



ELSEVIER

Contents lists available at [ScienceDirect](#)

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Signaling sarcasm: From hyperbole to hashtag

 Florian Kunneman^a, Christine Liebrecht^b, Margot van Mulken^a, Antal van den Bosch^{a,*}
^aCentre for Language Studies, Radboud University Nijmegen, The Netherlands

^bThe Amsterdam School of Communication Research ASCoR, University of Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 15 August 2013

Received in revised form 6 March 2014

Accepted 24 July 2014

Available online xxxx

Keywords:

Social media

Automatic sentiment analysis

Opinion mining

Sarcasm

Verbal irony

ABSTRACT

To avoid a sarcastic message being understood in its unintended literal meaning, in micro-texts such as messages on Twitter.com sarcasm is often explicitly marked with a hashtag such as '#sarcasm'. We collected a training corpus of about 406 thousand Dutch tweets with hashtag synonyms denoting sarcasm. Assuming that the human labeling is correct (annotation of a sample indicates that about 90% of these tweets are indeed sarcastic), we train a machine learning classifier on the harvested examples, and apply it to a sample of a day's stream of 2.25 million Dutch tweets. Of the 353 explicitly marked tweets on this day, we detect 309 (87%) with the hashtag removed. We annotate the top of the ranked list of tweets most likely to be sarcastic that do not have the explicit hashtag. 35% of the top-250 ranked tweets are indeed sarcastic. Analysis indicates that the use of hashtags reduces the further use of linguistic markers for signaling sarcasm, such as exclamations and intensifiers. We hypothesize that explicit markers such as hashtags are the digital extralinguistic equivalent of non-verbal expressions that people employ in live interaction when conveying sarcasm. Checking the consistency of our finding in a language from another language family, we observe that in French the hashtag '#sarcasme' has a similar polarity switching function, be it to a lesser extent.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In the general area of sentiment analysis, sarcasm is a disruptive factor that causes the polarity of a message to flip. Unlike a simple negation, a sarcastic message often conveys a negative opinion using only positive words – or even intensified, hyperbolic positive words. Likewise, but less frequently, sarcasm can flip the polarity of an opinion with negative words to the intended positive meaning. The detection of sarcasm is therefore important, if not crucial, for the development and refinement of sentiment analysis systems, but is at the same time a serious conceptual and technical challenge.

In this article we introduce a sarcasm detection system for tweets, messages on the microblogging service offered by Twitter.¹ In doing this we are helped by the fact that sarcasm appears to be a commonly recognized concept by many Twitter users, who explicitly mark their sarcastic messages by using hashtags such as '#sarcasm' or '#not'. Hashtags in tweets are explicitly marked keywords, and often act as categorical labels or metadata in addition to the body text of the tweet (Chang, 2010). By using the explicit hashtag any remaining doubt a reader may have is taken away: the message is not to be taken literally; it is sarcastic.

* Corresponding author. Address: Centre for Language Studies/CIW, Faculty of Arts, Radboud University Nijmegen, P.O. Box 9103, NL-6500 HD Nijmegen, The Netherlands. Tel.: +31 24 3611647.

E-mail address: a.vandenbosch@let.ru.nl (A. van den Bosch).

¹ <http://www.twitter.com>.

While such hashtags primarily function as conversational markers of sarcasm, they can be leveraged as annotation labels in order to generate a model of sarcastic tweets from the text co-occurring with these hashtags. A clear advantage of this approach is the easy acquisition of a vast amount of training data. On the other hand, its performance is dependent on the correctness of two assumptions: first that users who include one of the selected hashtags in their tweet actually intended to convey sarcasm and indeed intended to flip the polarity of the message, and second that the pattern of sarcasm in a tweet still holds when the hashtag is excluded from it as a training label. We set out to test these assumptions along with the quality of the resulting sarcasm detection system by applying it on a realistically large and unbiased sample of tweets (of which the vast majority is non-sarcastic) posted on the same day.

The hashtag as a marker of sarcasm has been leveraged in previous research to detect sarcasm in tweets (González-Ibáñez, Muresan, & Wacholder, 2011; Reyes, Rosso, & Buscaldi, 2012). One contribution of this paper to the existing body of work is that a sarcasm classifier is trained on several markers of sarcasm in tandem, the most frequent being '#not', and performance is assessed on a realistically large and unbiased sample of tweets. Furthermore, we provide insight into the role of hyperbole in sarcastic tweets, and we perform a cross-lingual comparison of the use of sarcasm in Twitter by annotating French tweets ending with '#sarcasme'.

This paper is structured as follows. In Section 1.1 we discuss the concepts of sarcasm, the broader category of verbal irony, and their communicative function according to the literature. In Section 2 we offer a brief survey of related work on the development of automatic detectors of polarity in social media. Our experimental setup is described in Section 3. We report on the results of our experiments in Section 4 and analyse our results in view of the theoretical work discussed earlier in Section 5. We summarize our results, draw conclusions, and identify points for future research in Section 6.

1.1. Definitions

Twitter members mark their sarcastic messages with different hashtags. As described in more detail in Section 3.1, we find that four words tend to be used as hashmarks in sarcastic posts: '#sarcasm', '#irony', '#cynicism' and '#not'. Although sarcasm, irony and cynicism are not synonymous, they have much in common. This is especially true for sarcasm and irony; many researchers treat those phenomena as strongly related (Attardo, 2007; Brown, 1980; Gibbs & O'Brien, 1991; Kreuz & Roberts, 1993; Mizzau, 1984; Muecke, 1969), and sometimes even equate the terms in their studies in order to work with a usable definition (Grice, 1978; Tsur, Davidov, & Rappoport, 2010). Cynicism is more mocking and tells us more about human beliefs than irony and sarcasm (Eisinger, 2000), but there is a close correlation between these concepts (Yoos, 1985). The hashtag '#not' is not the name of a rhetorical device or trope such as sarcasm, irony and cynicism, but it is a conventionally used meta-communication marker to indicate the message contains a shift in evaluative valence.

In psycholinguistics and cognitive linguistics sarcasm has been widely studied, often in relation with concepts such as cynicism, and with verbal irony as a broader category term. A brief overview of definitions, hypotheses and findings from communication studies regarding sarcasm and related concepts may help clarify what the hashtags convey.

In this study, we are interested in sarcasm as a linguistic phenomenon, and how we can detect it in social media messages. Yet, Brown (1980) warns that sarcasm 'is not a discrete logical or linguistic phenomenon' (p. 111), while verbal irony is. Indeed, Reyes and Rosso (2012) see sarcasm 'as specific extension[s] of a general concept of irony' (p. 755). In line with the extensive use of #sarcasm in tweets to mark verbal irony, we take the liberty of using the term sarcasm while verbal irony would be the more appropriate term. Even then, according to Gibbs and Colston (2007) the definition of verbal irony is still a 'problem that surfaces in the irony literature' (p. 584).

There are many different theoretical approaches to verbal irony. It should (a) be evaluative, (b) be based on incongruence of the ironic utterance with the co-text or context, (c) be based on a reversal of valence between the literal and intended meaning, (d) be aimed at some target, and (e) be relevant to the communicative situation in some way (Burgers, Van Mulken, & Schellens, 2011). Although it is known that irony is always directed at someone or something (the sender himself, the addressee, a third party, or a combination of the three, see Burgers et al. (2011) and Livnat (2004)) and irony is used relatively often in dialogic interaction (Gibbs, 2007), these two elements of irony are hardly examinable in the case of Twitter: the context of the Twitter messages is missing and it is inconvenient to investigate interaction. Therefore, it is hard to interpret the communicative situation and the target of the message. However, it is possible to analyse texts, such as tweets, on their evaluative meaning and a potential valence shift in the same way as Burgers et al. (2011) did. Burgers et al.'s own definition of verbal irony is 'an utterance with a literal evaluation that is implicitly contrary to its intended evaluation.' (p. 190).

Thus, a sarcastic utterance involves a shift in evaluative valence, which can go two ways: it could be a shift from a literally positive to an intended negative meaning, or a shift from a literally negative to an intended positive evaluation. Since Reyes, Rosso, and Veale (2013) also argue that users of social media often use irony in utterances that involve a shift in evaluative valence, we use the definition of verbal irony of Burgers et al. (2011) in this study on sarcasm, and we use both terms synonymously. The definition of irony as saying the opposite of what is meant is commonly used in previous corpus-analytic studies, and is reported to be reliable (Kreuz, Roberts, Johnson, & Bertus, 1996; Leigh, 1994; Srinarawat, 2005).

In order to ensure that the addressees detect the sarcasm in the utterance, senders use markers in their utterances. Attardo (2000) states that those markers are clues a writer can give that 'alert a reader to the fact that a sentence is ironical' (p. 7). The use of markers in written and spoken interaction may be different (Jahandarie, 1999). In spoken interaction, sarcasm is often marked with a special intonation (Attardo, Eisterhold, Hay, & Poggi, 2003; Bryant & Tree, 2005; Rockwell, 2007), air quotes (Attardo, 2000) or an incongruent facial expression (Attardo et al., 2003; Muecke, 1978; Rockwell,

Download English Version:

<https://daneshyari.com/en/article/10355183>

Download Persian Version:

<https://daneshyari.com/article/10355183>

[Daneshyari.com](https://daneshyari.com)