Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-)

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

Luís Sarmento
University of Porto, Faculty of Engineering, - DEI - LIACC
Porto, Portugal
las@fe.up.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 investigate the accuracy of a set of surface patterns in identifying ironic sentences in comments submitted by users to an on-line newspaper. The initial focus is on identifying irony in sentences containing *positive* predicates since these sentences are more exposed to irony, making their true polarity harder to recognize. We show that it is possible to find ironic sentences with relatively high precision (from 45% to 85%) by exploring certain *oral* or *gestural* clues in user comments, such as emoticons, onomatopoeic expressions for laughter, heavy punctuation marks, quotation marks and *positive* interjections. We also demonstrate that clues based on deeper linguistic information are relatively inefficient in capturing irony in user-generated content, which points to the need for exploring additional types of *oral* clues.

Categories and Subject Descriptors

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

General Terms

Design, Measurement

Keywords

irony detection, opinion mining, user-generated content

1. INTRODUCTION

In another paper, we propose a method based on a small set of manually crafted rules for automatically creating a reference corpus for opinion mining in user-generated content

(UGC) [6]. The reported results show that such polarity detection rules are able to identify *negative* opinions with relatively high precision (approximately 90%), but performance in detecting *positive* opinions is much lower (around 60% precision). When finding positive opinions, we observed that one of the major sources of error (about 35% of the cases) was related to *verbal irony*. Verbal irony is classically defined as the rhetorical process of intentionally using words or expressions for uttering a meaning different (usually the opposite) from the one they have when used literally. Nevertheless, this generic definition may be instantiated in slightly different ways depending on the perspectives and frameworks adopted (see, among others, [8] and [3]). Following Gibbs, verbal irony can be expressed by a variety of figurative devices, like sarcasm, hyperbole, rhetorical questions, jocularity, among other strategies, whose differences may be quite difficult to distinguish in practice [3].

In this paper, we adopt the term *irony* for referring to the specific case where a word or expression with prior *positive* polarity is figuratively used for expressing a *negative* opinion. We explore a set of relatively simple linguistic clues associated with the expression of irony in Portuguese. Our main goal is to investigate the productivity and accuracy of such clues in detecting irony in UGC. We intend to use such rules (in combination with previously developed opinion detection rules) to speed-up the construction of a reference corpus for opinion-mining which takes into account irony.

2. RELATED WORK

Despite some approaches to provide computational formulations of irony (e.g. [7]), there has been little work on the automatic detection of irony in text. However, there have been some attempts to tackle closely related problems, such as the detection of humorous text, hostile messages or, more generally, non-literal use of language. For example, Birke and Sarkar present a method for creating a corpus annotated with information regarding the *literal* and *non-literal* usage of verbs [1]. In a first stage, they use a weakly-supervised method to separate literal from non-literal usages. The operation is supported by two seed sets that contain examples representing both situations: the *literal feedback seed set* contains data from the Wall Street Journal (WSJ) and

the *non-literal feedback seed set* is composed of idiomatic and metaphoric expressions taken from dictionaries. Then, for a given sentence containing a verb to be tested, a word-based comparison is performed against all the sentences of each feedback set. The sentence is classified as either literal or non-literal, taking into account the set in which the most similar sentence was found. In a second stage, the method is improved by using an active learning strategy.

Mihalcea and Strapparava present an approach for automatically detecting humorous *one-liners* (i.e. short sentences) [5]. The authors started by building a corpus containing 16,000 humorous one-liners and an equal number of Reuters titles, proverbs from a on-line proverb collection, British National Corpus (BNC) sentences with word similar in content to the humorous one-liners, and sentences from the Open Mind Common Sense (OMCS) collection. Then several classification experiments where made using (i) humor specific stylistic features, such as the presence of alliteration, lexical antonymy and adult slang, (ii) content-based features (unigrams) and (iii) a combination of both features. Results show that a classification tree based on humor specific features is capable of differentiating one-liners from Reuters titles and BNC sentences, but does not separate one-liners from proverbs and OMCS sentences. Content based classification using Support Vector Machines and Naive-Bayes classifiers shows that is possible to clearly differentiate one-liners from all other types of sentences (except BNC sentences, which were chosen for being content similar to the one-liners). The combination of features provided marginal or no improvement. Performance analysis based on humor specific features showed that individual features lead to precision between 61% and 65%, with alliteration having the highest presence in the examples (52%).

Kreuz and Caucci studied the importance of several lexical factors in the identification of ironic/sarcastic statements [4]. They randomly collected from the Google Book Search a set of 100 hundred sentences containing the phrase "said sarcastically", and then removed from the sentence the adverb "sarcastically" to eliminate any explicit clue about the ironic content of the statement. They manually analyzed each sentence to check if it contains one of the following linguistic clues: (i) presence of adjectives and adverbs, (ii) presence of interjections, and (iii) usage of punctuation, such as exclamation points or question marks. Then, 101 participants were asked to rate these sentences along with a set of *control sentences*, according to how likely they seem to be ironic without providing any additional contextual information. Ratings where made using a seven-point scale (0 - not at all likely; 7 - very likely). The results show that sarcastic sentences where rated higher than control excerpts (4.85 vs. 2.89). They performed regression analysis in order to determine which lexical features could be used for predicting participant ratings: only the *presence of interjections* was considered a good predictor.

3. CLUES FOR IRONY IN UGC

In video/spoken discourse, especially in a conversational context, we are usually able to detect a variety of external clues (e.g. facial expression, intonation, pause duration) that enable the perception of irony. In written text, a set of more or less explicit linguistic strategies is also used to express irony. In the next subsections, we describe eight linguistic patterns that we have previously identified to be related to the expression of irony (Table 1). Some are specific to Portuguese (e.g. morphological patterns) while others seem to be language independent (e.g. emoticons).

All the patterns in this study restrict somehow the polarity of possible matching sequences, since we are particularly interested in recognizing irony in apparently *positive* sentences involving human named-entities (NE). Hence, most of these patterns contain a polarity constraint, represented by $[4\text{-Gram}^+]$, which requires the presence of at least one prior positive adjective or noun in a window of four words, while excluding the occurrence of *any* negative element in such window.

pattern	match:
P_{dim}	$(4\text{-Gram}^+\ NE_{dim}\ |\ NE_{dim}\ 4\text{-Gram}^+)$
P_{dem}	DEM NE 4-Gram^+
P_{itj}	ITJ_{pos} (DEM ADJ_{pos})* NE (?|!|...)
P_{verb}	NE (tu)* ser_{2s} 4-Gram^+
P_{cross}	(DEM|ART) (ADJ_{pos}|ADJ_{neut}) de NE
P_{punct}	4-Gram^+ (!!|!?|?!)
P_{quote}	"(ADJ_{pos}|N_{pos})\{1,2\}"
P_{laugh}	(LOL|AH|EMO^+)

Table 1: Patterns used in experiments.

3.1 P_{dim}: Diminutive Forms

Diminutives are commonly used in Portuguese, often with the purpose of expressing positive sentiments, like affect, tenderness and intimacy. However, they can also be sarcastically and ironically used for expressing an insult or depreciation towards the entity they represent. This is especially so when diminutives are found in NE mentioning well-known personalities, such as political entities (e.g. "Socratezinho" for the current Portuguese prime-minister, José Sócrates).

3.2 P_{dem}: Demonstrative Determiners

In Portuguese, the occurrence of any demonstrative form – namely, "este" (this), "esse" and "aquele" (that) – before an human NE usually indicates that such entity is being negatively or pejoratively mentioned. In some cases, demonstratives (DEM) are the unique explicit clue that signals the presence of irony (e.g. "Este Sócrates é muito amigo do Sr. Jack" / "This Sócrates is a very good friend of Mr. Jack").

3.3 P_{itj}: Interjections

Interjections abound in subjective texts, particularly in UGC, carrying on valuable information concerning authors' emotions, feelings and attitudes. We believe that some interjections can be used as potential clues for irony detection, when they appear in specific contexts, such as the ones represented in the Pattern P_{itj}. Since we are especially interested in recognizing irony in prior positive text, we confined our analysis to a small set of interjections that are commonly used to express positive sentiments, namely: "bravo", "força", "muito obrigado/a", "obrigado/a", "obrigadinho/a", "parabéns", "muitos parabéns" and "viva".

3.4 P_{verb}: Verb Morphology

The type of pronoun used for addressing people can also be an important clue for irony detection in UGC, especially in languages like Portuguese, where the choice of a specific

pronoun or way of expression (e.g. "tu" vs. "você", both translatable by "you") may depend on the degree of proximity/familiarity between the speaker and the NE it refers to. The pronoun "tu" is used in a familiar context (e.g. with friends and family). In our experiments, we analyze to what extent the use of the pronoun "tu" for addressing a well-know named entity can be used as a clue for irony detection in UGC. As represented in P_{verb}, the pronoun can be either explicitly referred in the text or it can be embedded in the morphology of the verb (which is in the second-person singular). We confined the analysis to the verb "ser" (to be).

3.5 P_{cross}: Cross-constructions

In Portuguese, evaluative adjectives with a prior positive or neutral polarity usually take a negative or ironic interpretation whenever they appear in *cross-constructions*, where adjectives relate to the noun they modify through the preposition "de" (e.g. "O comunista *do* ministro" / "The communist *of the* minister") [2]. Pattern P_{cross} recognizes cross-constructions headed by a positive or neutral adjective (ADJ_{pos} or ADJ_{neut}, respectively), which modify a human NE. Adjectives are preceded by a demonstrative (DEM) or an article (ART) determiner.

3.6 P_{punct}: Heavy Punctuation

In UGC, punctuation is frequently used both for verbalizing user immediate emotions and feelings and for intentionally signaling humoristic or ironic text. We assume that the presence in a sentence of a sequence composed of more than one exclamation point and/or question mark can be used as a clue for irony detection.

3.7 P_{quote}: Quotation Marks

Quotation marks are also frequently used to express and emphasize an ironic content, especially if the content has a prior positive polarity (e.g. positive adjective qualifying an entity). In our experiments, we tried to find possible ironic sentences by searching quoted sequences composed of one or two words, corresponding, at least one of them, to a positive adjective or noun.

3.8 P_{laugh}: Laughter Expressions

Internet slang contains a variety of widespread expressions and symbols that typically represent a sensory expression, suggesting different attitudes or emotions. In our experiments, we considered (i) the acronyms "lol" and corresponding variations (LOL), (ii) onomatopoeic expressions such as "ah", "eh" and "hi" (AH) and (iii) the prior *positive* emoticons ":)" ";-)" and ":P" (EMO$^+$). In this particular case, we did not constrain the polarity of elements contained in the sentence. We assume that laugh expressions are intrinsically positive or ironic.

4. EXPERIMENTAL SET-UP

We collected opinionated user posts from the web site of a popular Portuguese newspaper. This collection is composed of 8,211 news and corresponding comments posted by on-line readers. It includes about 250,000 user posts, totaling approximately one million sentences, in a period of five months (November 2008 to March 2009). On average, user comments have about four sentences. Named-entity recognition was performed by dictionary look-up, using a NE lexicon with 1,226 names of frequently mentioned politicians.

Pattern	# matches
P_{dim}	0
P_{dem}	42
P_{itj}	**127**
P_{verb}	22
P_{cross}	11
P_{punct}	**385**
P_{quote}	**697**
P_{laugh}	**548**

Table 2: Productivity of each pattern.

The NE lexicon was compiled by automatically extracting names from news feeds, which frequently contain recurrent structures (e.g. apposition, quotations) where such entities are mentioned in a rather explicit way. We then generated several possible diminutive forms for each NE, which were used for checking matches of pattern P_{dim}. For testing prior polarity of words, specifically adjectives and nouns, we created a sentiment lexicon with manually annotated polarities. The sentiment lexicon is composed by 3,533 adjectives and 2,522 nouns. In terms of polarity distribution, 55,5% of the entries have a negative prior polarity, 21,8% have a positive prior polarity and the remaining 22,7% are considered neutral.

We scanned the collection for matching each sentence containing at least one person name. For each pattern that matched at least 100 sentences, we performed manual evaluation using the following scheme to classify the matched sentences:

1. **ironic**: the matched content is ironic, or it signals the presence of irony in the sentence;

2. **not ironic**: the matched content is not ironic (i.e. is *literal*) and there is no irony in the sentence;

3. **undecided**: the context is not enough for deciding if irony is present or not in the matched sentence;

4. **ambiguous**: the matched content is ambiguous with other construction.

We have no precise metric of the frequency of irony in our collection. However, by inspecting of the a sample collection, we estimate that it is quite low (much less than 10%).

5. RESULTS AND ANALYSIS

As shown in Table 2, the most productive patterns are directly related to the use of punctuation marks and keyboard characters, which are ways of representing oral or gestural expressions in written text (P_{quote}, P_{laugh}, P_{puntc} and P_{itj}). On the other hand, patterns involving more structured linguistic knowledge – P_{dim}, P_{dem}, P_{verb} and P_{cross} – although theoretically well-grounded, have shown to be ineffective for detecting irony in UGC. In fact, and rather surprisingly, one of the patterns used P_{dim} did not match any sentence in the collection evaluated.

Coverage of patterns is extremely low (1.832 matches in approx. 1 million sentences, i.e. around 0,18%), since we are imposing quite restrictive constraints, both regarding (positive) polarity context and the presence of a name of a public figure. For example, if we remove the constraint

	ironic	not ironic	undecided	ambiguous
P_{itj}	44.88 %	13.39 %	40.94 %	0.79 %
P_{punct}	45.71 %	**27.53**%	26.75 %	0.00 %
P_{quote}	**68.29** %	**21.95** %	2.73 %	7.03 %
P_{laugh}	**85.40** %	0.55 %	11.13 %	2.92 %

Table 3: Results for patterns with 100+ matches.

[4-Gram$^+$] on the Pattern P_{dim}, we match 890 sentences. However, such sentences are mainly literal and express negative opinions. This turns out to be an interesting secondary product of our experiment, since apparently this pattern can be used to detect negative literal opinions quite efficiently. At this point, we are not concerned with increasing coverage of these patterns but with the precision that they achieve in recognizing irony.

Table 3 shows (i) the percentage of ironic sentences correctly identified, (ii) the percentage of literal sentences incorrectly recognized as ironic (iii) the percentage of the cases that we were not able to decide just by inspection of the matched sentence, and (iv) the cases that were incorrectly identified due to ambiguity.

Patterns P_{laugh} and P_{quote} obtained the best performance in irony detection, with precisions of 85.4% and 68.3%, respectively. The remaining patterns evaluated were able to correctly identify irony in about 45% of the sentences including the matched sequences. These numbers are significantly above the baseline that we previously established ($\ll 10\%$).

From Table 3, we can also observe that patterns P_{quote} and P_{punct} incorrectly identify literal sentences as ironic in approximately 22% and 28% of the cases, respectively. Regarding P_{quote}, we observed that errors mainly arise from two typical situations: (i) quotation marks are used to delimit a multiword expression that contains a prior positive word (e.g. "desde há séculos o chamavam 'Santo Contestável'" / "many centuries ago [they] called him 'Saint Contestável'"), and (ii) quotation marks are used to differentiate a technical term or brand, which also includes a prior positive word. For P_{punct}, we noticed that the main source of error is related to the reinforcement of rhetorical questions, not necessarily ironical (e.g. "Onde estão as alternativas democráticas ?!..." / "Where are the democratic alternatives?!") or negative statements.

The number of *undecided* cases is especially expressive in P_{itj} and P_{punct}, reaching approximately 41% and 27%, respectively. For example, P_{itj} matched many comments in which the context does not allow to determine if the statement is ironic or not, such as "parabéns Fonseca, por ter dito a verdade com tanta simplicidade" / "congratulations Fonseca, for having told the truth with so much simplicity". Deciding such cases would require analyzing much wider contexts (e.g. previous sentences), which was out of the scope of this exploratory work.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a set of possible linguistic clues for detecting irony, and explored their efficiency when processing user-generated content. We showed that it is possible to identify ironic opinions in comments that would otherwise be considered as positive, by using relatively simple linguistic patterns. We observed that the most productive patterns involve (i) emoticons and onomatopoeic expressions for laughter, (ii) heavy punctuation marks, (iii) quotation marks and (vi) *positive* interjections. Notably, all these patterns are somehow related with orality, which shows that ironic constructions are frequently signaled by oral clues. We do not claim that the clues that we found to be efficient in UGC can be applied in other text genres, such as news articles or literary text, with comparable success. Likewise, we do not know if the patterns that were found unproductive in UGC turn out to be effective in other text genres. This is a question for future work.

We believe that we can improve the coverage and precision of irony detection procedures mainly in two ways. Coverage can be increased by considering additional linguistic clues, namely the ones explored by Mihalcea and Strapparava for recognizing humor in texts (e.g. alliteration, rhyme, antonymic words or constructions, specific proverbs, etc.) [5]. Precision should also improve by considering the various different clues in combination, as we noticed that ironic sentences usually match more than one of the clues evaluated in this paper.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] J. Birke and A. Sarkar. Active learning for the identification of nonliteral language. In *Proceedings of the Workshop on Computational Approaches to Figurative Language, NAACL-HLT*, Rochester, NY, April 26 2007.

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

[3] R. W. Gibbs. Irony in talk among friends. *Metaphor and Symbol*, 15:5–27., 2000.

[4] R. J. Kreuz and G. M. Caucci. Lexical influences on the perception of sarcasm. In *Proceedings of the Workshop on Computational Approaches to Figurative Language, NAACL-HLT*, Rochester, NY, April 26 2007.

[5] R. Mihalcea and C. Strapparava. Learning to laugh (automatically): Computational models for humor recognition. *Journal of Computational Intelligence*, 2006.

[6] L. Sarmento, P. Carvalho, M. J. Silva, and E. de Oliveira. Automatic creation of a reference corpus for political opinion mining in user-generated content. In *TSA'09 - 1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*, Hong Kong, Nov. 6 2009.

[7] A. Utsumi. A unified theory of irony and its computational formalization. In *Proceedings of the 16th conference on Computational linguistics*, pages 962–967, Morristown, NJ, USA, 1996. Association for Computational Linguistics.

[8] D. Wilson and D. Sperber. On verbal irony. *Lingua*, 87: 53–76, 1992.