Automatic Creation of a Reference Corpus for Political Opinion Mining in User-Generated Content

Luís Sarmento
University of Porto, Faculty of Engineering, - DEI - LIACC
Porto, Portugal
las@fe.up.pt

Paula Carvalho
University of Lisbon, Faculty of Sciences, LASIGE
Lisboa, Portugal
pcc@di.fc.ul.pt

Mário J. Silva
University of Lisbon, Faculty of Sciences, LASIGE
Lisboa, Portugal
mjs@di.fc.ul.pt

Eugénio de Oliveira
University of Porto, Faculty of Engineering - DEI - LIACC
Porto, Portugal
eco@fe.up.pt

ABSTRACT

We propose and evaluate a method for automatically creating a reference corpus for training text classification procedures for mining political opinions in user-generated content. The process starts by compiling a collection of highly opinionated comments posted by users on an on-line newspaper. Then, we define and use a set of manually-crafted high-precision rules supported by a large sentiment-lexicon in order to identify sentences in each comment expressing opinions about political entities. Finally, the opinions found are propagated to the remainder sentences of the comment mentioning the same entities, thus increasing the number and variety of opinion-bearing sentences. Results show that most of the rules can identify negative opinions with very high precision, and these can be safely propagated to the remainder sentences in the comment in almost 100% of the cases. Due to problems arising from irony, the precision of identification drops for positive opinions, but several rules still reach high precision. Propagation of positive opinions is correct in about 77% of the cases, and most errors at this stage result from irony and polarity inversion throughout the comment.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing—*Linguistic processing*

General Terms

Algorithms, Design, Measurement

Keywords

opinion mining, user-generated content, reference corpus

1. INTRODUCTION

Recently, there has been much interest in developing tools for performing automatic opinion mining on the web [8], and more specifically in user-generated content (UGC), including web blogs, microblogs (e.g. Twitter) and user posts to media and consumer sites. This type of content is particularly interesting for opinion mining tasks because of the huge density of opinions expressed about a large variety of issues. However, mining UGC is particularly challenging due to the presence of multiple idiosyncrasies, such as spelling mistakes, ungrammatical constructions, absence of capitalization, abbreviations and neologisms. Linguistic-based approaches (e.g. use of lexico-syntactic rules) tend to be inefficient for mining such type of texts, especially in terms of *recall*. They aim at exhaustively capturing grammatical constructions rather than "deviant" forms, which are extremely frequent in UGC. To overcome recall problems, researchers have been using automated text classification procedures, capable of capturing relevant opinion information from a large variety of more or less complex features that can be extracted from text. However, training text classifiers requires annotated corpora, which are difficult and costly to create manually [1]. Additionally, once created for a particular text genre or semantic domain, these corpora are generally not applicable to train classifiers for detecting opinions in other text genres or domains (e.g. movie reviews vs. political debates).

Following previous work [11], in this paper we introduce an automatic method for creating (and quickly updating) a reference corpus for political opinion mining in UGC. We propose a three-step approach. First, we collect a large number opinionated comments posted by online readers of a popular on-line newspaper. Second, we apply a small library of hand-crafted lexico-syntactic opinion detection rules to those comments, in order to identify sentences or phrases containing explicit positive or negative opinions about a set of relevant political entities. Finally, we automatically propagate the opinions identified by the manually-developed rules to the remaining sentences of the comment, in order to increase the number and diversity of sentences conveying opinions, including sentences exhibiting some of the previously mentioned deviant forms or idiosyncrasies (non-grammatical phrases, out-of-lexicon words, abundant ironic constructions, etc.). Therefore, this work answers the two following questions:

1. Do lexico-syntactic opinion detection rules achieve a good precision in detecting both positive and negative opinions about political entities?

2. Can we safely propagate the opinion found by these rules to other sentences of the user comments?

Positive answer to both questions validates the proposed approach for automatically creating a reference corpus for training text classification methods.

The remainder of this paper is organized as follows. In the next section, we revise related work. In Section 3, we describe the *sentiment lexicon* and the small set of *opinion detection rules* developed for identifying and classifying opinionated constructions. The experiments conducted are described in Section 4. Results are presented and discussed in Section 5. We conclude this paper by highlighting the main results obtained and by pointing out directions for further research.

2. RELATED WORK

The problem of developing reference corpora for training and testing sentiment classifiers has been receiving some attention. MPQA is one of the most important corpora available for sentiment analysis in English [14]. It contains about 10,000 sentences collected from world press articles, whose *private states* (opinions, emotions, sentiments, speculations, evaluations, etc.) were manually annotated and carefully revised. A key point of this work is that annotations are performed at word and phrase level, rather than at sentence (e.g. [15]) or document level (e.g. [9], [12]), as commonly done. Also, annotation was carried out taking into account the context, which is crucial for resolving possible ambiguities and accurately identifying polarity (as already demonstrated by Yu and Hatzivassiloglou [15]). All private-state expressions identified in the MPQA corpus were associated to a fine-grained annotation scheme that includes the source of the private-state, the target involved and specific properties, such as intensity, significance and type of attitude.

A different approach for obtaining a reference corpus is followed by Pang et al. [9]. The authors selected a collection of movie reviews where user ratings were explicitly expressed (e.g. “4 stars”), and automatically converted them into positive, negative or neutral classifications. This corpus is comprised of a total of 753 negative and 1,301 positive reviews, which were used for training Support Vector Machines (SVM) with word features, in order to classify reviews. The results obtained were considered good in relation to a set of human-generated baselines. Nevertheless, such approach requires that each opinionated text is associated to a rating value, which is not the case of the texts we collected and analyzed in the scope of this study. On the other hand, the corpus is domain-dependent (or even community-dependent), which means that it is inadequate for training opinion mining systems that deal with texts from other genres or domains. Also, the unit of analysis considered in this corpus is the document, making it impossible to account for the frequent polarity variations that can be found throughout the document.

Franky and Manurung tried to repeat the experiments of Pang et al. for the Indonesian language [5]. Due to the absence of training corpora in this language, they applied *machine translation* tools for translating the original English corpus into Indonesian, and then used the translated version to train the classifiers. Several machine translation options – ranging from commercial tools to simple word-by-word translation – and text classification methods were experimented. Notably, the best results obtained were similar to those obtained in the original experiment.

Several completely different approaches have been proposed to circumvent the difficulties related to building larger and less domain-dependent corpora, both for English and other languages. Wiebe and Riloff propose a rule-based method for automatically generating a corpus for training *subjective* and *objective* sentence classifiers [13]. Non-annotated sentences were automatically classified as *subjective*, if containing two or more of the *strong subjective clues* previously specified by the authors, or *objective*, whenever they do not present any of those clues. The rule-based classifier was found to have a relatively high precision in identifying subjective sentences ($\simeq 90\%$) and slightly lower in identifying objective sentences ($\simeq 80\%$), while recall ranged from 30-34%. The authors proceeded in creating an *initial training set* with about 100,000 subjective and objective sentences found in English articles from a world press collection based on such rule-based classifiers. In a next step, they used the initial training set for learning additional patterns capable of identifying subjective and objective sentences. Adding such patterns to the rule-based classifier allowed an increase in recall, at a cost of a slight drop in precision (but still over 80%). An expanded training set was then automatically created using a *self-training* procedure around a naive-Bayes classifier. The procedure starts by applying the initial training set for learning patterns, which are then used for finding new subjective and objective sentences in the non-annotated corpus. The new data is then used to train the naive-Bayes classifiers, which in turn are applied for classifying *all* sentences in the corpus. This much larger set of subjective and objective sentences is then used as the starting point for repeating the entire process (extracting patterns, increasing training set and training naive-Bayes classifier). According to the authors, this procedure results in a large increase in classification recall without significantly affecting precision, leading to the highest F-measure when compared to the combination of rule-based classification augmented with extracted patterns.

Devitt and Ahmad argue that attempts to build reference corpus for opinion mining so far are either limited by size constraints, due to the amount of human annotation required, or by the specificity of the domain, making them inadequate for training robust and wide-scope opinion classifiers [1]. Thus, they propose to use extrinsic data sources in order to overcome such limitations. For example, matching news about the stock market with the stock index for long periods is an inexpensive way to build a large corpus of sentiment-bearing news, with objective numeric data “expressing” such sentiment: when the stock index is falling, news about the stock market are mostly negative (and vice-versa). This automatically aligned data can then be used for training opinion classifiers and price “predictors”, given the current news. However, experiments by Généreux et al. in using such strategy to predict changes in stock prices were rather disappointing [7].

Our strategy for building a reference corpus is, in a certain way, similar to the one by Wiebe and Riloff, but we are focused on subjective opinions and in distinguishing them in terms of polarity, i.e. *positive* vs. *negative*. We also take the sentence as the minimum unit of analysis and classification, but unlike their approach, opinionated text is restricted to a set of relevant political entities (that we also aim at identifying). Regarding the linguistic patterns used, we explored a very restricted range of predicative constructions, which do not take into account information concerning constituent delimitation and classification. This option also differs from Riloff and Wiebe, where a very large set of syntactic constructions identified by a shallow parser are exploited [10]. Our approach is directly motivated by the challenges related to *user-generated content*, which contains multiple idiosyncrasies that raise several problems to traditional parsing technologies.

3. LINGUISTIC RESOURCES

In this section, we describe the linguistic resources used for developing high-precision *opinion identification rules*, namely a sentiment lexicon and a small library of syntactic-semantic patterns where polarity-bearing adjectives or nouns may occur. The resources described next were developed for Portuguese, which is the target language of the current work.

3.1 Sentiment lexicon

We started by exploring the syntactic and semantic information available in previously developed linguistic resources [2]. These comprise 4,250 intransitive adjectives, and 3,554 morpho-syntactically associated nouns (e.g. estúpido/estupidez; stupid/stupidity). Both predicates[1] are characterized for co-occurring with human subjects and for having no complements.

We have then selected the possible polar adjective and nouns from these resources, and manually classified each predicate according to its predictable polarity, which may be 0, 1, or -1. These codes represent a neutral, positive or negative semantic orientation, respectively. The sentiment lexicon has also been enriched with some new entries, collected from diverse corpora. At the present, it is composed by 6,055 entries: 3,533 adjectives and 2,522 nouns. In terms of polarity distribution, 55,5% of the entries were classified as negative, 21,8% as positive and the remaining 22,7% as neutral. These numbers show a considerable discrepancy between the negative and positive entries in the lexicon, but no polarity constraints were imposed when selecting lexical data.

3.2 Lexico-syntactic patterns

We manually developed a set of syntactic-semantic patterns, describing typical elementary constructions related to the expression of opinion about named entities (NE). We selected a set of four *base patterns*, whose nuclear element can be an adjective or a predicate noun comprised in the sentiment lexicon (represented below, respectively, as $A^{-/+}$ and $N^{-/+}$ to denote either negative or positive polarity). The chosen base patterns are:

R1 [NE] [V_{cop}] [Art_{ind}] [N_{hum}] [$A^{-/+}$] (e.g. "Sócrates é um político traiçoeiro" / "Sócrates is a treacherous politician")

R2 [NE] [V_{ser}] [Art_{ind}] [($A^{-/+}$|$N^{-/+}$)] (e.g. "Menezes é um oportunista", "MFL é uma desgraça" / "Menezes is an opportunist", "MFL is a disgrace")

R3 [NE] [V_{cop}] [$A^{-/+}$] (e.g. "Sócrates está nervoso" / "Sócrates is nervous")

R4 [Art_{def}] [$A^{-/+}$] [(do|da|dos|das)] [NE] (e.g. "O aldrabão do Sócrates" / "The liar of Sócrates")

Adjectives and nouns represented in R1-R3 are related with their subject, a NE, through a *copula verb* (V_{cop}) or a *support verb* (V_{sup}), respectively. In R4, the adjective occurs in pre-nominal position, and relates with the noun it modifies through the preposition "de" (of). We have also taken into consideration the presence of optional modifiers in some of these syntactic constructions, such as *adverbs of negation*, *intensifiers* and other *adverbs*, which may influence the polarity of the construction. We assigned a reference code to each pattern, which comprises information regarding (i) the base pattern (R1 – R4), (ii) the inclusion of optional adverbs that do not affect the polarity information assigned to base pattern and (iii) the presence of an adverb of negation, which reverses the polarity information assigned to base pattern. For example, R1A is a variation of R1, where adjective is preceded by an intensifier adverb, while R2BN is a variation of R2, where predicate noun is modified by an adverb of *negation*. In this study, we used a total of 17 patterns, including the base patterns above described and corresponding variations.

The *opinion detection rules* are made by associating each of these patterns with a *polarity assignment rule*, which takes into account the polarity of the adjectives and nouns in the pattern – $A^{-/+}$ and $N^{-/+}$ – as described in the sentiment lexicon, and the presence of negative adverbs that invert the polarity of the overall construction.

4. EXPERIMENTAL SET-UP

We collected opinionated user posts from the web site of one of the most popular Portuguese newspapers. We obtained a collection of 8,211 news and the corresponding comments posted by on-line readers. The collection covers a period of five months (November 2008 to March 2009) and includes about 250,000 user posts, totaling approximately one million (simple and complex) sentences. On average, user comments have about four sentences. We can thus hypothesize that, in this collection, direct application of our method can potentially find on average *three* new opinionated sentences for each one matching a high-precision pattern.

We also compiled a dictionary containing names of frequently mentioned politicians for instantiating the patterns described in Section 3.2. This dictionary was obtained by identifying names in RSS news feeds, compiled over the same five month period. The dictionary contains 1,226 names that are found in 129,251 sentences of comment collection. We

[1]By predicate, we mean the nuclear element of each elementary sentence, which can be either a verb, an adjective or a noun. Following the Lexicon-Grammar Theory [6], elementary sentences correspond to the basic unit of analysis and meaning, which are constituted by a predicate and its mandatory arguments.

run the opinion detection rules over the set of these sentences, in order to find opinion-bearing sentences. For each rule, we manually checked if the matched constructions actually carry an opinion about a specific entity, as predicted by the rule. At this stage, evaluation was made by only considering the candidate sentence matched, disregarding the rest of the text in the comment. Errors were classified according to the following categories:

E1 the predicate (i.e. $A^{-/+}$ or $N^{-/+}$) is syntactically or semantically ambiguous (e.g. predicates can be both intransitive and transitive, exhibiting different meanings and polarities according to their transitivity);

E2 NE is an element of a complex noun phrase (e.g. NE are included in a coordinating or possessive construction);

E3 errors arising from the inability to detect dependencies between clauses in the sentence (e.g. the recognized construction is an argument of another predicate in the sentence, conditioning their polarity);

E4 the construction, or the possible sentence that includes it, is ironic;

E5 opinion interpretation presupposes knowledge of the world (such as references to other entities) and cannot be interpreted only by the information explicitly mentioned in text.

E6 other errors.

In a second evaluation step, we selected the best performing pattern variations in order to verify if the opinion expressed in the sentence matching the high-precision rule could be propagated to the remaining sentences of the comment. In this second stage, we only considered the comments where the candidate sentence was correctly classified in terms of polarity. Propagation results were evaluated according to the following possibilities:

P^+ opinion found can be propagated to other clauses or sentences of the comment that mention the same entity;

P^- opinion found *cannot* be propagated to other clauses or sentences of the comment that mention the same entity, since they express at least an opposite opinion about that entity;

P^0 opinion found cannot be propagated to the other clauses or sentences of the comment because the entity at stake is only mentioned once (nevertheless the opinion found *is not inconsistent* with the remaining text of the comment);

P^I out of context, opinion found seems to be correctly classified, but the remaining text gives evidence that the matching construction is used ironically;

P^A the remainder of the comment does not allow deciding whether the recognized opinion can be extendable to the other phrases or sentences, due to ambiguity or vagueness in text.

5. RESULTS AND ANALYSIS

In this section we present the results regarding both stages of our experiment, namely the performance of *opinion detection* using manually developed rules and the precision of *opinion propagation* to the remainder sentences of the comment.

5.1 Opinion Detection

Table 1 presents the precision in opinion detection for all pattern variations applied to the test collection. We explicitly differentiate candidate sentences expressing negative opinions – Table 1 a) – and positive opinions – Table 1 b). The overall number of sentence matches with all patterns is 1,056. Not surprisingly, this number covers a small percentage (0.82%) of all possible potential matches in the collection (129,251 sentences). Also, only 12 of the 17 pattern variations applied to the collection matched at least one sentence.

There are two significant differences in the results obtained while detecting negative and positive opinions. First, the number of matched sentences expressing negative opinions (692) is almost the double of those expressing positive opinions (363); additionally, negative sentences instantiate a higher number of lexico-syntactic patterns (12 vs. 9). These differences may be due to the empirical evidence that readers commenting news seem to be more prone to express disagreement than support, regarding the issue at stake. Therefore, it is likely that the collection of comments we are exploring is unbalanced, containing considerably more (and more varied) negative opinions than positive opinions. Another factor contributing to this difference might be related to the higher number of entries in the sentiment lexicon with negative polarity, as mentioned in Section 3.1.

Second, the precision of identifying negative opinions ($\simeq$ 89%) is significantly higher than the precision of identifying positive opinions ($\simeq$ 60%). One reason for this discrepancy, which we will explore in the next section, is the frequent use of *irony* in comments: in many cases, a positive predicate is used to express a negative opinion. Interestingly, if we only consider the results obtained by using pattern variations that contain an *adverb of negation*, the precision in identifying positive opinions (i.e. by negating negative predicates) is higher than average, while the precision of identifying negative opinions (i.e. by negating positive predicates) drops below the average.

Looking at the results of individual rules, one can observe that some pattern variations – R1A, R1C, R2BN and R3AN – almost do not instantiate (support < 3.0%), both for negative and positive opinions. There are also patterns that seem to be dependent on polarity. For example, R4 applies to negative opinions, but not to positive ones. R2B (negative opinions) and R3B (positive opinions) correspond to the most productive rules. However, the precision of these rules is quite different: while precision is very high (93.7%) for negative opinions (R2B), for positive opinions (R3B) precision decreases significantly (52.5%). Results also demonstrate that the patterns obtaining a precision rate inferior to 0.9 in negative opinion identification are the most generic, regarding their syntactic structure (e.g. [NE] [V_{cop}] [$A^{-/+}$]).

5.2 Error Analysis for Opinion Detection

Table 2 details the errors found while detecting negative and positive opinions. Errors of types E1, E2 and E3 are common both to positive and negative opinions. As previ-

a) *negative* opinions

	#	# correct	precision %	support %
R1A	3	3	100	0.43
R1BN	1	1	100	0.14
R1B	87	83	95.4	12.6
R1C	2	2	100	0.28
R2A	1	0	0	0.14
R2B	191	179	93.7	27.6
R2BN	2	2	100	0.28
R3AN	4	4	100	0.58
R3A	39	33	84.6	5.6
R3BN	21	17	81.0	3.0
R3B	177	140	79.1	25.6
R4	164	151	92.1	23.7
$\sum$	692	615	88.9	100

b) *positive* opinions

	#	# correct	precision %	support %
R1A	9	7	77.8	2.5
R1B	105	81	77.1	28.9
R1C	3	2	66.7	0.82
R2B	40	14	35.0	11.0
R2BN	4	1	25.0	1.10
R3A	46	23	50.0	12.7
R3AN	4	4	100	1.10
R3B	122	64	52.5	33.6
R3BN	30	20	66.7	8.3
$\sum$	363	217	59.5	100

Table 1: Precision results for candidate sentences expressing *negative* and *positive* opinions.

a) *negative* opinions

	E1	E2	E3	E4	E5	E6
R1B	1	2	1	0	0	0
R2A	0	1	0	0	0	0
R2B	2	7	3	0	0	0
R3A	0	5	1	0	0	0
R3BN	1	3	0	0	0	0
R3B	5	23	8	1	0	0
R4	10	0	0	0	3	0
$\sum$	19	41	13	1	3	0
%	24.7	53.2	16.9	1.3	3.9	0

b) *positive* opinions

	E1	E2	E3	E4	E5	E6
R1A	2	0	0	0	0	0
R1B	5	2	3	8	1	5
R1C	0	0	1	0	0	0
R2B	2	3	4	13	1	3
R2BN	0	0	1	2	0	0
R3A	2	5	3	11	0	2
R3B	6	16	14	15	2	5
R3BN	1	1	5	3	0	0
$\sum$	18	27	31	52	4	15
%	12.2	18.4	21.1	35.4	2.7	10.2

Table 2: Errors for candidate sentences expressing *negative* and *positive* opinions.

ously mentioned, E1 concerns different kinds of lexical ambiguities, such as the one illustrated by the following example (EG1):

EG1 *"O objectivo do Sócrates..."* [2]

The lexical unit "objectivo" ("objective" / "goal") can be either a noun, which is neutral in terms of polarity, or an adjective, which is classified as positive in the sentiment lexicon. Both categories may potentially occur in a structure like EG1, but only the nominal analysis is adequate in the context of the sentence ("O objectivo do Sócrates é ser reeleito" / "The objective of Sócrates is to be reelected").

Errors marked as E2 and E3 are related to constituent identification and dependency, respectively, as illustrated in the examples below:

EG2 *"[Comparar Estaline ou Hitler a] Bush é tão absurdo...." / "[Comparing Stalin or Hitler to] Bush is so absurd. . . "*

EG3 *"[Agora inventaram que] Sócrates era um suspeito..." / "[Now they made up that] Sócrates was a suspect..."*

[2]In Portuguese, the illustrated construction can be interpreted as "The objective of Sócrates..." or "Sócrates, the objective,..."

In the example EG2, "Bush" was identified as the subject of "absurdo" ("absurd"), but it is a constituent of a comparative construction, which corresponds to the true subject of the sentence. In the example EG3, the construction "Sócrates era um suspeito" ("Sócrates was a suspect") was incorrectly classified as negative, since it is an argument of another predicate ("inventar" / "to make up"), whose meaning alters the interpretation of polarity of the argument. Such type of errors could be solved either by using a more sophisticated parser capable of performing constituent delimitation and resolving argument dependency, or by imposing finer constrains on the lexico-syntactic patterns so as to filter out such situations.

Most of the errors found regarding the detection of positive opinions are related to irony (E4). For example, the irony implicit in EG4, illustrated below, is reinforced by the unexpected use of the name "José de Sousa" for mentioning the Portuguese Prime-minister, instead of using his well-known name, "José Sócrates":

EG4 *"José de Sousa é um honesto cidadão" / "José de Sousa is an honest citizen"*

We also found some cases where the decision about opinion polarity depends on the knowledge of the world (E6), as illustrated by the example EG5:

	% P^+	% P^-	% P^0	% P^I	% P^A
R1B$^-$	79.5	0	20.5	0	0
R2B$^-$	68.4	1.2	28.6	0	1.8
R1B$^+$	28.4	12.3	46.9	8.7	3.7

Table 3: Precision in propagating the opinion identified by rules R1B$^-$, R2B$^-$ and R1B$^+$ to the remainder sentences of the comment.

EG5 *"Chavez era bom pra governar o BPP e o BPN!" / "Chavez would be a good manager for BBP and BPN!"*

For interpreting this construction, one needs to know that "BPP" and "BPN" are Portuguese banks that have recently been involved in financial scandals. This suggests that our semantic lexicon should also include names of people and organizations with a strong polarity bias.

5.3 Opinion Propagation

In order to evaluate the automatic opinion propagation, we selected the most successful rules (i.e. those exhibiting both significant support and high-precision in detecting opinions at sentence level) and manually verified if the opinion found in each case could be propagated to the rest of the sentences in the comment. We only evaluated the propagation of opinions previously found to be correct (at sentence level), and we followed the typology proposed in Section 4 (P^+, P^-, P^0, P^I and P^A). For negative opinions we chose rules R1B (83 correct cases / 12.6% support) and R2B (179 correct cases / 27.6% support), henceforth mentioned as R1B$^-$ and R2B$^-$, respectively. For positive opinions, we opted for Rule R1B (81 correct cases / 28.9% support), henceforth mentioned as R1B$^+$.

Results show that propagation of negative opinions seems to be less problematic than propagation of positive opinions (see Table 3). In fact, rule R1B$^-$ leads to almost perfect propagation, since the opinion found by this rule was always extensible to the remaining sentences of the comment when these mention the entity at stake. Such a situation occurs in about 80% of all cases. The remaining cases (20%) refer to situations where the entity is not mentioned again in the comment, but additional opinions conveyed (i.e. about other entities) are not inconsistent with the previously identified opinion.

Regarding rule R2B$^+$, the fraction of correct and valid propagation decreases to 68.4%. However, only 3.0% of the cases are actually related with incorrect propagation (1.2%) or ambiguity (1.8%). Again, the remainder 28.6% involves situations where there are no additional mentions to the entity at stake, but the other opinions found are not inconsistent.

For Rule R1B$^+$ the fraction of correct and valid propagations drops further to 28,4%. While there is still a significant percentage of cases (47%) in which the propagation does not occur because there are no additional mentions to the entity at stake (i.e. P^0), about 25% of the cases involve *incorrect* propagation. In 12.3% of the cases the other sentences in the comment express inconsistent opinions about the entity. In 3.7% of the cases the context does not allow to clearly decide if the detected opinion can be propagated to the remaining sentences in which the entity is mentioned. In 8.7% of the cases opinions could not be correctly propagated since, after analyzing the complete context, we could conclude that the opinion was, in fact, ironic.

Since we are testing propagation for the purpose of building a reference corpus for opinion mining, we might consider the cases where the entities were not mentioned again in the comment as a *valid* source of example sentences (i.e. P^0 cases). First, they do not contradict the previously found opinion. Second, mentions to other entities found it the remaining sentence of the comments might implicitly carry opinion information consistent with the previously found opinion. For example, by mentioning another entity (person or organization) users can in fact be supporting the previously identified opinion, even without making any more direct references to such entity. Therefore, if we assume all P^0 cases as valid, then the precision achieved in propagating negative opinions reaches almost 100% for R1B$^-$ and R2B$^-$ and 77.3% for R1B$^+$.

5.4 Error Analysis for Opinion Propagation

A more detailed analysis of the errors revealed that there are recurrent situations that are problematic for the propagation of opinions, especially in what concerns *positive* opinions. The first, is sudden *polarity inversion* somewhere along the comment. The typical situation is that of a reader starting his/her comment by expressing a mildly positive opinion about a certain entity, usually regarding a secondary or lateral issue, and then expressing a negative opinion about that entity on a more specific issue. A similar situation for the case of movie reviews is reported by Pang et al. [9].

The second is related to the inability to detect irony in the initial opinion. What is apparently a positive sentence (usually the starting sentence in the comment) is in fact an ironic opinion, and the remainder of the comment then continues by expressing negative opinions. We noticed that ironic opinions frequently include *very positive* adjectives or nouns, such as for example "O José Sócrates é [fabuloso]." ("José Sócrates is [fabulous]."). These very strong adjectives or nouns might be a clue for detecting potentially ironic opinions.

6. CONCLUSION AND FUTURE WORK

The experiments conducted in this study show that we can positively answer the two questions posed in Section 1. The evaluation results show that the performance of lexico-syntactic rules depends on the polarity conveyed in the recognized text. In case of negative opinion detection, the precision rate reached about 90% while for positive opinionated text the precision decreased to 60%. Our experiments also demonstrate that it is possible to propagate the opinion found by the lexico-syntactic patterns to other sentences of the user comments mentioning the same entity, increasing the number and diversity of annotated sentences. Again, propagation success is higher for negative opinions (almost 100%) than for positive opinions (around 77%).

Error analysis showed that there is room for making simple improvements, both at the detection and propagation stages. Improving opinion detection will involve the enlargement of the sentiment lexicon, namely by annotating sentiment verbs, and some common and proper names. We expect to achieve this by following a bootstrapping approach [4]. We also intend to enhance the lexico-syntactic rules, by refining the constraints on the lexico-syntactic patterns and describing other relevant syntactic patterns. Additionally, we wish to develop mechanisms for detecting polarity inversion along the comment, so as to inhibit propagation when such situations are found. One idea for this consists in trying to detect the point in the comment that divides it in two sections that exhibit the maximum difference between the ratio of positive and negative predicates (e.g. many positive predicates in the first half of the comment vs. many negative predicates in the second half). Special attention will be given to treating the problems related to irony. The results of a first experiment conducted on automatic irony detection are presented in another paper [3].

We believe that, given the increasing amount of user-generated content, the approach we proposed in this paper is a valid and promising option for easily building a reference corpus for the purpose of training opinion classifiers targeting such content.

7. ACKNOWLEDGEMENTS

Work partially supported by grants SFRH/BD/23590/2005 and SFRH/BPD/45416/2008 from FCT (Portuguese research funding agency). We also thank FCT for its LASIGE and LIACC multi-annual support.

8. REFERENCES

[1] A. Devitt and K. Ahmad. Sentiment analysis and the use of extrinsic datasets in evaluation. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA).

[2] P. Carvalho. *Análise e representação de construções adjectivais para processamento automático de texto. Adjectivos Intransitivos Humanos*. PhD thesis, University of Lisbon, 2007.

[3] P. Carvalho, L. Sarmento, M. J. Silva, and E. de Oliveira. Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-). In *TSA'09 - 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*, Hong Kong, Nov. 6 2009.

[4] A. Esuli and F. Sebastiani. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT.*, 2006.

[5] Franky and R. Manurung. Machine learning-based sentiment analysis of automatic indonesian translations of english movie reviews. In *Proceedings of the International Conference on Advanced Computational Intelligence and Its Applications 2008 (ICACIA 2008)*, Depok, Indonesia, 1-2 September 2008 2008.

[6] M. Gross. Les bases empiriques de la notion de prédicat sémantique. *Langages*, (63):7–51, Sept. 1981.

[7] M. Généreux, T. Poibeau, and M. Koppel. Sentiment analysis using automatically labelled financial news. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA).

[8] B. Pang and L. Lee. *Opinion Mining and Sentiment Analysis.* Now Publishers Inc, July 2008.

[9] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86, Morristown, NJ, USA, 2002. Association for Computational Linguistics.

[10] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of EMNLP-03, 8th Conference on Empirical Methods in Natural Language Processing*, Sapporo, JP, 2003.

[11] M. J. Silva, P. Carvalho, L. Sarmento, E. de Oliveira. and P. Magalhães. The Design of OPTIMISM, an Opinion Mining System for Portuguese Politics. In *Proceedings EPIA'09, 14th Portuguese Conference on Artificial Intelligence (local proceedings)*, Aveiro, Portugal, 2009.

[12] M. Thomas, B. Pang, and L. Lee. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 327–335, 2006.

[13] J. Wiebe and E. Riloff. Creating subjective and objective sentence classifiers from unannotated texts. In *Proceeding of CICLing-05, International Conference on Intelligent Text Processing and Computational Linguistics.*, volume 3406 of *Lecture Notes in Computer Science*, pages 475–486, Mexico City, MX, 2005. Springer-Verlag.

[14] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210, 2005.

[15] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136, Morristown, NJ, USA, 2003. Association for Computational Linguistics.