Computational rule-based model for Irony Detection in Italian Tweets

Simona Frenda
FICLIT - University of Bologna, Italy
simona.frenda@gmail.com

Abstract

English. In the domain of Natural Language Processing (NLP), the interest in figurative language is enhanced, especially in the last few years, thanks to the amount of linguistic data provided by web and social networks. Figurative language provides a non-literary sense to the words, thus the utterances require several interpretations disclosing the play of signification. In order to individuate different meaning levels in case of ironic texts detection, it is necessary a computational model appropriated to the complexity of rhetorical artifice. In this paper we describe our rule-based system of irony detection as it has been presented to the SENTIPOLC task of EVALITA 2016, where we ranked third on twelve participants.

Italiano. *Nell'ambito del Natural Language Processing (NLP) l'interesse per il linguaggio figurativo è particolarmente aumentato negli ultimi anni, grazie alla quantità d'informazione linguistica messa a disposizione dal web e dai social network. Il linguaggio figurativo conferisce alle parole un senso che va oltre quello letterale, pertanto gli enunciati richiedono interpretazioni plurivoche che possano svelare i giochi di significato del discorso. Nel caso specifico del riconoscimento automatico di un testo ironico, infatti, determinare la presenza di diversi gradi di significazione esige un modello computazionale adeguato alla complessità dell'artificio retorico. In questo articolo descriviamo il nostro sistema "rule-based" dedito al riconoscimento dell'ironia che ha partecipato al task SENTIPOLC di EVALITA 2016, nel quale ci siamo classificati terzi su dodici partecipanti.*

1 Introduction

The amount of texts available on the web and especially in social networks has become a source of linguistic information especially for the Sentiment Analysis. For instance, on Twitter, where the length of tweets is limited (140 characters), users are encouraged to use some creative devices in order to communicate their opinions. In particular they express their emotions or feelings through some morphosyntactic elements or conventional expedients, such as: emoticons, hashtags, heavy punctuation, etc. It seems that these elements represent a substitution of typical gestures and tones of oral communication. In this research we used some linguistic features, frequently found in ironic tweets, as referent points to create the rules of our irony detection system in Italian tweets.

The results we gained are promising and reveal the features considered can be good ironic clues to identify ironic texts.

In the following section we synthetically describe the state of art about irony detection. In the third and fourth sections we present our approach, describing the linguistic resources used and data processing. The fifth section contains the description of linguistic features, and finally in the sixth section we present the results obtained in SENTIPOLC evaluation.

2 Related Work

Although the difficulties of research, it is evident in the literature an attempt to understand this linguistic phenomenon and develop some computational models to detect or generate irony.

In the 90s Lessard and Levison (1992, 1993)[1] and Binsted and Ritchie (1994, 1997)[2] developed the first joke generators and recently Stock and Strapparava (2006) realized HAHAcronym, a system designed to generate and re-analyze the acronyms, considering semantic opposition and rhythm criteria.

The research described by Utsumi (1996) was one of the first approaches to automatic irony processing, even though it was too abstract for a computational framework. In 2009, Veale and Hao noted that English figurative comparisons

1Ritchie (2009: 73).
2Ritchie (2009: 73).

(as X as Y) are often used to express ironic opinions, especially when the marker "about" is present (about as X as Y). Recently, Reyes et al. (2013) produced a multidimensional model for detecting irony on Twitter based on four conceptual features: signatures (pointedness, counterfactuality, and temporal compression), unexpectedness (temporal imbalance and contextual imbalance), style and emotional scenarios (activation, imagery, and pleasantness described by Whissel, 2009[3]). Barbieri and Saggion (2014) proposed a model based on a group of seven sets of lexical and semantic features of the words in a tweet: frequency, written-spoken style, intensity of adverbs and adjectives, structure (punctuation, length, emoticons), sentiments, synonyms and ambiguity.

Karoui et al. (2015) focused on the presence of negation markers as well as on both implicit and explicit opposition in French ironic tweets. Moreover, this research highlights the importance of surface traits in ironic texts, such as: punctuation marks (González-Ibáñez et al., 2011), sequence or combination of exclamation and question marks (Carvalho et al., 2009; Buschmeier et al., 2014), tweet length (Davidov et al., 2010), interjections (González-Ibáñez et al., 2011), words in capital letters (Reyes et al., 2013), emoticons (Buschmeier et al., 2014), quotations (Tsur et al., 2010)[4], slang words (Burfoot and Baldwin, 2009)[5] and opposition words, as "but" or "although" (Utsumi, 2004)[6].

Carvalho et al. (2009) distinguished eight "clues" for irony detection in some comments (each consisting of about four sentences) from a Portuguese online newspaper. Their attention focused on positive comments because in a previous research they showed that positive sentences are more subjected to irony and it is more difficult to recognize their true polarity. So the idea is to identify the irony in apparently positive sentences that require the presence of at least one positive adjective or noun in a window of four words. Carvalho et al. (2009) based their model on both oral and gestural "clues" of irony, such as: emoticons, heavy punctuation, quotation marks, onomatopoeic expressions for laughter and positive interjections and, on the other hand, on specific morphosyntactic constructions, such as: the diminutive form of NE, the demonstrative determiners before NE, the pronoun "*tu*" specifically referred or embedded in the morphology of the verb "*ser*".

Our work proposes an adaptation for some of these clues, increased by other surface features, to Italian irony detection in Twitter.

3 Methodology

Approaching the detection of irony in tweets means to understand how people, especially net users, make irony. We try to approach this hard work by analyzing the corpus of tweets and identifying possible ironic clues. Once identified, surface features common to ironic tweets are inserted as binary rules in our system.

Our rule-based system, written in Perl, finds ironic features (described in section 5) in tweets and consequently distinguishes the ironic ones from the non-ironic.

In the following sections we describe resources used, data processing, ironic clues and the results obtained in the EVALITA 2016 SENTIPOLC task.

4 Analysis of corpus

For this research we used a corpus of tweets provided by SENTIPOLC organizers (Barbieri et al., 2016). This training set is composed of 7410 tweets labeled according to the criteria of subjectivity, overall and literal polarity (positive/neutral/negative/mixed), irony and political topic.

4.1 Resources

For the analysis and processing of Italian tweets we used some linguistic resources available online, such as:

- *Sentiment Lexicon* LOD (*Linked Open Data*). Developed by the Institute for Computational Linguistics "A. Zampolli", it contains 24.293 lexical entries annotated with positive/negative/neutral polarity.
- Morph-it! (Zanchetta and Baroni, 2005). It is a lexicon of inflected forms of 34.968 lemma (extracted from the corpus of "La Repubblica") with their morphological features.

A tweet is composed of different essential elements for linguistic analysis, as interjections and emoticons. We therefore developed a lexicon of interjections and a list of emoticons described summarily below:

3Reyes et al. (2013: 249).
4Karoui et al. (2015).
5Karoui et al. (2015).
6Karoui et al. (2015).

- The interjections, extracted from Morph-it! and Treccani[7], are manually annotated with their polarity. The annotation has been developed with the support of Vocabolario Treccani, while the sentiment lexicon has been used to label improper interjections (see Table 1).
- The emoticons, extracted from Wikipedia, are subdivided in EMOPOS, EMONEG and EMOIRO, according to the classification of Di Gennaro et al. (2014) and Wikipedia description[8], especially for the ironic annotation (see Table 2).

Positive	Negative	Neutral
evviva	mah	boh
urrà	macché	mhm
complimenti	bah	chissà
congratulazioni	puah	beh

Table 1: Example of annotated lexicon of interjections.

Label	Emoticon
EMOPOS	=) =] :D (-: [-: (-; [-; :-> :) :-) (; ;)
EMONEG	:[=(:-(:'(:-/ :/ :-> :\> :/ =/ =\ :L =L :S
EMOIRO	^^ ^.^ :P xP ^3^ ^L^ ^_^ ^-^ ^w^

Table 2: Example of annotated list of emoticons.

4.2 Data Processing

Incoming file processed by our system has been previously lemmatized and syntactically annotated by TreeTagger (Schmid, 1994) with Italian tagset provided by Baroni.

Nevertheless, before syntactic analysis, we applied the rules of substitution and elimination of some textual elements, in order to clean up the texts and avoid hampering the process of POStagging and lemmatization of TreeTagger. In particular:

- the label EMOPOS replaces positive emoticons;
- the label EMONEG replaces negative emoticons;
- the label EMOIRO replaces ironic emoticons;
- the characters of url are removed.

This method allows us to clean up the texts from those characters that may hinder the analysis of data and ironic clues retrieval.

5 Features

In section 2 we have presented the research of Carvalho et al. (2009) which demonstrated how the most productive patterns (with a precision from 45% to 85%) are the ones related to orality and gesture, as emoticons or expressions for laughter. Based on this analysis, we try to recognize ironic tweets with a system designed to find ironic clues into the texts. Some of these clues are adapted to Italian language from Portuguese, while some other features are individuated during the analysis of the tweets.

All of these features are used as binary rules in our system to classify the texts in ironic and non-ironic.

5.1 Positive Interjections

Ameka (1992)[9] describes the interjections as "relatively conventionalized vocal gestures which express a speaker's mental state, action or attitude or reaction to a situation". These linguistic elements are used as simple ways to communicate user's feelings or moods.

In previous researches interjections were represented as good humor clues. Kreuz and Caucci (2007) tried to determine if specific lexical factors might suggest the interpretation of a statement as sarcastic. They demonstrated with a test that the presence of interjections is a good predictor for the readers. They provided a group of students with some extracts from various works, a part of which originally contained the word "sarcastically". Students were able to classify correctly the extracts where the word "sarcastically" was deleted thanks to the interjections.

Carvalho et al. (2009) noted that positive interjections has very often an ironical use in apparently positive utterances.

Taking into consideration these precedent researches, we consider improper and proper interjections annotated with positive polarity (see Table 1 in section 4.1). Improper interjections are

7http://www.treccani.it

8Wikipedia version of the 6th of June.

9Lindbladh (2015: 1).

usually followed by exclamations or question marks, which suggest a rising intonation ("*sicuro!*"), whereas proper ones (or onomatopoeic expressions) are sometimes added to the phrase without any punctuation characters ("*ah dimenticavo*", "*ah comunque*").

5.2 Expressions with "*che*"

The adjective or pronoun "*che*" can be used with exclamatory intention in expressions such as "*che ridere*", "*che educato*", "*che sorpresa*". Like interjections, these expressions are used as marks to express user's emotions and their ironic intent.

5.3 Pronoun "*tu*" and Verb Morphology

The use of pronoun "*tu*" and its morphological inflection of the verb "*essere*" expresses a high degree of proximity between the user and the person it refers to (Carvalho et al., 2009). For instance, if this person is a popular politician, this degree of familiarity is fake or artificial and it is usually used ironically in the tweets.

5.4 Disjunctive Conjunction

In the training set we note how disjunctive conjunctions ("*o*", "*oppure*") are used to introduce an alternative between two propositions or concepts which may belong to very different semantic domains (for example: *In televisione stamattina: i cartoni animati o Mario Monti.[…]*). This strange combination of ideas surprises the readers and suggests them a possible ironic interpretation of the message.

5.5 Onomatopoeic Expressions for laughter

Onomatopoeic expressions for laughter (the most diffused are "*ahah*", "*hehe*" and "*ihih*") are usually used in humorous texts (Carvalho et al., 2009; Buschmeier et al., 2014) with their variants (in capital letters or with repetitions). They represent some marks which inform the reader about the user's mood and also suggest that the tweet must be interpreted in a figurative sense.

5.6 Ironic Emoticons

Users utilize emoticons to show their facial expressions as well as their emotions in the texts. Tavosanis (2010) presents a macro-classification of emoticons: expressive, decorative/pleasant and of morphosyntactic substitution, which stand for a word or a whole phrase.

In our research we only consider expressive emoticons which add information about the user's mood. In particular we focus on the ironic emoticons, those which express joking or ironic intention (see section 4.1). We have distinguished EMOIRO from EMOPOS because positive emoticons (considered in Carvalho et al., 2009 and González-Ibáñez et al., 2011) are frequently used to express a humorous intention, not specifically ironic.

5.7 Hashtag

Hashtag is a special element in the syntax of tweets used to connect those ones containing the same keywords (which may be a part of the speech) or phrases as *#mobbastaveramenteperò*.

The user communicates through hashtags several information about events, people they refers to and the topic of message. We focus on hashtags that may suggest to the readers an ironic connotation of the message as #lol and #ironia, and on others that we extracted from ironic tweets in the training set: #stranezze, #Ahahahahah, #benecosì, etc.

5.8 Regional Expressions

It seems that regional expressions are utilized by users in ironic texts to underline their own mood and emotions. In particular, common constructions deriving from local use may be: "*annamo bene*", "*namo bene*" and "*ce*" followed by the verb (e.g. "*ce vuole*", "*ce sta*", "*ce potrebbe*"), as in this ironic tweet: "*@zdizoro t'appassionerà sapè che nel prossimo governo #Monti ce potrebbe rimanè MaryStar Gelmini, come n'incrostazione*".

5.9 Quotation Marks

We focus on the use of quotation marks as a sign for the readers to interpret non-literally the content of text. In fact, in the social networks these elements are frequently used to underline the possible different meanings of the word between quotation marks, and emphasize the ironic content.

5.10 Heavy Punctuation

In web communication the punctuation plays an important role in the expression of the emotions and feelings. Several researches (González-Ibáñez et al., 2011; Kreuz and Caucci, 2007; Carvalho et al., 2009a; Buschmeier et al., 2014; Davidov et al. 2010; Karoui et al., 2014) considered the punctuation as a surface feature to signal humorous texts. In particular we focus on combi-

nation of question and exclamation marks to irony detection.

6 Results

Our system is evaluated on the SENTIPOLC official test data composed of 3000 tweets and the values of precision, recall and average F-score are calculated using the evaluation tool provided by the organizers (Barbieri et al., 2016). As we can see from Table 3, official results of our system are promising, although our research in this domain has to be improved.

Rank	F-score
1	0.548
2	0.5412
3	**0.5251**
4	0.5162
5	0.5133
6	0.4992
7	0.4961
8	0.4872
9	0.481
10	0.4761
11	0.4728
12	0.4725

Table 3: Official results and ranking of Irony Detection sub-task.

7 Conclusion

In this paper we have described our computational model based on linguistic features which have proven to be good clues for the identification of ironic texts. Nonetheless, in future works we plan to examine in depth semantic inconsistencies and ambiguities, amusing wordplay and rhymes that may surprise the reader. In conclusion, we think that a good detection of irony is possible if all the levels of linguistic analysis are considered.

References

Francesco Barbieri and Horacio Saggion. 2014. Modelling Irony in Twitter, Features Analysis and Evaluation. *Language Resources and Evaluation conference, LREC*. Reykjavik, Iceland.

Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli and Viviana Patti. 2016. Overview of the EVALITA 2016 SENTIment POLarity Classification Task. In Pierpaolo Basile, Anna Corazza, Franco Cutugno, Simonetta Montemagni, Malvina Nissim, Viviana Patti, Giovanni Semeraro and Rachele Sprugnoli, editors, *Proceedings of Third Italian Conference on Computational Linguistics (CLiC-it 2016) & Fifth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2016)*. Associazione Italiana di Linguistica Computazionale (AILC).

Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger. 2014. An impact analysis of features in a classification approach to irony detection in product reviews. *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Baltimore, Maryland, USA. 42–49.

Paula Carvalho, Luís Sarmento, Mário J. Silva and Eugénio De Oliveira. 2009. Clues for detecting irony in user-generated contents: Oh...!! it's "so easy" ;-). *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. ACM. 53–56.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. *Proceedings of the Fourteenth Conference on Computational Natural Language Learning*, CoNLL '10. Stroudsburg, PA, USA. Association for Computational Linguistics. 107–116.

Pierluigi Di Gennaro, Arianna Rossi and Fabio Tamburini. 2014. The FICLIT+CS@UniBO System at the EVALITA 2014 Sentiment Polarity Classification Task. *Proceedings of the Fourth International Workshop EVALITA 2014*. Pisa University Press.

Roberto González-Ibáñez, Smaranda Muresan and Nina Wacholder. 2011. Identifying Sarcasm in Twitter: A Closer Look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: shortpapers*. Portland, Oregon. 581–586.

Jihen Karoui, Farah Benamara Zitoune, Veronique Moriceau, Nathalie Aussenac-Gilles and Lamia Hadrich Belguith. 2015. Detection automatique de l'ironie dans les tweet en francais. *22eme Traitement Automatique des Langues Naturelles.* Caen.

Roger J. Kreuz and Gina M. Caucci. 2007. Lexical Influences on the Perception of Sarcasm. *Proceedings of the Workshop on Computational Approaches to Figurative Language*. Rochester, NY. 1–4.

Sara Lindbladh. 2015. La semantica e pragmatica dei segnali discorsivi italiani – un confronto tra bene,

va bene, be' e va be'. *Seminarium 27 oktober*. Universita di Uppsala, Sweden.

Antonio Reyes, Paolo Rosso and Tony Veale. 2013. A multidimensional approach for detecting irony in-Twitter. *Language Resources and Evaluation*. 47: 239–268.

Graeme Ritchie. 2009. Can computers create humor? *AI Magazine*. Volume 30, No. 3. 71-81.

Helmut Schmid. 1994. Probabilistic Part-of-Speech-Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Oliviero Stock and Carlo Strapparava. 2006. Laughing with HAHAcronym, a computational humor system. *Proceedings of the Twenty-First National Conference on Artificial Intelligence* (AAAI-06), Boston, Massachusetts.

Mirko Tavosanis. 2010. *L'italiano del web*. Carocci. Roma.

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

Tony Veale and Yanfen Hao. 2009. Support structures for linguistic creativity: A computational analysis of creative irony in similes. *Proceedings of CogSci 2009, the 31st annual meeting of the cognitive science society*. 1376–1381.

Eros Zanchetta and Marco Baroni. 2005. Morph-it! A free corpus-based morphological resource for the Italian language. *Proceedings of Corpus Linguistics 2005*. University of Birmingham, Birmingham, UK.