Distinguishing between Irony and Sarcasm in Social Media Texts: Linguistic Observations

Maria Khokhlova
St.Petersburg State University
St.Petersburg, Russia
m.khokhlova@spbu.ru

Viviana Patti
Università di Torino
Turin, Italy
patti@di.unito.it

Paolo Rosso
Universitat Politècnica de València
Valencia, Spain
prosso@dsic.upv.es

Abstract—**Automatic detection of figurative constructions used in texts can be quite a tricky task. In order to make it possible for a program to detect such constructions in a text it is necessary to understand what principles underlie sarcastic and ironic utterances and thus can be used for their identification. The paper deals with the linguistic differentiation between irony and sarcasm based on the analysis of tweets. We discuss the linguistic peculiarities of eight corpora marked with #irony and #sarcasm hashtags.**

I. Introduction

Social media texts have their own characteristics. First of all one can name their conciseness and laconism. Such texts are quite difficult material for interpretation also because of mistakes, misspellings, incorrect grammar, various abbreviations and abruptness. Authors like to use colloquial style for their messages and to keep their idiolects. Among other techniques they can make jokes or mock at something balancing between literal and figurative meanings. Even in oral speech (let alone written texts) we can face with difficulties when interpreting ironic or sarcastic exclamations. But in dialogues facial expressions or voice pitch can sometimes be hints to this kind of information suggesting more than one interpretation of an utterance. Computational models for irony detection in social media have been proposed, mostly focused on Twitter, but only a few preliminary studies investigate the differences between irony and sarcasm [1]. The current work aims to contribute to this subject from a linguistic perspective.

The rest of the paper is organized as follows. In the next section we dwell on the notions of "irony" and "sarcasm" and report on related work. In Section 3 we describe our data, while Section 4 presents the results of the analysis. Finally, Section 5 closes the paper with conclusions and suggestions for future work.

II. Related work

Irony and sarcasm are both ways (or figures of speech) of saying something when the opposite is meant [2]. Both concepts have very fuzzy boundaries and the difference between these two forms of expression is sometimes very subtle. They have in common the absence of an explicit marker [3] that points to the fact that the literal meaning is negated and thus we deal with ambiguity.

Sarcasm and irony represent certain inherent traits of natural language and are wide-spread in English texts. The former can be traced back to the Greek word "σαρκασμός" that means "to tear flesh", e.g. to speak in this manner. While the latter comes from the Greek word "εἰρωνεία" that can be translated as "dissimulation". In dictionaries we can sometimes find definitions only for one of the terms not for both ([4] has dictionary entry only on "irony"). Irony can be defined wider than sarcasm, the latter is sometimes viewed as its type. Sarcasm is often interpreted as malicious irony that is intended to mock at somebody [3]. Sarcastic utterances can express indignation and hate in their highest form. In [5] the authors point to the fact that sarcastic utterances have usually a victim as their target. Whereas ironic sentences are not aimed at anybody, they have general nature.

Many authors described psycholinguistic experiments that involved identification of sarcastic and ironic sentences [5], [6], [7]. Figurative language poses a serious challenge to automatic systems that achieve good results in case of literal language (that is fairly predictable). But nevertheless automatic detection of irony and sarcasm in texts becomes more and more popular among researchers. Correctly identified ambiguous utterances can improve performance of software for opinion mining and sentiment analysis [8]. A special task SemEval-2015 was dedicated to the sentiment analysis of figurative language in Twitter [9].

Within the past decade a number of systems dealing with sarcasm detection have been designed. Some approaches in literature [10], [11], [12] propose methods to automatically extract irony and sarcasm using frequent and typical expression as features. In [13] the authors criticize this approach and propose a computational model for detecting sarcasm in tweets that involves seven sets of lexical features without the use of words and patterns of words. They claim that sarcastic expressions are language-specified and thus it would be better to avoid them when we design an automatic system.

ISBN 978-952-68397-6-9 (paperback), ISBN 978-952-68397-7-6 (PDF)

The huge amount of information streaming from social media platforms is increasingly attracting the attention of many kinds of researchers and linguistic analysis of social media has become a relevant topic. Social media texts are difficult to interpret as they are rather concise. It is a trickier task in case of tweets when a text is limited to the 140 characters, so the surrounding context is poor and a sentence cannot be reliably judged as either ironic or sarcastic. Thus it is necessary to manually analyze tweets that have already been tagged with #irony or #sarcasm by the authors in order to make inferences that can be used for automatic detection of these traits in other data.

A number of works focus on exploring corpora and analyze texts on the surface level [14], [15], [16], [17]. Such analysis involves the interpretation of a range of features on lexical, morphological and syntactic levels that can be efficiently defined and used in machine-learning algorithms. Few preliminary studies addressed the task to investigate the differences between irony and sarcasm. A contribution on this line is given in [1], where authors analyze messages explicitly tagged by users with #irony, #sarcasm and #not in order to test the hypothesis to deal with different linguistic phenomena, with a special focus on the role of features related to the multi-faceted affective information expressed in such texts. The current work aims to further contribute to this subject.

III. Data and Methodology

The aim of our research was to study sarcastic and ironic utterances in corpora. Taking into account the fact mentioned in Section II that sarcasm can be viewed as a kind of irony, in our research we will focus on the distinction of sarcasm and other types of irony. In our study we used texts written via the microblogging platform Twitter (tweets) marked with #irony and #sarcasm hashtags. We made an assumption that data labeled by the users is reliable (the authors deliberately assigned the hashtags and employed certain linguistic means) [18] and thus can be used for the comparison between irony and sarcasm. Altogether we have analyzed six corpora (described in detail in [18]), their common volume is over 800 thousand tokens (see Table I).

TABLE I. List of corpora

Corpus	Number of tweets	Number of tokens
Irony TwBarbieri 2014	50,000 (10,000 ironic, 40,000 non-ironic)	144,072
Irony TwReyes 2013	40,000 (10,000 ironic, 30,000 non-ironic)	160,757
Not	10,000	111,800
Sarcasm TwPtacek 2014 Imbalanced	100,000 (25,000 sarcastic, 75,000 non-sarcastic)	282,874
Sarcasm TwRiloff 2013	3,200	6,241
Sarcasm TwBarbieri 2014	50,000 (10,000 sarcastic, 40,000 non-sarcastic)	126,750

Below we exemplify tweets from our corpora, users' spelling is kept.

Examples of tweets marked with #sarcasm are as follows:

I always know its gonna be a good day when I wake up late and literally have 10 minutes to get ready.

Dear Teva, thank you for waking me up every few hours by howling. Your just trying to be mother natures alarm clock.

I just love getting a random schedule change and not even knowing about it!:)

Examples of tweets marked with #irony are as follows:

I just got 100% on a stress test for Health class I was stressing about. Irony!

My American History teacher is foreign. .

I work in publishing and my boss is currently stealing e-books to take on holiday.

We used the TwitIE software for tokenizing and tagging microblog texts as it is specially designed for this type of texts and achieves high accuracy in their processing [19]. TwitIE uses the Penn Treebank Project tagset [20]:

A	DT	a
driving	VBG	drive
school	NN	school
teacher	NN	teacher
just	RB	just
hit	VB	hit
my	PRP$	my
sisters	NN	sister
car	NN	car

We used the Sketch Engine system [21], [22] for the analysis of our corpora (concordance, word list, thesaurus and word sketch tools).

In our study we used the NRC Word-Emotion Association Lexicon also called EmoLex [23], [24]. This lexicon is a list of 14,182 English words (unigrams) that belong to two sentiments (positive and negative) and are labelled with eight Plutchik's primary emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) [25]. Table II gives an example of the word "nobility" and the assigned sentiments and emotions in NRC Emotion Lexicon (1 — true; 0 — false).

TABLE II. EXAMPLE OF NRC EMOTION LEXICON

Word	Sentiment or emotion	Value
nobility	anger	0
nobility	anticipation	1
nobility	disgust	0
nobility	fear	0
nobility	joy	0
nobility	negative	0
nobility	positive	1
nobility	sadness	0
nobility	surprise	0
nobility	trust	1

In our research we tried to consider lexical properties of the texts, the choice of words, constructions and their order.

As preliminary observation we can say that special attention should be paid to the following linguistic properties of tweets:

1) punctuation marks and emoticons;

2) usage of special lexis;

3) n-grams;

4) polarity lexis.

IV. ANALYSIS

A. N-grams

On the first stage of our analysis we studied high frequency words and phrases in two types of corpora.

Ironic texts tend to have more constructions with negation, e.g.: I don't, I can't, don't know.

"I don't want to be average." Is such an average thought.

These texts have more proper names (names of companies, trademarks, personal names etc.) and this fact supports [26] in mentioning that irony is inherent to communication with intimates. Sarcastic texts do not use so much proper names as it is the case of ironic texts.

The highest frequency of the verb "love" can be found in sarcastic texts (5,074.88 ipm (item per million) vs. 1,540.85 ipm in ironic texts). The authors of sarcastic texts use a number of constructions with the verb, e.g. *I just love…*, *... love when people…*, *I love being ...*

For the construction "I + ADV + love" the following forms tend to be the most frequent ones in sarcastic texts: <u>*I just love*</u> (406 examples, e.g. *I just love getting a random schedule change and not even knowing about it!:)*); <u>*I absolutely love*</u> (28 sentences, e.g. *I absolutely love when people hang up the phone on me*). In #irony corpus we find only 31 example of this construction. Corpora with #not hashtag also tend to have high frequency of the constructions with "love" that makes them quite similar to the sarcastic texts.

Sarcastic texts also are more egocentric (authors tend to use much more "I" pronouns than in other texts).

B. Parts of speech

We studied also the ratio of parts of speech in both corpora (see Table III). In ironic texts we find more nouns (not only proper names). Other parts of speech in general are more frequent in sarcastic texts. We can suppose that #sarcasm corpus is more emotional as we see more adjectives and adverbs.

TABLE III. RATIO OF PARTS OF SPEECH IN CORPORA

tag	Corpus	
	irony	sarcasm
nouns	166383 (37.26%)	175882 (32.61%)
verbs	82288 (18.43%)	106317 (19.72%)
adjectives	30763 (6.89%)	41992 (7.79%)
prepositions	43888 (9.23%)	45865 (8.5%)
pronouns	35703 (8.0%)	52411 (9.72%)
articles	19313 (4.33%)	37684 (7.0%)
adverbs	27178 (6.09%)	44769 (8.3%)
other	9.77%	6.36%

C. Case Study: the verb "love":

- Irony corpus:

Obj >> ing_complement;

love when < love how;

love + Obj (concrete concepts, heterogeneous in their structure, having one hit in this construction in the corpus).

I love how I spend so much time educating my patients about correct diets and I have a horrible diet.

I love how my best friend and I are interested in another set of best friends #twins

I love wasting paper to print my APES homework

My UX friends are going to love the title of Jakob's new post

The more I love my iPhone , the more I want a new one.

In this corpus we find much more examples with love + object (Obj) than with –ing complement. In the examples objects are represented by concrete concepts, although heterogeneous in their structure: bird, song, shark. Construction *love how* prevails over *love when*.

- Sarcasm corpus:

Obj ≈ ing_complement (*love being ignored*);

love when ≈ love how;

love + Obj (beverages, weather, smell, day, night, homogeneous, several hits of the same example).

I just love waking up to find myself home alone.

I love how the freezing wind cuts right thru my apt's closed/locked windows & blows my blinds so hard they knock stuff over #brrrr

@SmallTwnProbs: I love warm beer #smalltownproblems

Don't you just love people on trains who have incredibly loud phone call conversations that the whole coach can hear IJS..

I love those people who follow you and when you follow them back they unfollow you. ???

The sarcastic texts appear to have the same number of constructions both with object and with -ing complement. Among objects in love + object construction we find various semantic groups: beverages, day/night, weather etc. We find the construction *I just love* tends to be a certain trait of sarcastic tweets having 771.64 ipm hits vs 22.81 in ironic texts. These groups are more homogeneous and also each of the nouns has several hits in sarcastic texts. *love when* and *love how* constructions also have quite the same frequency.

- Corpus without hashtags:

Obj << ing_complement (love being, love waking up, love getting);

love when > love how;

love + Obj (abstract concepts, homogeneous, several hits of the same example)

yeah totally I love getting blamed for everything

I love when I have a week off from school and I wake up at 6am

To the generator outside my window, thanks for the wonderful sleep last night. I love Mondays. I really do.

I just love pointless meetings

How I love the smell of worms when it rains I picked the best day to do my running around...

In the #not corpus we find more constructions with –ing complement than with objects (they tend to be abstract ones). The sarcasm corpus tend to be similar to the #not corpus. The construction *I just love* is also very frequent in this corpus.

D. Comparison with NRC Emotion Lexicon

We got frequency lists for the sarcasm and irony corpora and compared them to the EmoLex list. Below (Table IV) you can find an example.

Table V shows the percentage of positive and negative lexis that was evaluated with the help of EmoLex list.

TABLE IV. EXAMPLE OF EMOLEX

word	sentiment or emotion	frequency 1 (#irony)	frequency 2 (#sarcasm)
achievement	trust	1	2
acid	negative	2	1
acquire	positive	2	1
action	positive	17	8
actual	positive	12	9
addiction	negative	5	2
admire	positive	1	2

TABLE V. POSITIVE AND NEGATIVE DISTRIBUTION AMONG IRONY AND SARCASM CORPORA

Corpus	Positive	Negative
Irony	45.7%	54.3%
Sarcasm	48%	52%

As we see sarcastic texts are slightly more positive than ironic texts. However negative polarity prevails in both types of corpora (this is the common tendency in social media texts). It is worth mentioning that the presence of negative or positive words in a tweet does not always provide a reliable conclusion of their attitude, nevertheless such information can be a hint for the interpretation of the mechanisms underlying emotional attitude.

TABLE VI. DISTRIBUTION OF EMOTIONS AMONG IRONY AND SARCASM CORPORA

	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust
Irony	14.20%	11.49%	11.16%	16.22%	9.65%	13.21%	7.25%	16.82%
Sarcasm	13.26%	12.1%	10.85%	15.93%	10.52%	13.23%	7.45%	16.66%

E. Hashtags

People like using more hashtags in ironic texts giving references to other authors or objects and the texts are more structured. In these tweets people give more replies to other users, and on the contrary #sarcasm corpora include “rhetorical” texts that are not aimed at anything and don’t imply any answer.

F. Structure

We find the data in the irony corpus that varies in its form suggest that people use irony on different topics (that seem important for them) but they use sarcasm when talking about usual concepts (drinks, pastime etc.), so there could be "frames" they tend to be consistent with. Words-tokens ratio tends to be lower in sarcastic texts, so the lexis the authors use is not so divergent as in ironic texts. Also there are more non-words and punctuation marks in tweets marked with #sarcasm. They appear to be more emotional (a high number of interjections and exclamation marks prove this fact).

G. Other results

Some tweets tagged by #irony or #sarcasm are difficult to interpret as ironic or sarcastic. And hashtags are the only indication on figurative meaning.

V. Conclusion and future work

In our preliminary study we analyzed #sarcasm and #irony corpora in order to detect their differences that can be used for improving machine-learning algorithms in future. We believe that such linguistic analysis not involving automatic techniques is a crucial step in works that deal with distinguishing between irony and sarcasm.

Psycholinguistic experiments can be seen as one of the next stages of the investigation, e.g. comparison of the labels assigned by the informants and by the authors of tweets.

We can enhance the results comparing our corpora with other "clear" tweets that include neither #sarcasm nor #irony hashtags. It would be also interesting to trace if the same constructions and lexis is inherent to ironic and sarcastic texts of another language, for example, Italian, Spanish or Russian. To understand to what extent this figurative language in this sense is language dependent. It is also worth studying the differences the phenomena of irony and sarcasm in the same language of native speakers from different countries (for English it can be USA, UK, Australia etc., for Spanish users from Spain, Mexico, Argentina etc.).

VI. Acknowledgements

The work of Paolo Rosso has been partially funded by the SomEMBED TIN2015-71147-C2-1-P MINECO research project and by the Generalitat Valenciana under the grant ALMAMATER (PrometeoII/2014/030).

We are grateful to anonymous reviewers for their valuable comments.

References

[1] E. Sulis, D. I. Hernández Farías, P. Rosso, V. Patti and G. Ruffo, Figurative messages and affect in Twitter: Differences between #irony, #sarcasm and #not, Knowledge-Based Systems. In Press. Available online. http://dx.doi.org/10.1016/j.knosys.2016.05.035

[2] H.P. Grice, "Logic and Conversation", *Syntax and Semantics*, vol. 3, Speech Acts, ed. by P. Cole and J. L. Morgan. New York: Academic Press, 1975, pp. 41–58.

[3] R. Giora, "On irony and negation", *Discourse processes*, Vol. 19 (2), 1995, pp. 239-264.

[4] P.H. Matthews, *Oxford Concise Dictionary of Linguistics*. Oxford: Oxford U.P., 2007.

[5] R.J. Kreuz and S. Glucksberg, "How to be sarcastic: The echoic reminder theory of verbal irony", *Experimental Psychology: General*, vol. 118, 1989, pp.374-386.

[6] R.J. Kreuz and G.M. Caucci, "Lexical influences on the perception of sarcasm", *in Proc. of the NAACL workshop on Computational Approaches to Figurative Language (HLT–NAACL 2007)*. New York: ACL, 2007, pp. 1777-1806.

[7] C.J. Lee and A.N. Katz, "The differential role of ridicule in sarcasm and irony", *Metaphor and Symbol*, Vol. 13 (1), 1998, pp. 1-15.

[8] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval,* vol. 2, nos. 1-2, 2008, pp. 1-135.

[9] A. Ghosh, G. Li, T. Veale, P. Rosso, E. Shutova, J. Barnden and A. Reyes, "Semeval-2015 task 11: Sentiment analysis of figurative language in Twitter", *in Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, , Denver, Colorado, June. Association for Computational Linguistics, 2015, pp. 470–478

[10] D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcastic sentences in twitter and amazon", *in Proc. of the Fourteenth Conference on Computational Natural Language Learning*. Stroudsburg: ACL. 2010. pp. 107-116.

[11] R. González-Ibáñez, S. Muresan and N. Wacholder, "Identifying sarcasm in Twitter: a closer look", *in Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*, vol. 2. Stroudsburg: ACL, 2011, pp. 581-586.

[12] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert and R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation", *in Proc. of the 2013 Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, pp. 704–714.

[13] F. Barbieri, H. Saggion and F. Ronzano, "Modelling Sarcasm in Twitter, a Novel Approach", *in Proc. of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Baltimore:ACL, pp. 50-58.

[14] A. Reyes and P. Rosso, "Making Objective Decisions from Subjective Data: Detecting Irony in Customers Reviews", *Journal on Decision Support Systems*, vol. 53, issue 4, 2012, pp. 754–760.

[15] A. Reyes, P. Rosso and T. Veale, "A Multidimensional Approach for Detecting Irony in Twitter", in *Language Resources and Evaluation*, vol. 47, issue 1, 2013, pp. 239-268.

[16] E. Filatova, "Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing", *in Proceedings of Language Resources and Evaluation Conference*, 2012, pp. 392–398.

[17] Ch. Liebrecht, F. Kunneman, A. van den Bosch, "The perfect solution for detecting sarcasm in tweets #not", *in Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 29–37.

[18] D. Hernández Farías, V .Patti and P. Rosso "Irony Detection in Twitter: The Role of Affective Content", *in ACM Transactions on Internet Technology*, in press.

[19] K. Bontcheva, L. Derczynski, A. Funk, M.A. Greenwood, D. Maynard and N. Aswani, 2013. "TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text". In Proceedings of the International Conference on Recent Advances in Natural Language Processing, ACL.

[20] Penn Treebank Project, Web: https://www.cis.upenn.edu/~treebank/

[21] A. Kilgarriff, P. Rychly, P. Smrz, D. Tugwell, "The Sketch Engine", *in Proc EURALEX 2004*, Lorient, France, pp. 105–116.

[22] Sketch Engine, Web: http://www.sketchengine.co.uk

[23] S. Mohammad and P. Turney, "Crowdsourcing a Word-Emotion

Association Lexicon", *Computational Intelligence,* vol. 29 (3), 2013, pp. 436-465.

[24] NRC Word-Emotion Association Lexicon, Web: http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[25] R. Plutchik, 2001, "The nature of emotions", *in Am. Scientist* 89 (4), pp. 344–350 .

[26] Ju. Jorgensen, "The functions of sarcastic irony in speech", in *Journal of Pragmatics*, 26, 1996, pp. 613-634.