT, synsets where the resulting polarity is the most common, and a synset with ambiguous polarity (*null*).

### 5.2 Evaluation

In order to assess the results produced by the initial polarity assignment, we asked two human judges, both native speakers of Portuguese, to independently classify a sample of 390 synsets according to their polarity. Their goal was to select the polarity that the concept denoted by each synset transmits. All the synsets were randomly collected from the 7,556 synsets with polarities automatically assigned in this initial step, but the automatic polarity was not shown to the judges. Also, all the synsets of the sample had more than one lemma. First, this procedure is suited precisely for synsets with more than one lemma. Second, assuming that the sentiment labels in SentiLex-PT are correct, there was no need to evaluate

**Table 2** Evaluation of the initial assignment step.

| Sample | Reference | Target | P(+) | P(-) | P(0) | P(all) | Kappa |
|---|---|---|---|---|---|---|---|
| 390 sets | H1 | Aut | 0.92 | 0.88 | 0.38 | 0.86 | 0.73 |
| 390 sets | H2 | Aut | 0.72 | 0.75 | 0.50 | 0.73 | 0.53 |
| 390 sets | H1 | H2 | 0.77 | 0.81 | 0.82 | 0.80 | 0.66 |

any of the 1,315 polarised synsets with only one lemma[4].

We recall that each of the other synsets is a group of lemmas that, all together, denote a concept. Therefore, even if there is one or more lemmas that, in different contexts, may transmit different polarities, there should only be one meaning shared between all the lemmas and, in this context, only one polarity common to all of them. This fact is important because it minimises the ambiguity issues of polarity classification, as compared to the classification of single lemmas, outside a context.

Concerning the illustration of the aforementioned phenomena, we first present two meanings of the Portuguese word *queda*, which might either denote a downfall/tumble (negative) or an ability/capacity (typically positive), respectively in the following synsets:

- ***queda**, tombo, trambolhão, choque, baque, boléu*
- ***queda**, jeiteira, vocação, qualidade, aptidão, jeito, habilidade, capacidade*

Similarly, in the following synsets, the adjective *simples* might respectively refer to something simple/easy (typically positive) or an ignorant/uneducated (negative) person:

- ***simples**, fácil, desintrincado*
- ***simples**, inculto, bronco, ignorante, burgesso, néscio, desiluminado*

Table 2 presents the results of this evaluation. Besides the number of evaluated synsets (Sample), 'Reference' indicates the reference set of polarised synsets, which can be viewed as a golden set, while 'Target' is the evaluated set. In the later columns, 'H1' stands for the first human judge, 'H2' for the second human and 'Aut' refers to the synsets with polarities assigned automatically. In the same table, we provide the accuracy of our target according to the reference, which is the proportion of matches per polarity value (P(+), P(-), P(0)), the total accuracy (P(all)), and the agreement between the reference and the target, expressed by the Cohen's Kappa coefficient (Kappa). Although not common, this includes the agreement between the judges annotation and the system.

Evaluation shows that the initial assignment is an adequate and straightforward approach for moving from polarised lemmas to polarised synsets/meanings. Using the judge H1 as reference, the accuracy of the initial polarity assignment is 86%, and it is 73% against judge H2. Curiously, even though the agreement between the judges was good (0.66) [8], H1 had higher agreement with the system than with H2.

Accuracy is always higher than 70% for positive and negative synsets, but it is lower for the neutral synsets. Besides suggesting that it is not easy to identify objective/neutral synsets automatically, this highlights the fact that it is not easy for humans as well. Due to these problems, several works (e.g. [26, 23, 21]) just classify words as either positive or negative. In fact, the difference of accuracy between the two annotators is explained by their

[4] We did not consider the possibility that the sense of the lemma in the Onto.PT synset was different than in SentiLex-PT. The main reason for this is that, in Onto.PT, single-lemma synsets are unique. In fact, for rare situations, such as words with two completely different senses without synonyms, single-lemma synsets might merge different senses.

different sensibilities towards neutral synsets – the amount of synsets classified this way is 2.5 times higher for H2 than for H1. If neutral synsets were ignored, the accuracies would actually be 97% and 96% respectively for H1 and H2, with $\kappa = 0.96$.

## 6 Results of Polarity Propagation

On the proposed approach, the second step for creating a synset-based polarity lexicon is polarity propagation. However, before propagating polarities, blindly, through all types of relations, we only selected the types which seemed adequate for propagation. Then, we evaluated the result of their propagation in the first iteration, where four types of candidate relations revealed to be inadequate for this task. Only the remaining five types were used for propagation.

### 6.1 Candidate Relations

Regarding the analysis of several examples, as those in section 3.3, we decided not to use hypernymy nor meronymy for polarity propagation. There are works [17] where hypernymy is used for polarity propagation, but only when this relation is held between adjectives. This is not the case for Onto.PT, where hypernymy only connects nouns. On the other hand, we believed that several types of Onto.PT relations could transmit polarities more consistently, namely:

- Relations connecting an object or a process to a resulting/goal state or another object/process (causation, producer, purpose);
- Relations connecting properties, qualities or states with nouns or adjectives (refers-to, has-quality, has-state);
- Relations connecting nouns or adjectives with adverbs (manner).

Additionally, we believed that there were types of relations connecting positive with negative synsets, which thus transmit an inverted polarity. In Onto.PT, two types of relations fit in that group:

- Those connecting synsets with an opposite meaning (antonymy, which, in Onto.PT 0.3, only occurs between adjectives);
- Relations connecting nouns or verbs with manners that do not characterise them (manner-without).

For each of the aforementioned relation types, Table 3 presents an illustrative example, together with the number of instances in Onto.PT 0.3.

### 6.2 Selection of the Adequate Relations

Instead of using all the selected relations in polarity propagation, once again, we asked two human judges to independently classify the polarity of Onto.PT synsets of seven samples. Each sample contained synsets connected by one of the seven relation types in Table 3, to those polarised in the initial assignment. For evaluation purposes, the automatic results of propagation were compared to the manual classifications. Table 4 shows the results of this evaluation. The number of evaluated synsets differs according to the reference because, when judges could not attribute a well-defined meaning to a synset, they did not classify it. This happened because, although the judges were advised to look in online dictionaries for unknown meanings, Onto.PT contains some unfrequent words, as well as unfrequent senses of well-known words.

**Table 3** Relations of Onto.PT, evaluated for propagation.

**Has-quality:** 2,219 instances
{*inábil*} → {*desajeitamento,desjeito,inabilidade,desabilidade*}
{*unskilful*} → {*clumsiness,inability,lack_of_skill*}

**Has-state:** 573 instances
{*estável,permanente,efectivo*} → {*beatitude,paz,concordia,tranquilidade*}
{*stable,permanent*} → {*bliss,peace,tranquility*}

**Manner-of:** 3,924 instances
{*avidamente,vorazmente,sofregamente*} → {*sede,avidez,sofreguidão,avareza,cobiça,...*}
{*greedily*} → {*greed*}

**Antonym-of:** 687 instances
{*inconcludente,inconclusivo*} → {*liquidante,conclusivo,terminativo,terminante*}
{*inconclusive*} → {*conclusive,terminative*}

**Manner-without:** 316 instances
{*caladamente,silenciosamente,secretamente, ...*} → {*exposição,manifestação,declaração*}
{*quietly,silently,secretly*} → {*exposition,expression,declaration*}

**Causation-of:** 12,148 instances
{*causticar,cauterizar,calcinar,...*} → {*combustão,cremação,cauterização,calcinação,...*}
{*to_etch,to_cauterise*} → {*combustion,cauterization*}

**Producer-of:** 2,335 instances
{*destilação_de_petróleo*} → {*gasolina*}
{*oil_distillation*} → {*gasoline*}

**Purpose-of:** 16,918 instances
{*explorar_espaço*} → {*cosmonave,astronave,espaçonave,nave*}
{*to_explore_the_space*} → {*spacecraft,spaceship*}

**Refers-to:** 37,491 instances
{*caricaturesco,caricatural*} → {*caricatura,cartoon,cartum*}
{*caricatural*} → {*caricature,cartoon*}

On the performed evaluation, depending on the relation, the judge's agreement is between moderate and good [8]. Yet, we believe that we can rely on these results to conjecture on the adequacy of the Onto.PT relations for polarity propagation. As discussed earlier, the task of classifying the polarity of synsets depends on the judge's intuition. For instance, in the previous evaluation, we noticed different sensibilities towards the classification of the synsets as neutral. As for the actual evaluation results, Table 4 shows that manner-of and manner-without were the best performing relation types, which indicates that means and adjectives transmit their polarity to their corresponding adverbs. Not just these two, but all five relation types with best accuracy denote a pattern, as they all connect at least one adjective or adverb synset to another synset. Given that both adjectives and adverbs are used as modifiers, they are often connected to qualities and thus to sentiment, which explains this pattern. On the other hand, purpose-of and producer-of had the lowest accuracy. In fact, purpose-of is not as semantically well-defined as the other relations because it can connect very different things. To give an idea, it relates an action (verb), which can either be a general purpose (e.g. *to desinfect, to calculate, to censor, to dissociate*) or just something one can do with (e.g. *to punish, to transport, to climb, to spend, to entertain*), for instance, an instrument (e.g. *desinfectant, whip*), a concrete object (e.g. *van, stairs*), an abstract means (e.g. *credit, calculation, satire*), a human entity (e.g. *clown*), or a property (e.g. *dissociation*). And we should recall that the same instrument/means can be used either for positive or negative actions. As for the producer-of relation, most of its instances relate fruits and vegetables with their trees, which are rarely related to sentiment.

**Table 4** Evaluation of different relations in the first propagation iteration.

| Relation | Sample | Ref | Target | P(+) | P(-) | P(0) | P(all) | Kappa |
|---|---|---|---|---|---|---|---|---|
| Causation-of | 99 sets | H1 | Aut | 0.36 | 0.69 | 0.00 | 0.59 | 0.26 |
| | 100 sets | H2 | Aut | 0.63 | 0.42 | 0.00 | 0.56 | 0.26 |
| | 99 sets | H1 | H2 | 0.66 | 0.80 | 0.50 | 0.68 | 0.45 |
| Producer-of | 79 sets | H1 | Aut | 0.29 | 0.33 | 0.00 | 0.30 | 0.09 |
| | 77 sets | H2 | Aut | 0.29 | 0.43 | 0.00 | 0.36 | 0.14 |
| | 77 sets | H1 | H2 | 0.73 | 0.75 | 0.93 | 0.84 | 0.73 |
| Purpose-of | 99 sets | H1 | Aut | 0.29 | 0.23 | 0.83 | 0.29 | 0.12 |
| | 98 sets | H2 | Aut | 0.24 | 0.20 | 0.50 | 0.23 | 0.05 |
| | 97 sets | H1 | H2 | 0.57 | 0.77 | 0.86 | 0.78 | 0.57 |
| Refers-to | 118 sets | H1 | Aut | 0.37 | 0.54 | 0.33 | 0.48 | 0.17 |
| | 119 sets | H2 | Aut | 0.54 | 0.56 | 0.27 | 0.53 | 0.23 |
| | 117 sets | H1 | H2 | 0.52 | 0.78 | 0.61 | 0.67 | 0.48 |
| Has-quality | 85 sets | H1 | Aut | 0.85 | 0.81 | 0.17 | 0.78 | 0.60 |
| | 85 sets | H2 | Aut | 0.85 | 0.75 | 0.50 | 0.76 | 0.60 |
| | 85 sets | H1 | H2 | 0.90 | 0.90 | 0.57 | 0.85 | 0.75 |
| Has-state | 60 sets | H1 | Aut | 0.38 | 0.80 | 0.60 | 0.67 | 0.37 |
| | 60 sets | H2 | Aut | 0.50 | 0.77 | 0.40 | 0.67 | 0.38 |
| | 60 sets | H1 | H2 | 0.43 | 0.86 | 0.60 | 0.72 | 0.48 |
| Manner-of | 90 sets | H1 | Aut | 0.90 | 0.75 | 0.22 | 0.74 | 0.56 |
| | 90 sets | H2 | Aut | 0.83 | 0.75 | 0.00 | 0.70 | 0.48 |
| | 90 sets | H1 | H2 | 0.88 | 0.84 | 0.29 | 0.81 | 0.67 |
| Antonym-of | 60 sets | H1 | Aut | 0.52 | 0.79 | 0.43 | 0.62 | 0.42 |
| | 60 sets | H2 | Aut | 0.55 | 0.71 | 0.43 | 0.60 | 0.40 |
| | 60 sets | H1 | H2 | 0.71 | 0.92 | 0.74 | 0.80 | 0.70 |
| Manner without | 85 sets | H1 | Aut | 0.62 | 0.82 | 0.00 | 0.74 | 0.51 |
| | 85 sets | H2 | Aut | 0.66 | 0.82 | 0.00 | 0.75 | 0.51 |
| | 85 sets | H1 | H2 | 0.70 | 0.85 | 0.56 | 0.78 | 0.59 |

According to our interpretation, these results make sense, so we relied on them for selecting the relations to use. This means that polarity was propagated only through the five relation types with accuracy higher than 60%, namely: manner-of, has-quality, has-state, antonymy and manner-without.

## 6.3 Polarity Propagation through Selected Relations

After selecting the adequate relation types, the polarities assigned in the first step were propagated until every synset, connected directly or indirectly through one of the five selected types, had been reached. The algorithm ran for eight iterations and then stopped, with 10,318 polarised synsets. This number is lower than if all the relations in Table 4 are used (43,468), but we preferred to have a smaller but more reliable polarity lexicon. We can however, in the future, generate larger polarity lexicons, in a trade-off for lower reliability.

Table 5 has two examples of polarity propagation from the initial assignment until iteration 3, through different semantic relations. In the same table, 'Iter' is the iteration number, or 0 for the initial assignment, and 'Pol' is the propagated polarity.

# 7 Overall Evaluation

In order to complement the evaluation of the generated lexicon, we performed one last evaluation, where the polarity of 500 synsets, polarised in iterations 1 to 8, was compared, once again, with the polarity given manually by two human judges. The results of this evaluation are shown in Table 6, all together, and in Table 7, according to the iteration and starting with the evaluation of the 390 synsets polarised in the initial assignment (same as Table 2). As expected, accuracy becomes lower for higher iterations. The agreement becomes

**Table 5** Examples of initial assignment (0) and propagation.

| Iter | Pol | Propagation |
|---|---|---|
| 0 | - | **(adj)** *incorrigível* (-), *destravancado*, *irregenerável*, *indisciplinável*, *insubordinável*<br>(*incorrigible* (-), *unregenerable*, *undisciplinable*, *rebellious*) |
| 1 | - | **Has-quality**<br>→ **(n)** *incorrigibilidade*, *irreparabilidade*<br>(*incorrigibility*, *irreparability*) |
| 2 | - | **Quality-of**<br>→ **(adj)** *irrecuperável*, *irremediável*, *incompensável*, *irreparável*, *insubstituível*, *insuprível*<br>(*irrecoverable*, *irremediable*, *not_compensable*, *irreparable*, *irreplaceable*) |
| 3 | - | **Has-manner**<br>→ **(adv)** *irreparavelmente*<br>(*irreparably*) |
| 0 | + | **(n)** *concordância*, *consentimento*, *autorização*, *beneplácito*, *tolerância* (+), *permissão*, *licença*<br>(*agreement*, *consent*, *permission*, *tolerance* (+)) |
| 1 | + | **Has-manner**<br>→ **(n)** *outorgadamente*<br>(*consentingly*) |
| 2 | + | **Manner-of**<br>→ **(adj)** *deferido*, *outorgado*, *concedido*<br>(*deferred*, *granted*) |
| 3 | - | **Antonym-of**<br>→ **(adj)** *impedido*, *tolhido*, *vedado*, *proscrito*, *negado*, *defeso*, *interdito*, *proibido*, *inconcesso*<br>(*prevented*, *fenced*, *denied*, *forbidden*, *prohibited*) |

**Table 6** Evaluation of iterations 1 to 8.

| Sample | Ref | Target | P(+) | P(-) | P(0) | P(all) | Kappa |
|---|---|---|---|---|---|---|---|
| 500 sets | H1 | Aut | 0.65 | 0.66 | 0.38 | 0.63 | 0.42 |
| 500 sets | H2 | Aut | 0.68 | 0.65 | 0.42 | 0.64 | 0.43 |
| 500 sets | H1 | H2 | 0.75 | 0.84 | 0.62 | 0.75 | 0.62 |

lower as well for higher iterations. Nevertheless, until iteration 4, it is always higher than 0.54. These results also suggest that, for more reliable results, the algorithm should stop before there are no more reachable synsets, for instance, in iteration 2, or maybe 4. Also in Table 7, we present the combined accuracy for the whole lexicon, which is about 78.9% or 70% using H1 or H2 respectively as reference. This value considers the number of synsets polarised in the initial step, polarised through propagation, and the accuracy of those steps.

For the analysis of these results, it is worth reminding that Onto.PT is a resource created automatically and in development. Consequently, it is not 100% reliable. For instance, it contains a few incorrect relations, which have a negative impact on the propagation results. On the other hand, Onto.PT tends to keep growing and to improve its reliability, which will consequently have a positive impact on future polarity lexicons generated by the proposed approach. In fact, there are more recent version of this resource, released after 0.3.

## 8 Concluding Remarks

We have described an approach for moving from the polarity of single lemmas to the polarity of concepts/meanings, represented as wordnet synsets. This kind of approach is an alternative to tedious and time-consuming annotation tasks, performed by humans.

**Table 7** Evaluation and quantity of synsets according to iteration (0 to 5).

| Iteration | Synsets | Sample | Ref | P(all) | Kappa |
|---|---|---|---|---|---|
| 0 | 7,556 | 390 | H1<br>H2 | 0.86<br>0.73 | 0.66 |
| 1 | 1,971 | 330 | H1<br>H2 | 0.71<br>0.72 | 0.63 |
| 2 | 549 | 112 | H1<br>H2 | 0.48<br>0.50 | 0.57 |
| 3 | 163 | 38 | H1<br>H2 | 0.45<br>0.34 | 0.55 |
| 4 | 56 | 15 | H1<br>H2 | 0.60<br>0.60 | 0.54 |
| 5 | 16 | 4 | H1<br>H2 | 0.25<br>0.00 | 0.00 |
| ... | | | | | |
| **Total** | 10,318 | 890 | H1<br>H2 | 0.79<br>0.70 | 0.65 |

The proposed approach is language independent, but we have applied it to a Portuguese wordnet-like resource. Though relatively straightforward, the accuracy of the polarised synsets confirms that this approach is quite solid. In its first step, we have shown that synsets can be polarised according to the sum of the polarity of their lemmas alone, with accuracies close to 90%. In the second step, polarities were propagated for several iterations, through a set of selected relation types. For selecting the relations to use, we measured the accuracy of different types for this task and observed that relations connecting adjectives or adverbs to other synsets were more suitable for this. We also concluded that the accuracy of polarity attribution decreased for higher iterations.

In the end, we obtained a polarity lexicon for Portuguese derived from Onto.PT, with 10,318 polarised synsets, and polarities between 70% and 79% accurate, depending on the judge. As far as we know, while writing this paper, the resulting polarity lexicon was the first synset-based resource of this kind targeting Portuguese. Though, very recently, we became aware that, taking advantage of the alignment between OpenWordNet.PT and Princeton WordNet, SentiWordNet polarities have been assigned to OpenWordNet.PT synsets [22]. Although this effort might suffer from issues regarding the translation of lexicons from one language to another (different languages represent different socio-cultural realities, they do not cover exactly the same part of the lexicon and, even where they seem to be common, several concepts are lexicalised differently [10]), it is another relevant contribution for the development of Portuguese SA applications. We should recall that, if combined with WSD techniques, synset-based polarity lexicons will improve the attribution of a polarity to words in context, and thus the polarity classification of various kinds of text. Therefore, following the free availability of Onto.PT, the result of assigning polarities to a more recent version of Onto.PT is available from this project's website (`http://ontopt.dei.uc.pt`).

Even though the obtained results are interesting, additional work is needed, in order to get a more reliable resource. Therefore, we end by leaving some lines for further work. First, looking at the obtained results, it might be a good idea to decrement polarity strength in each propagation iteration. This would enable the algorithm to stop either when there are no more reachable synsets or when the propagated polarity is below some threshold. Second, more than adequate relation types for polarity propagation, we believe that there are combinations of types that lead to good polarity propagation, and combinations that don't. It would be interesting to learn these combinations, which can be seen as paths (e.g. A *hyponym-of* B *has-quality* C), semi-automatically, by selecting paths between synsets with

correct polarities. Finally, we believe that polarity represented by three parameters (positive, negative and neutral) is limitative and should be rethought. For instance, instead of being represented by only one value, polarity can be represented by a real value from 0 to 1, or by three parameters, respectively indicating the positive, negative and neutral strength, in a similar fashion of what is done in SentiWordNet. This could be achieved by propagating polarity using a method as PageRank, in a similar fashion to existing work for English [4]. If the polarity values obtained this way are normalised, it will be possible to compare them to the results obtained with the approach proposed in this article.

**Acknowledgements.** This work was supported by the iCIS project (CENTRO-07-ST24-FEDER-002003), co-financed by QREN, in the scope of the Mais Centro Program and European Union's FEDER.

## References

1 Rodrigo Agerri and Ana García-Serrano. Q-wordnet: Extracting polarity from WordNet senses. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, LREC 2010, La Valletta, Malta, 2010. ELRA.

2 A. R. Balamurali, Aditya Joshi, and Pushpak Bhattacharyya. Harnessing wordnet senses for supervised sentiment classification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, EMNLP 2011, pages 1081–1091, Edinburgh, Scotland, UK, 2011. ACL Press.

3 Andrea Esuli and Fabrizio Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, LREC 2006, pages 417–422, 2006.

4 Andrea Esuli and Fabrizio Sebastiani. PageRanking WordNet synsets: An application to opinion mining. In *Proceedings of 45th Annual Meeting of the Association for Computational Linguistics*, ACL'07, pages 424–431, Prague, Czech Republic, 2007. ACL Press.

5 Christiane Fellbaum, editor. *WordNet: an electronic lexical database (Language, Speech, and Communication)*. The MIT Press, 1998.

6 Cláudia Freitas. Sobre a construcção de um léxico da afetividade para o processamento computacional do português. *Revista Brasileira de Linguística Aplicada*, 13(4):1013–1059, 2013.

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

8 Annette M. Green. Kappa statistics for multiple raters using categorical classifications. In *Proceedings of the 22nd Annual Conference of SAS Users Group*, San Diego, USA, 1997.

9 Vasileios Hatzivassiloglou and Kathleen R. Mckeown. Predicting the semantic orientation of adjectives. In *Proceedings of 35th Annual Meeting of the Association for Computational Linguistics*, ACL 1997, pages 174–181, Madrid, ES, 1997. ACL Press.

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

11 Minqing Hu and Bing Liu. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA, 2004. ACM.

12 Nobuhiro Kaji and Masaru Kitsuregawa. Building lexicon for sentiment analysis from massive collection of HTML documents. In *Proceedings of the Joint Conference on Empir-*

*ical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL 2007, pages 1075–1083, 2007.

13 Jaap Kamps, Robert J. Mokken, Maarten Marx, and Maarten de Rijke. Using WordNet to measure semantic orientation of adjectives. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume IV of *LREC 2004*, pages 1115–1118, Paris, France, 2004. ELRA.

14 Hiroshi Kanayama and Tetsuya Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 355–363, Sydney, Australia, 2006. ACL Press.

15 Jungi Kim, Hun-Young Jung, Yeha Lee, and Yeha Lee. Conveying subjectivity of a lexicon of one language into another using a bilingual dictionary and a link analysis algorithm. *International Journal of Computer Processing Of Languages*, 22(02-03):205–218, 2009.

16 Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, COLING 2004, pages 1267–1373, Geneva, Switzerland, 2004. ACL Press.

17 Isa Maks and Piek Vossen. Different approaches to automatic polarity annotation at synset level. In *Proceedings of the 1st International Workshop on Lexical Resources*, WoLeR'11, Ljubljana, Slovenia, 2011.

18 Erick G. Maziero, Thiago A. S. Pardo, Ariani Di Felippo, and Bento C. Dias-da-Silva. A base de dados lexical e a interface web do TeP 2.0 – Thesaurus Eletrônico para o Português do Brasil. In *VI Workshop em Tecnologia da Informação e da Linguagem Humana (TIL)*, pages 390–392, 2008.

19 Roberto Navigli. Word sense disambiguation: A survey. *ACM Computing Surveys*, 41(2):1–69, 2009.

20 Bo Pang and Lillian Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, January 2008.

21 António Paulo-Santos, Hugo Gonçalo Oliveira, Carlos Ramos, and Nuno C. Marques. A bootstrapping algorithm for learning the polarity of words. In *Proceedings of Computational Processing of the Portuguese Language – 10th International Conference (PROPOR 2012)*, volume 7243 of *LNCS*, pages 229–234, Coimbra, Portugal., April 2012. Springer.

22 Alexandre Rademaker, Valeria De Paiva, Livy Maria Real Gerard de Melo, Coelho, and Maira Gatti. Openwordnet-pt: A project report. In *Proceedings of the 7th Global WordNet Conference*, GWC 2014, pages 383–390, Tartu, Estonia, jan 2014.

23 Delip Rao and Deepak Ravichandran. Semi-supervised polarity lexicon induction. In *Proceedings of 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL 2009, pages 675–682, Athens, Greece, 2009. ACL Press.

24 Mário J. Silva, Paula Carvalho, and Luís Sarmento. Building a sentiment lexicon for social judgement mining. In *Proceedings of 10th International Conference on Computational Processing of Portuguese, PROPOR 2012*, LNCS/LNAI, Coimbra, Portugal, April 2012. Springer.

25 Marlo Souza, Renata Vieira, Debora Busetti, Rove Chishman, and Isa Mara Alves. Construction of a portuguese opinion lexicon from multiple resources. In *Proceedings of 8th Brazilian Symposium in Information and Human Language Technology*, STIL 2011, Cuiabá, Brazil, 2011.

26 Peter D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL 2002, pages 417–424, Philadelphia, PA, USA, 2002. ACL Press.