The perfect solution for detecting sarcasm in tweets #not

Christine Liebrecht
Centre for Language Studies
Radboud University Nijmegen
P.O. Box 9103
NL-6500 HD Nijmegen
c.liebrecht@let.ru.nl

Florian Kunneman
Centre for Language Studies
Radboud University Nijmegen
P.O. Box 9103
NL-6500 HD Nijmegen
f.kunneman@let.ru.nl

Antal van den Bosch
Centre for Language Studies
Radboud University Nijmegen
P.O. Box 9103
NL-6500 HD Nijmegen
a.vandenbosch@let.ru.nl

Abstract

To avoid a sarcastic message being understood in its unintended literal meaning, in microtexts such as messages on Twitter.com sarcasm is often explicitly marked with the hashtag '#sarcasm'. We collected a training corpus of about 78 thousand Dutch tweets with this hashtag. Assuming that the human labeling is correct (annotation of a sample indicates that about 85% of these tweets are indeed sarcastic), we train a machine learning classifier on the harvested examples, and apply it to a test set of a day's stream of 3.3 million Dutch tweets. Of the 135 explicitly marked tweets on this day, we detect 101 (75%) when we remove the hashtag. We annotate the top of the ranked list of tweets most likely to be sarcastic that do not have the explicit hashtag. 30% of the top-250 ranked tweets are indeed sarcastic. Analysis shows that sarcasm is often signalled by hyperbole, using intensifiers and exclamations; in contrast, non-hyperbolic sarcastic messages often receive an explicit marker. 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.

1 Introduction

In the general area of sentiment analysis, sarcasm plays a role as an interfering factor that can flip the polarity of a message. Unlike a simple negation, a sarcastic message typically conveys a negative opinion using only positive words – or even intensified positive words. 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 paper 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 well-understood concept by Twitter users, as seen by the relatively accurate use of an explicit marker of sarcasm, the hashtag '#sarcasm'. Hashtags in messages on Twitter (tweets) are explicitly marked keywords, and often act as categorical labels or metadata in addition to the body text of the tweet. By using the explicit hashtag any remaining doubt a reader may have is taken away: the message is intended as sarcastic.

In communication studies, sarcasm has been widely studied, often in relation with, or encompassed by concepts such as irony as a broader category term, and in particular in relation with (or synonymous to) verbal irony. A brief overview of definitions, hypotheses and findings from communication studies regarding sarcasm and verbal irony may help clarify what the hashtag '#sarcasm' conveys.

1.1 Definitions

Many researchers treat irony and sarcasm as strongly related (Attardo, 2007; Brown, 1980; Gibbs and O'Brien, 1991; Kreuz and Roberts, 1993; Muecke, 1969; Mizzau, 1984), and sometimes even equate the terms in their studies in order to work with an usable definition (Grice, 1978; Tsur et al., 2010). We are interested in sarcasm as a linguistic phenomenon, and how we can detect it in social me-

[1] http://www.twitter.com

Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 29–37,
Atlanta, Georgia, 14 June 2013.

dia messages. Yet, Brown (1980) warns that sarcasm 'is not a discrete logical or linguistic phenomenon' (p. 111), while verbal irony is; 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. Burgers (2010), who provides an overview of approaches, distinguishes a number of features in ironic utterances that need to be included in an operational definition of irony: (1) irony is always implicit (Giora, 1995; Grice, 1978), (2) irony is evaluative (Attardo, 2000; Kotthoff, 2003; Sperber and Wilson, 1995), it is possible to (3) distinguish between a non-ironic and an ironic reading of the same utterance (Grice, 1975; Grice, 1978), (4) between which a certain type of opposition may be observed (see also Kawakami, 1984, 1988, summarized in (Hamamoto, 1998; Partington, 2007; Seto, 1998). Burgers' own definition of verbal irony is 'an evaluative utterance, the valence of which is implicitly reversed between the literal and intended evaluation' (Burgers, 2010, p. 19).

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 et al. (2012b) also argue that users of social media often use irony in utterances that involve a shift in evaluative valence, we use Burgers' (2010) definition of verbal irony 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 et al., 1996; Leigh, 1994; Srinarawat, 2005).

Irony is used relatively often in dialogic interaction. Around 8% of conversational turns between American college friends contains irony (Gibbs, 2007). According to Gibbs (2007), group members use irony to 'affirm their solidarity by directing comments at individuals who are not group members and not deemed worthy of group membership' (p. 341). When an individual sees a group's normative standards violated, he uses sarcasm to vent frustration. Sarcasm is also used when someone finds a situation or object offensive (Gibbs, 2007). Sarcasm or irony is always directed at someone or something; its target. A target is the person or object against whom or which the ironic utterance is directed (Livnat, 2004). Targets can be the sender himself, the addressee or a third party (or a combination of the three). Burgers (2010) showed that in Dutch written communication, the target of the ironic utterance is often a third party. These findings may be interesting for our research, in which we study microtexts of up to 140 characters from Twitter.

Sarcasm in written and spoken interaction may work differently (Jahandarie, 1999). In spoken interaction, sarcasm is often marked with a special intonation (Attardo et al., 2003; Bryant and Tree, 2005; Rockwell, 2007) or an incongruent facial expression (Muecke, 1978; Rockwell, 2003; Attardo et al., 2003). Burgers (2010) argues that in written communication, authors do not have clues like 'a special intonation' or 'an incongruent facial expression' at their disposal. Since sarcasm is more difficult to comprehend than a literal utterance (Gibbs, 1986; Giora, 2003; Burgers, 2010), it is likely that addressees do not pick up on the sarcasm and interpret the utterances literally. Acoording to Gibbs and Izett (2005), sarcasm divides its addressees into two groups; a group of people who understand sarcasm (the so-called group of *wolves*) and a group of people who do not understand sarcasm (the so-called group of *sheep*). In order to ensure that the addressees detect the sarcasm in the utterance, senders use linguistic markers in their utterances. According to Attardo (2000) those markers are clues a writer can give that 'alert a reader to the fact that a sentence is ironical' (p. 7). On Twitter, the hashtag '#sarcasm' is a popular marker.

1.2 Intensifiers

There are sarcastic utterances which would still be qualified as sarcastic when all markers were removed from it (Attardo et al., 2003), for example the use of a hyperbole (Kreuz and Roberts, 1995). It may be that a sarcastic utterance with a hyperbole ('fantastic weather' when it rains) is identified as sarcastic with more ease than a sarcastic utterance without a hyperbole ('the weather is good' when it rains). While both utterances convey a lit-

erally positive attitude towards the weather, the utterance with the hyperbolic 'fantastic' may be easier to interpret as sarcastic than the utterance with the non-hyperbolic 'good'. Such hyperbolic words which strengthen the evaluative utterance are called intensifiers. Bowers (1964) defines language intensity as 'the quality of language which indicates the degree to which the speaker's attitude toward a concept deviates from neutrality' (p. 416). According to Van Mulken and Schellens (2012), an intensifier is a linguistic element that can be removed or replaced while respecting the linguistic correctness of the sentence and context, but resulting in a weaker evaluation. A commonly used way to intensify utterances is by using word classes such as adverbs ('very') or adjectives ('fantastic' instead of 'good'). It may be that senders use such intensifiers in their tweets to make the utterance hyperbolic and thereby sarcastic, without using a linguistic marker such as '#sarcasm'.

1.3 Outline

In this paper we describe the design and implementation of a sarcasm detector that marks unseen tweets as being sarcastic or not. We analyse the predictive performance of the classifier by testing its capacity on test tweets that are explicitly marked with the hashtag #sarcasme (Dutch for 'sarcasm'), left out during testing, and its capacity to rank likely sarcastic tweets that do not have the #sarcasme mark. We also provide a qualitative linguistic analysis of the features that the classifier thinks are the most discriminative. In a further qualitative analysis of sarcastic tweets in the test set we find that the use of an explicit hashtag marking sarcasm occurs relatively often without other indicators of sarcasm such as intensifiers or exclamations.

2 Related Research

The automatic classification of communicative constructs in short texts has become a widely researched subject in recent years. Large amounts of opinions, status updates and personal expressions are posted on social media platforms such as Twitter. The automatic labeling of their polarity (to what extent a text is positive or negative) can reveal, when aggregated or tracked over time, how the public in general thinks about certain things. See Montoyo et al. (2012) for an overview of recent research in sentiment analyis and opinion mining.

A major obstacle for automatically determining the polarity of a (short) text are constructs in which the literal meaning of the text is not the intended meaning of the sender, as many systems for the detection of polarity primarily lean on positive and negative words as markers. The task to identify such constructs can improve polarity classification, and provide new insights into the relatively new genre of short messages and microtexts on social media. Previous works describe the classification of irony (Reyes et al., 2012b), sarcasm (Tsur et al., 2010), satire (Burfoot and Baldwin, 2009), and humor (Reyes et al., 2012a).

Most common to our research are the works by Reyes et al. (2012b) and Tsur et al. (2010). Reyes et al. (2012b) collect a training corpus of irony based on tweets that consist of the hashtag #irony in order to train classifiers on different types of features (signatures, unexpectedness, style and emotional scenarios) and try to distinguish #irony-tweets from tweets containing the hashtags #education, #humour, or #politics, achieving F1-scores of around 70. Tsur et al. (2010) focus on product reviews on the World Wide Web, and try to identify sarcastic sentences from these in a semi-supervised fashion. Training data is collected by manually annotating sarcastic sentences, and retrieving additional training data based on the annotated sentences as queries. Sarcasm is annotated on a scale from 1 to 5. As features, Tsur et al. look at the patterns in these sentences, consisting of high-frequency words and content words. Their system achieves an F1-score of 79 on a testset of product reviews, after extracting and annotating a sample of 90 sentences classified as sarcastic and 90 sentences classified as not sarcastic.

In the two works described above, a system is tested in a controlled setting: Reyes et al. (2012b) compare irony to a restricted set of other topics, while Tsur et al. (2010) took from the unlabeled test set a sample of product reviews with 50% of the sentences classified as sarcastic. In contrast, we apply a trained sarcasm detector to a real-world test set representing a realistically large sample of tweets posted on a specific day of which the vast majority is not sarcastic. Detecting sarcasm in social media is,

arguably, a needle-in-a-haystack problem (of the 3.3 million tweets we gathered on a single day, 135 are explicitly marked with the hashtag #sarcasm), and it is only reasonable to test a system in the context of a typical distribution of sarcasm in tweets. Like in the research of (Reyes et al., 2012b), we train a classifier based on tweets with a specific hashtag.

3 Experimental Setup

3.1 Data

For the collection of tweets for this study we make use of a database provided by the Netherlands e-Science Centre, consisting of a substantial portion of all Dutch tweets posted from December 2010 onwards.[2] From this database, we collected all tweets that contained the marker '#sarcasme', the Dutch word for sarcasm with the hashtag prefix. This resulted in a set of 77,948 tweets. We also collected all tweets posted on a single day, namely February 1, 2013.[3] This set of tweets contains approximately 3,3 million tweets, of which 135 carry the hashtag #sarcasme.

3.2 Winnow classification

Both the collected tweets with a #sarcasme hashtag and the tweets that were posted on a single day were tokenized and stripped of punctuation. Capitals were not removed, as they might be used to signal sarcasm (Burgers, 2010). We made use of word uni-, bi- and trigrams as features. Terms that occurred three times or less or in two tweets or less in the whole set were removed, as well as the hashtag #sarcasme. Features were weighted by the χ^2 metric.

As classification algorithm we made use of Balanced Winnow (Littlestone, 1988) as implemented in the Linguistic Classification System.[4] This algorithm is known to offer state-of-the-art results in text classification, and produces interpretable per-class weights that can be used to, for example, inspect the highest-ranking features for one class label. The α and β parameters were set to 1,05 and 0,95 respectively. The major threshold ($\theta+$) and the minor threshold ($\theta-$) were set to 2,5 and 0,5. The number of iterations was bounded to a maximum of three.

3.3 Experiment

In order to train the classifier on distinctive features of sarcasm in tweets, we combined the set of 78 thousand sarcasm tweets with a random sample of other tweets posted on February 1, 2013 as background corpus. We made sure the background corpus did not contain any of the 135 explicitly marked sarcasm tweets posted that day. As the size of a background corpus can influence the performance of the classifier (in doubt, a classifier will be biased by the skew of the distribution of classes in the training-data), we performed a comparitive experiment with two distributions between sarcasm-labeled tweets and background tweets: in the first variant, the division between the two is 50%–50%, in the second, 25% of the tweets are sarcasm-labeled, and 75% are background.

4 Results

To evaluate the outcome of our machine learning experiment, we ran two evaluations. The first evaluation focuses on the 135 tweets with explicit #sarcasme hastags posted on February 1, 2013. We measured how well these tweets were identified using the true positive rate (TPR), false positive rate (FPR, also known as recall), and their joint score, the area under the curve (AUC). AUC is a common evaluation metric that is argued to be more resistant to skew than F-score, due to using TPR rather than precision (Fawcett, 2004). Results are displayed in Table 1. The first evaluation, on the variant with a balanced distribution of the two classes, leads to a retrieval of 101 of the 135 sarcasm-tweets (75%), while nearly 500 thousand tweets outside of these were also classified as being sarcastic. When a quarter of the training tweets has a sarcasm label, a smaller amount of 76 sarcasm tweets are retrieved. The AUC scores for the two ratios indicates that the 50%–50% balance leads to the highest AUC score (0.79) for sarcasm. Our subsequent analyses are based on the outcomes when using this distribution in training.

Besides generating an absolute winner-take-all classification, our Balanced Winnow classifier also assigns scores to each label that can be seen as its

[2] http://twiqs.nl/

[3] All tweets from February 1, 2013 onwards were removed from the set of sarcasm tweets.

[4] http://www.phasar.cs.ru.nl/LCS/

Pos/Neg Ratio Training Examples	Label	# Training tweets	# Test tweets	TPR	FPR	AUC	Classified	Correct
50/50	sarcasm	77,948	135	0,75	0,16	0,79	487,955	101
	background	77,499	3,246,806	0,79	0,25	0,77	2,575,206	2,575,173
25/75	sarcasm	77,948	135	0,56	0,05	0,75	162,400	76
	background	233,834	3,090,472	0,92	0,43	0,74	2,830,103	2,830,045

Table 1: Scores on the test set with two relative sizes of background tweets (TPR = True Positive Rate, FPR = False Positive Rate, AUC = Area Under the Curve

confidence in that label. We can rank its predictions by the classifier's confidence on the 'sarcasm' label and inspect manually which of the top-ranking tweets is indeed sarcastic. We generated a list of the 250 most confident 'sarcasm'-labeled tweets. Three annotators (the authors of this paper) made a judgement for these tweets as being either sarcastic or not. In order to test for intercoder reliability, Cohen's Kappa was used. In line with Siegel and Castellan (1988), we calculated a mean Kappa based on pairwise comparisons of all possible coder pairs. The mean intercoder reliability between the three possible coder pairs is substantial ($\kappa = .79$).

When taking the majority vote of the three annotators as the golden label, a curve of the precision at all points in the ranking can be plotted. This curve is displayed in Figure 1. As can be seen, the overall performance is poor (the average precision is 0.30). After peaking at 0.50 after 22 tweets, precision slowly decreases when descending to lower rankings. During the first five tweets, the curve is at 0.0; these tweets, receiving the highest overall confidence scores, are relatively short and contain one strong sarcasm feature in the classifier without any negative feature.

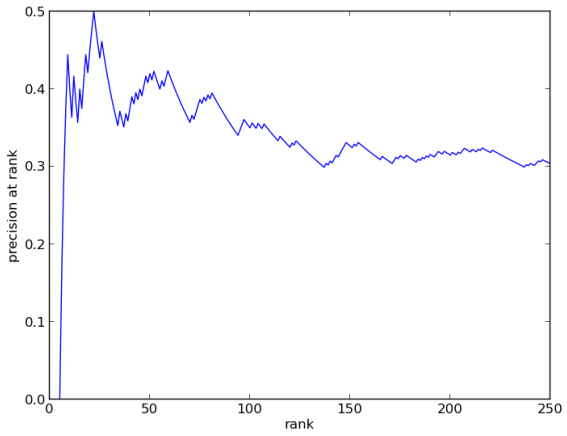

Figure 1: Precision at $\{1 \ldots 250\}$ on the sarcasm class

5 Analysis

Our first closer analysis of our results concerns the reliability of the user-generated hastag #sarcasme as a golden label, as Twitter users cannot all be assumed to be experts in sarcasm or understand what sarcasm is. The three annotators who annotated the ranked classifier output also coded a random sample of 250 tweets with the #sarcasme hashtag from the training set. The average score of agreement between the three possible coder pairs turned out to be moderate ($\kappa = .54$). Taking the majority vote over the three annotations as the reference labeling, 85% (212) of the 250 annotated #sarcasme tweets were found to be sarcastic.

While the classifier performance gives an impression of its ability to distinguish sarcastic tweets, the strong indicators of sarcasm as discovered by the classifier may provide additional insight into the usage of sarcasm by Twitter users: in particular, the typical targets of sarcasm, and the different linguistic markers that are used. We thus set out to analyze the feature weights assigned by the Balanced Winnow classifier ranked by the strength of their connection to the sarcasm label, taking into account the 500 words and n-grams with the highest positive weight towards the sarcasm class. These words and n-grams provide insight into the topics Twitter users are talking about: their targets. People often talk about school and related subjects such as homework, books, exams, classes (French, chemistry, physics), teachers, the school picture, sports day, and (returning from) vacation. Another popular target of sar-

casm is the weather: the temperature, rain, snow, and sunshine. Apart from these two common topics, people tend to be sarcastic about social media itself, holidays, public transport, soccer, television programs (The Voice of Holland), celebrities (Justin Bieber), the church, the dentist and vacuum cleaning. Many of these topics are indicative of the young age, on average, of Twitter users.

The strongest linguistic markers of sarcastic utterances are markers that can be seen as synonyms for *#sarcasme*, such as *sarcasme* (without #), *#ironie* and *ironie* (irony), *#cynisme* and *cynisme* (cynicism), or words that are strongly related to those concepts by marking the opposite of the expressed utterance: *#humor*, *#LOL*, *#joke* (grapje), and *#NOT*.

Second, the utterances contain much positive exclamations that make the utterance hyperbolic and thereby sarcastic. Examples of those markers in Dutch are (with and without # and/or capitals): *jippie, yes, goh, joepie, jeej, jeuj, yay, woehoe*, and *wow*.

We suspected that the sarcastic utterances contained intensifiers to make the tweets hyperbolic. The list of strongest predictors show that some intensifiers are indeed strong predictors of sarcasm, such as *geweldig* (awesome), *heerlijk* (lovely), *prachtig* (wonderful), *natuurlijk* (of course), *gelukkig* (fortunately), *zoooo* (soooo), *allerleukste* (most fun), *fantastisch* (fantastic), and *heeel* (veeery). Besides these intensifiers many unmarked positive words occur in the list of strongest predictors as well, such as *fijn* (nice), *gezellig* (cozy), *leuk* (fun), *origineel* (original), *slim* (smart), *favoriet* (favorite), *nuttig* (useful), and chill. Considerably less negative words occur as strong predictors. This supports our hypothesis that the utterances are mostly positive, while the opposite meaning is meant. This finding corresponds with the results of Burgers (2010), who show that 77% of the ironic utterances in Dutch communication are literally positive.

To inspect whether sarcastic tweets are always intensified to be hyperbolic, we need to further analyse the sarcastic tweets our classifier correctly identifies. Analyzing the 76 tweets that our classifier correctly identifies in the top-250 tweets the classifier rates as sarcastic, we see that intensifiers do not dominate in occurrence; supporting numbers are listed in Table 2. About one in three sarcastic tweets, 34.2%, are not hyperbolic at all: they are only explicitly marked, most of the times with a hashtag. A majority of 59.2% of the tweets does contain hyperbole-inducing elements, such as an intensifier or an exclamation, or combinations of these elements. A full combination of explicit markers, intensifiers, and exclamations only rarely occurs, however (2.6%). The three categories of predictive word types do cover 93.4% of the tweets.

Type	Relative occurrence (%)
Marker only	34.2
Intensifier only	9.2
Exclamation only	17.1
Marker + Intensifier	10.5
Marker + Exclamation	9.2
Intensifier + Exclamation	10.5
Marker + Intensifier + Exclamation	2.6
Other	6.6
Total	*100*

Table 2: Relative occurrence (%) of word types and their combinations in the tweets annotated as sarcastic by a majority vote.

6 Conclusion

In this study we developed and tested a system that detects sarcastic tweets in a realistic sample of 3.3 million Dutch tweets posted on a single day, trained on a set of nearly 78 thousand tweets, harvested over time, marked by the hashmark #sarcasme by the senders. The classifier is able to correctly detect 101 of the 135 tweets among the 3.3 million that were explicitly marked with the hashtag, with the hashtag removed. Testing the classifier on the top 250 of the tweets it ranked as most likely to be sarcastic, it attains only a 30% average precision. We can conclude that it is fairly hard to distinguish sarcastic tweets from literal tweets in an open setting, though the top of the classifier's ranking does identify many sarcastic tweets which were not explicitly marked with a hashtag.

An additional linguistic analysis provides some insights into the characteristics of sarcasm on Twitter. We found that most tweets contain a literally

positive message, take common teenager topics as target (school, homework, family life) and further contain three types of words: explicit markers (the word *sarcasme* and pseudo-synonyms, with or without the hashmark #), intensifiers, and exclamations. The latter two categories of words induce hyperbole, but together they only occur in about 60% of sarcastic tweets; in 34% of the cases, sarcastic tweets are not hyperbolic, but only have an explicit marker, most of which hashtags. This indicates that the hashtag can and does replace linguistic markers that otherwise would be needed to mark sarcasm. Arguably, extralinguistic elements such as hashtags can be seen as the social media equivalent of non-verbal expressions that people employ in live interaction when conveying sarcasm. As Burgers (2010) show, the more explicit markers an ironic utterance contains, the better the utterance is understood, the less its perceived complexity is, and the better it is rated. Many Twitter users already seem to apply this knowledge.

Although in this research we focused on the Dutch language, our findings may also apply to languages similar to Dutch, such as English and German. Future research would be needed to chart the prediction of sarcasm in languages that are more distant to Dutch. Sarcasm may be used differently in other cultures (Goddard, 2006). Languages may use the same type of marker in different ways, such as a different intonation in spoken sarcasm by English and Cantonese speakers (Cheang and Pell, 2009). Such a difference between languages in the use of the same marker may also apply to written sarcastic utterances.

Another strand of future research would be to expand our scope from sarcasm to other more subtle variants of irony, such as understatements, euphemisms, and litotes. Based on Giora et al. (2005), there seems to be a spectrum of degrees of irony from the sarcastic 'Max is exceptionally bright' via the ironic 'Max is not exceptionally bright', the understatement 'Max is not bright' to the literal 'Max is stupid'. In those utterances, there is a gap between what is literally said and the intended meaning of the sender. The greater the gap or contrast, the easier it is to perceive the irony. But the negated *not bright* is still perceived as ironic; more ironic than the literal utterance (Giora et al., 2005). We may need to combine the sarcasm detection task with the problem of the detection of negation and hedging markers and their scope (Morante et al., 2008; Morante and Daelemans, 2009) in order to arrive at a comprehensive account of polarity-reversing mechanisms, which in sentiment analysis is still highly desirable.

References

S. Attardo, J. Eisterhold, J. Hay, and I. Poggi. 2003. Visual markers of irony and sarcasm. *Humor*, 16(2):243–260.

S. Attardo. 2000. Irony as relevant inappropriateness. *Journal of Pragmatics*, 32(6):793–826.

S. Attardo. 2007. Irony as relevant inappropriateness. In R. W. Gibbs, R. W. Gibbs Jr., and H. Colston, editors, *Irony in language and thought: A cognitive science reader*, pages 135–170. Lawrence Erlbaum, New York, NY.

J. W. Bowers. 1964. Some correlates of language intensity. *Quarterly Journal of Speech*, 50(4):415–420, December.

R. L. Brown. 1980. The pragmatics of verbal irony. In R. W. Shuy and A. Shnukal, editors, *Language use and the uses of language*, pages 111–127. Georgetown University Press, Washington, DC.

G. A. Bryant and J. E. Fox Tree. 2005. Is there an ironic tone of voice? *Language and Speech*, 48(3):257–277.

C. Burfoot and T. Baldwin. 2009. Automatic satire detection: Are you having a laugh? In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 161–164. Association for Computational Linguistics.

C. F. Burgers. 2010. *Verbal irony: Use and effects in written discourse*. Ipskamp, Nijmegen, The Netherlands.

H. S. Cheang and M. D. Pell. 2009. Acoustic markers of sarcasm in Cantonese and English. *The Journal of the Acoustical Society of America*, 126:1394.

T. Fawcett. 2004. ROC graphs: Notes and practical considerations for researchers. Technical Report HPL-2003-4, Hewlett Packard Labs.

R. W. Gibbs and H. Colston. 2007. Irony as persuasive communication. In R. W. Gibbs, R. W. Gibbs Jr., and H. Colston, editors, *Irony in language and thougt: A cognitive science reader*, pages 581–595. Lawrence Erlbaum, New York, NY.

R. W. Gibbs and C. Izett. 2005. Irony as persuasive communication. In H. Colston and A. Katz, editors, *Figurative language comprehension: Social and cultural influences*, pages 131–151. Lawrence Erlbaum, New York, NY.

R. W. Gibbs and J. O'Brien. 1991. Psychological aspects of irony understanding. *Journal of pragmatics*, 16(6):523–530.

R. W. Gibbs. 1986. On the psycholinguistics of sarcasm. *Journal of Experimental Psychology: General*, 115(1):3.

R. W. Gibbs. 2007. On the psycholinguistics of sarcasm. In R. W. Gibbs, R. W. Gibbs Jr., and H. Colston, editors, *Irony in language and thougt: A cognitive science reader*, pages 173–200. Lawrence Erlbaum, New York, NY.

R. Giora, O. Fein, J. Ganzi, N. Levi, and H. Sabah. 2005. On negation as mitigation: the case of negative irony. *Discourse Processes*, 39(1):81–100.

R. Giora. 1995. On irony and negation. *Discourse processes*, 19(2):239–264.

R. Giora. 2003. *On our mind: Salience, context, and figurative language*. Oxford University Press.

C. Goddard. 2006. "lift your game Martina!": Deadpan jocular irony and the ethnopragmatics of Australian English. *APPLICATIONS OF COGNITIVE LINGUISTICS*, 3:65.

H. Grice. 1975. Logic and conversation. In P. Cole and J. Morgan, editors, *Speech acts: Syntax and semantics*, pages 41–58. Academic Press, New York, NY.

H. Grice. 1978. Further notes on logic and conversation. In P. Cole, editor, *Pragmatics: syntax and semantics*, pages 113–127. Academic Press, New York, NY.

H. Hamamoto. 1998. Irony from a cognitive perspective. In R. Carston and S. Uchida, editors, *Relevance theory: Applications and implications*, pages 257–270. John Benjamins, Amsterdam, The Netherlands.

K. Jahandarie. 1999. *Spoken and written discourse: A multi-disciplinary perspective*. Greenwood Publishing Group.

H. Kotthoff. 2003. Responding to irony in different contexts: On cognition in conversation. *Journal of Pragmatics*, 35(9):1387–1411.

R. J. Kreuz and R. M. Roberts. 1993. The empirical study of figurative language in literature. *Poetics*, 22(1):151–169.

R. J. Kreuz and R. M. Roberts. 1995. Two cues for verbal irony: Hyperbole and the ironic tone of voice. *Metaphor and symbol*, 10(1):21–31.

R. Kreuz, R. Roberts, B. Johnson, and E. Bertus. 1996. Figurative language occurrence and co-occurrence in contemporary literature. In R. Kreuz and M. MacNealy, editors, *Empirical approaches to literature and aesthetics*, pages 83–97. Ablex, Norwood, NJ.

J. H Leigh. 1994. The use of figures of speech in print ad headlines. *Journal of Advertising*, pages 17–33.

N. Littlestone. 1988. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2:285–318.

Z. Livnat. 2004. On verbal irony, meta-linguistic knowledge and echoic interpretation. *Pragmatics & Cognition*, 12(1):57–70.

M. Mizzau. 1984. *L'ironia: la contraddizione consentita*. Feltrinelli, Milan, Italy.

A. Montoyo, P. Martínez-Barco, and A. Balahur. 2012. Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*.

R. Morante and W. Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. In *Proceedings of the Workshop on BioNLP*, pages 28–36. Association for Computational Linguistics.

R. Morante, A. Liekens, and W. Daelemans. 2008. Learning the scope of negation in biomedical texts. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 715–724.

D. C. Muecke. 1969. *The compass of irony*. Oxford Univ Press.

D. C. Muecke. 1978. Irony markers. *Poetics*, 7(4):363–375.

A. Partington. 2007. Irony and reversal of evaluation. *Journal of Pragmatics*, 39(9):1547–1569.

A. Reyes, P. Rosso, and D. Buscaldi. 2012a. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*.

A. Reyes, P. Rosso, and T. Veale. 2012b. A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, pages 1–30.

P. Rockwell. 2003. Empathy and the expression and recognition of sarcasm by close relations or strangers. *Perceptual and motor skills*, 97(1):251–256.

P. Rockwell. 2007. Vocal features of conversational sarcasm: A comparison of methods. *Journal of psycholinguistic research*, 36(5):361–369.

K.-i. Seto. 1998. On non-echoic irony. In R. Carson and S. Uchida, editors, *Relevance theory: Applications and implications*, pages 239–255. John Benjamins, Amsterdam, The Netherlands.

S. Siegel and N. Castellan. 1988. *Nonparametric statistics for the behavioral sciences*. McGraw Hill, New York.

D. Sperber and D. Wilson. 1995. *Relevance: Communication and cognition*. Blackwell Publishers, Oxford, UK, 2nd edition.

D. Srinarawat. 2005. Indirectness as a politeness strategy of Thai speakers. In R. Lakoff and S. Ide, editors, *Broadening the horizon of linguistic politeness*, pages 175–193. John Benjamins, Amsterdam, The Netherlands.

O. Tsur, D. Davidov, and A. Rappoport. 2010. Icwsm–a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pages 162–169.

M. Van Mulken and P. J. Schellens. 2012. Over loodzware bassen en wapperende broekspijpen. gebruik en perceptie van taalintensiverende stijlmiddelen. *Tijdschrift voor taalbeheersing*, 34(1):26–53.